Université de Montréal

Essays in Functional Econometrics and Financial Markets

Par
Idriss TSAFACK

Département de sciences économiques
Faculté des arts et sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en sciences économiques

Juillet, 2020

Université de Montréal
Faculté des Études Supérieures et Postdoctorales

Cette thèse intitulée :

Essays in Functional Econometrics and Financial Markets

présentée par :

Idriss TSAFACK

sera évaluée par un jury composé des personnes suivantes :

René Garcia,	président du jury
Marine Carrasco,	directrice de recherche
Benoit Perron,	membre du jury
Joann Jasiak,	examinatrice externe

A ma bien aimée Lucienne Talba, mes parents Paul Tsafac, et Jeannette Donfack, aussi à mes freres et soeurs.

Remerciements

Je remercie grandement Marine Carrasco, ma directrice de recherche pour sa disponibilité sans mesure ainsi que ses conseils précieux et encouragements tout au long de cette thèse. J'ai enormément appris de sa rigueur et la tenacité dans les procédures scientifiques.

J'aimerais envoyer mes remerciements à Benoit Perron, pour ses conseils et son encadrement tout au long de cette période de thèse. Je tiens également à remercier René Garcia qui a bien voulu accepter de faire partie de mon comité de thèse depuis ma deuxième année de thèse. Je remercie aussi Marine Carrasco, Benoit Perron et William McCausland pour leurs soutiens durant mon année de marché de travail. Mes remerciements sont également adressés aux enseignants, chercheurs et au personnel administratif du département des sciences économiques et du CIREQ. Cette thèse a été réalisée avec l'appui financier du département et du CIREQ.

Je remercie également mes amis et collègues avec qui j'ai partagé cette étape importante de mon parcours académique. Merci à Lucienne, Leonel, Magnim, N'golo, Samuel, Marlène, Fatim, Marius, Herve Zeida, Herve Zongo, Marius, David, Jonathan, George, Louphou, William, Zerbo, Firmin, Ghislain, Felicien ainsi qu'à tous les autres collègues du CIREQ et du département avec qui nous avons participé aux ateliers de discussions des étudiants de doctorat. Je remercie également toutes les personnes dont j'ai fait la connaissance pendant mon séjour à Montréal et qui m'ont beaucoup soutenu et aidé. Enfin j'envoie un merci spécial à ma bien aimée, mes parents, mes frères et soeurs pour leur soutien inconditionnel tout au long de mes études.

Résumé

Dans cette thèse, j'exploite le cadre d'analyse de données fonctionnelles et développe l'analyse d'inférence et de prédiction, avec une application à des sujets sur les marchés financiers. Cette thèse est organisée en trois chapitres.

Le premier chapitre est un article co-écrit avec Marine Carrasco. Dans ce chapitre, nous considérons un modèle de régression linéaire fonctionnelle avec une variable prédictive fonctionnelle et une réponse scalaire. Nous effectuons une comparaison théorique des techniques d'analyse des composantes principales fonctionnelles (FPCA) et des moindres carrés partiels fonctionnels (FPLS). Nous déterminons la vitesse de convergence de l'erreur quadratique moyen d'estimation (MSE) pour ces méthodes. Aussi, nous montrons cette vitesse est sharp. Nous découvrons également que le biais de régularisation de la méthode FPLS est plus petit que celui de FPCA, tandis que son erreur d'estimation a tendance à être plus grande que celle de FPCA. De plus, nous montrons que le FPLS surpasse le FPCA en termes de prédiction avec moins de composantes.

Le deuxième chapitre considère un modèle autorégressif entièrement fonctionnel (FAR) pour prèvoir toute la courbe de rendement du S&P 500 a la prochaine journée. Je mène une analyse comparative de quatre techniques de Big Data, dont la méthode de Tikhonov fonctionnelle (FT), la technique de Landweber-Fridman fonctionnelle (FLF), la coupure spectrale fonctionnelle (FSC) et les moindres carrés partiels fonctionnels (FPLS). La vitesse de convergence, la distribution asymptotique et une stratégie de test statistique pour sélectionner le nombre de retard sont fournis. Les simulations et les données réelles montrent que les méthode FPLS performe mieux les autres en terme d'estimation du paramètre tandis que toutes ces méthodes affichent des performances similaires en termes de prédiction.

Le troisième chapitre propose d'estimer la densité de neutralité au risque (RND) dans le contexte de la tarification des options, à l'aide d'un modèle fonctionnel. L'avantage de cette approche est qu'elle exploite la théorie d'absence d'arbitrage et qu'il est possible d'éviter toute sorte de paramétrisation. L'estimation conduit à un problème d'inversibilité et la technique fonctionnelle de Landweber-Fridman (FLF) est utilisée pour le surmonter.

Mots-clés

Regression fonctionnelle, analyse de données fonctionnelles, modèle autoregressif fonctionnel, big data, régularisation, composantes principales fonctionnelle, moindres carrés partiels, Landweber-Fridman, Tikhonov, estimation, prédiction, prévision, S&P 500, options, probabilité de neutralité au risque, marchés financiers.

Abstract

In this thesis, I exploit the functional data analysis framework and develop inference, prediction and forecasting analysis, with an application to topics in the financial market. This thesis is organized in three chapters.

The first chapter is a paper co-authored with Marine Carrasco. In this chapter, we consider a functional linear regression model with a functional predictor variable and a scalar response. We develop a theoretical comparison of the Functional Principal Component Analysis (FPCA) and Functional Partial Least Squares (FPLS) techniques. We derive the convergence rate of the Mean Squared Error (MSE) for these methods. We show that this rate of convergence is sharp. We also find that the regularization bias of the FPLS method is smaller than the one of FPCA, while its estimation error tends to be larger than that of FPCA. Additionally, we show that FPLS outperforms FPCA in terms of prediction accuracy with a fewer number of components.

The second chapter considers a fully functional autoregressive model (FAR) to forecast the next day's return curve of the S&P 500. In contrast to the standard AR(1) model where each observation is a scalar, in this research each daily return curve is a collection of 390 points and is considered as one observation. I conduct a comparative analysis of four big data techniques including Functional Tikhonov method (FT), Functional Landweber-Fridman technique (FLF), Functional spectral-cut off (FSC), and Functional Partial Least Squares (FPLS). The convergence rate, asymptotic distribution, and a test-based strategy to select the lag number are provided. Simulations and real data show that FPLS method tends to outperform the other in terms of estimation accuracy while all the considered methods display almost the same predictive performance.

The third chapter proposes to estimate the risk neutral density (RND) for options pricing with a functional linear model. The benefit of this approach is that it exploits directly the fundamental arbitrage-free equation and it is possible to avoid any additional density parametrization. The estimation problem leads to an inverse problem and the functional Landweber-Fridman (FLF) technique is used to overcome this issue.

Keywords

Functional regression, functional data analysis, functional autoregressive model, big data, functional principal component, functional partial least squares, Landweber-Fridman, Tikhonov, estimation, prediction, forecasting, S&P 500, options, risk neutral density, financial markets.

Table des matières

Liste des tableaux

Table des figures

Chapitre 1

Theoretical Comparison of the Functional Principal Component Analysis and Functional Partial Least Squares *

1.1 Introduction

During the last decade, the world has experienced a huge evolution in technology, which made possible the extraction of very new insights from large data sets. One of the current goals in finance and economics is to find a way to take advantage of the availability of such huge data. Nowadays, we can observe the development of new statistical and econometric tools, such as the functional regression models. Those models have been widely used in the field of chemometric, medicine, and meteorology.

In this paper, we consider a functional regression model with a scalar response Y and a functional predictor variable $X(t)$.

$$Y = \int_S \beta(t)X(t)dt + \varepsilon$$

One of the main interests in such a model is the estimation of the slope function $\beta(t)$ and the prediction of the response variable. The main characteric of this model is that the predictor variable $X(t)$ belongs to a Hilbert space, which is an infinite-dimensional space, thereby causing the problem of functional parameter estimation in an infinite-dimensional space. The estimation of such functional parameter in a high-dimensional space leads to an ill-posed problem. To tackle this issue, most of the research usually suggest the use of some dimension reduction methods such as the Functional Principal Component Analysis

∗. This chapter is co-authored with Marine Carrasco. We thank René Garcia and Benoit Perron for their helpful comments.

(FPCA) (see Cai et al. (2006), Hall and Hosseini-Nasab (2006), Hall and Horowitz (2007) and Cardot et al. (1999)). Basically, the main point of this method is to run a regression of the response variable on a limited number of selected latent components. Those components are constructed in such a way that they capture the most important information related to the variation of the functional predictor variable.

On the other hand, a research by Jolliffe (1982) shows that one can sometimes deal with situations where the selected components by this method do not necessarily contribute well to the prediction or forecast of the response variable. This is why the Functional Partial Least Squares (FPLS) was introduced by Preda and Saporta (2005) and Reiss and Ogden (2007) in the functional regression context in order to extract the components that capture not only the most important information concerning the variations of the predictor variable, but also are the most able to predict or forecast the response variable. This approach has recently gained some popularity in the economic and financial fields ; then theoretical properties have remained elusive because of the nonlinearity of the estimator.

The aim of this paper is to compare the FPLS and FPCA techniques in the context of a functional linear regression with a scalar response. Both methods can be viewed as regularization techniques. To study the properties of FPLS, we use orthogonal polynomial theory. We derive the analytical formula of the slope function estimator in terms of a sequence of residual polynomials, the eigenvalues, and the eigenfunctions of the covariance operator of the predictor variable. We derive the convergence rate of the conditional MSE for both the FPCA and FPLS approaches under some regularity conditions. The rate of convergence of the conditional MSE with FPLS is the same as the one obtained with FPCA, which is $O_p(n^{-\mu/(\mu+2)})$. Also we show that this convergence rate is sharp. Moreover, we can notice from the theoretical comparison that the squared bias of the MSE obtained by using the FPLS is usually lower than the one obtained by the FPCA approach, while the estimation error obtained using the FPLS method tends to be larger than the one obtained by FPCA and may explode rapidly as the number of retained components increases. This result is due to the fact that the estimation error depends on the smallest root of the residual polynomial and this root has usually an exponential decay rate as the number of selected components increases. Additionally, we show that FPLS tends to outperform the FPCA method in terms of prediction and the optimal number of factors extracted via FPLS is usually smaller than or equal to that obtained by the FPCA. We derive Monte Carlo simulations and show that we replicate the theoretical results. For the empirical application, we consider the prediction of the next day return of the $S\&P$ 500 based on the previous day cumulative return and we confirm the estimation and prediction performance of the concerned methods.

Our paper is related to the literature concerning the theoretical properties of the estimated slope parameter in the functional linear regression context. More specifically, this reseach is related to the papers by Preda and Saporta (2005), Reiss and Ogden (2007), Delaigle and Hall (2012), Febrero-Bande et al. (2017). Indeed, concerning the study of the

consistency results with the FPLS methods, we can notice that only one paper attempted to derive such results, that is Delaigle and Hall (2012). In fact, the authors derived the rate of convergence of the mean squared prediction error (MSPE) by considering the formulation that they called "the Alternative PLS (APLS)". Moreover, the approach used by Delaigle and Hall (2012) is not useful when we wish to derive some comparisons of such methods theoretically. The difference between our paper and the one of Delaigle and Hall (2012) is that we adopt a different approach to observe the FPLS in such a way that we can compare the FPCA and FPLS based on theoretical results. Moreover, we derive the rate of convergence of the MSE for both methods under some conditions of the predictor and slope functions and we express it in terms of the sample size. Furthermore, Febrero-Bande et al. (2017) derived a comparison of the FPLS and FPCA approaches based on a Monte Carlo simulations and observed that the FPLS technique is more robust than the FPCA method most of the time and that the FPLS outperforms the FPCA in terms of estimation and prediction. They also stated that "due to the nonlinearity of the FPLS approach, its MSPE is theoretically intractable" (even after the results by Delaigle and Hall (2012)). Lastly we contribute to the literature by comparing theoretically both methods in terms of the number of retained components. This paper is also related to inverse problem literature and especially the book by Engl et al. (1996). The authors derived the theoretical results of the PLS also called conjugate gradient for an ill-posed problem. The difference with their research is that they considered a fixed design context where the covariance operator of predictor variables and the response variable are observed and the error between the estimated and the true response variable is bounded. This is not the case in our paper.

There is a large literature discussing the partial least squares (PLS) method on the multivariate linear regression models. This method is usually introduced when the number of predictor variables is larger than the number of observations and/or when there is a multicolinearity issue in the regression model. We can identify the main authors such as Wold et al. (1984), Höskuldsson (1988), Naes and Martens (1985), Helland (1988), Helland (1990), Helland and Almøy (1994), De Jong (1993). Lingjaerde and Christophersen (2000) derived some geometric and analytic properties of the filter factor of the PLS method in the context of multivariate linear regression. Some of the authors such as Helland (1990) and Rosipal and Krämer (2005) compared the PLS and the PCA based on their filter factor structure and asserted that the filter factor obtained with PLS is not a decreasing and convex function as we usually observe on other regularization methods. For the consistency results analysis, we can identify the papers of Helland (1990) and more recently, the works of Blazere et al. (2014a), Kelly and Pruitt (2015) and Carrasco and Rossi (2016). Besides the FPCA and FPLS methods, there are other regularization methods in the literature, such as the Functional Ridge Regression (FRR) (see Benatia et al. (2017) and Hall and Horowitz (2007)) and the nonparametric method for functional linear regression as was suggested by Ferraty and Vieu (2006), Cardot et al. (2003), Cardot et al. (2007),Aguilera et al. (2010), and Park and Qian (2012) among

others. For a review of the different models of functional data analysis and their empirical applications, see Ferraty and Vieu (2006), and Ramsay and Silverman (2007).

The rest of the paper is organized as follows. In Section 1.2, we present the theoretical model and the estimation methods. Section 1.3 is devoted to deriving the rate of convergence of the conditional MSE for both methods and a comparison of the two methods. Section 1.4 is dedicated to the evaluation of both techniques based on simulation results. In Section 1.5 we provide an empirical analysis based on real data in the U.S. stock market and evaluate the results in relation to what we obtained in the theoretical analysis. Section 1.6 concludes. The proofs of the main results are given in Appendix.

1.2 The theoretical model and estimator

This section is devoted to the presentation of the model and the estimation methods of the functional parameter. More precisely, we will first present the model setting and then the estimation methods.

1.2.1 The basics of the functional linear model

Consider a square-integrable functional random variable $X = (X(t))_{t\in S}$ (Which means that $X \in \mathbb{H} = L^2(S)$ and $\int_S E(X^2) < \infty$, the space of square integrable functions mapping from a compact interval S of $\mathbb{R}$ to the set of real numbers) and a real random variable Y that is supposed to be zero mean ($E(X(t)) = 0$ for each $t \in S$ and $E(Y) = 0$).

We assume that the sample $((X_1, Y_1), ..., (X_n, Y_n))$ is a sequence of independent pairs following the same distribution as (X,Y). We consider the functional linear model where $(Y_i)_{i=1...n}$ is the scalar response and $(X_i)_{i=1...n}$ is the functional predictor variable of the regression for each observation i.

We define $\mathbb{H}$ as a Hilbert-space endowed with an inner product $< ., . >$ and a norm $||.||$, which are respectively defined as follows : $< f, g >= \int_S f(t)g(t)dt$ and $||f|| = \left(\int_S f^2\right)^{1/2}$. Let us consider the inner product $< ., . >_n$ such that for each arbitrary element u and v of $\mathbb{R}^n$ we have $||v||_n^2 = v'v/n$ and $< u, v >_n = u'v/n$. Then, the functional linear model is presented as follows :

$$Y_i = \int_S \beta(t)X_i(t)dt + \varepsilon_i \tag{1.1}$$

where $(\beta(t))_{t\in S}$ is a function that belongs to the space $\mathbb{H}$ and $\varepsilon_i; i \in \{1, ..., n\})$ is independent and identically distributed such that $\mathbb{E}(\varepsilon_i|X) = 0$ and $\mathbb{E}(\varepsilon_i^2|X) = \sigma^2 < \infty$ for each $i \in \{1, ..., n\}$).

The covariance operator K of X is an operator mapping from $\mathbb{H}$ to $\mathbb{H}$ defined as

follows :

$$\begin{aligned} K: \quad \mathbb{H} &\to \mathbb{H} \\ f &\to K(f) = \mathbb{E}[(X_i \otimes X_i)(f)] = \mathbb{E}[< X_i, f > X_i] \end{aligned}$$

where $\otimes$ is the tensor product operator defined from $\mathbb{H}$ to $\mathbb{H}$. For all $t \in S$, we have

$$K(f)(t) = \int_S k(s,t) f(s) ds$$

and $k(s,t) = \mathbb{E}[X_i(s)X_i(t)]$ is the kernel of K. Let $(\lambda_j, v_j)_{j\geq 1}$ be the eigensystem of K such that v_j is the eigenfunction associated with the eigenvalue λ_j. A consistent estimator of K is given by the sample covariance operator $\hat{K}$ written as

$$\hat{K} = \frac{1}{n} \sum_{i=1}^{n} X_i \otimes X_i \tag{1.2}$$

Since $(X_i)_{i=1,...n}$ is square-integrable, K is a compact operator and its eigensystem $(\lambda_j, v_j)_{j\geq 0}$ is such that $\lambda_1 \geq \lambda_2 \geq ... > 0$.

Let us define the operators T_n and T_n^* as follows

$$\begin{aligned} T_n: \quad \mathbb{H} &\to \mathbb{R}^n \\ \phi &\to T_n(\phi) = \left[< X_1, \phi >, ..., < X_n, \phi > \right]' \end{aligned}$$

and

$$\begin{aligned} T_n^*: \quad \mathbb{R}^n &\to \mathbb{H} \\ g &\to T_n^*(g) = \frac{1}{n} \sum_{i=1}^{n} X_i g_i. \end{aligned}$$

It is easy to see that T_n^* is the adjoint of T_n, $\hat{K} = T_n^* \circ T_n$ and $(\sqrt{\hat{\lambda}_j}, \hat{v}_j, \hat{u}_j)_{j\geq 1}$ is the singular value decomposition of the operator T_n. We have $T_n(\hat{v}_j) = \left[\int_S X(t)\hat{v}_j(t)dt \right]_{i=1,...,n} = \sqrt{\hat{\lambda}_j}\hat{u}_j$ and $T^*(\hat{u}_j) = X'\hat{u}_j/n = \sqrt{\hat{\lambda}_j}\hat{v}_j$. $(\hat{u}_j)_{j\geq 1}$ are the eigenfunctions associated with the eigenvalues $(\hat{\lambda}_j)_{j\geq 1}$ for the operator $T_n \circ T_n^*$. By the same token, $(\hat{v}_j)_{j\geq 1}$ are the eigenvectors associated with the eigenvalues $(\hat{\lambda}_j)_{j\geq 1}$ for the operator $T_n^* \circ T_n = \hat{K}$. Then, $(\hat{\lambda}_j, \hat{v}_j)_{j\geq 1}$ can be seen as the empirical counterpart of the system $(\lambda_j, v_j)_{j\geq 1}$.

The cross-covariance function between X and Y is defined by :

$$\begin{aligned} C_{xy}: \quad S &\to \mathbb{R} \\ t &\to C_{xy}(t) = \mathbb{E}[X_i(t)Y_i] \end{aligned}$$

Premultiplying each side of Equation (1.1) by $X_i(s)$ and taking the expectation on both sides leads to the following equation :

$$C_{xy}(s) = K\beta(s) \tag{1.3}$$

for each s in S. The main interest is to estimate the functional slope β. If the operator K were invertible, we could estimate β using $\beta(s) = K^{-1}C_{xy}(s)$ for each s in S, but this is not possible because we deal with an ill-posed problem. Indeed, K is a bounded operator mapping from an infinite dimensional space $\mathbb{H}$ to $\mathbb{H}$. This means that it's direct inverse is not continuous and K is not invertible in all the space $\mathbb{H}$ but only on a subset of $\mathbb{H}$. Then, estimating $\hat{\beta}$ by $\hat{K}^{-1}\hat{C}_{xy}$ would lead to an unstable estimator of the functional parameter. Then, the main challenge is to get a stable estimator of the slope function $\beta(t)$. The next section is devoted to present the estimation of the functional slope using the standard FPCA and FPLS methods.

1.2.2 The Functional Principal Component Analysis

The FPCA is a method that consists in regressing the response variable on a finite number of uncorrelated latent components. Those latent components are linear combinations of the predictor variables selected in such a way that the most important information concerning the variation of the predictor variable is captured. This means that the latent components are nothing else than the eigenfunctions of the covariance operator K. Since the $v_j, j = 1, 2, ...,$ form an orthonormal basis in $\mathbb{H}$, the functional parameter β can be written in the space spanned by $v_j, j = 1, 2, ...,$ as follows

$$\beta(t) = \sum_{j=1}^{\infty} \beta_j v_j(t) \tag{1.4}$$

where

$$\beta_j = \int_0^1 \beta(t)v_j(t)dt \tag{1.5}$$

are the Fourier coefficients.

Moreover, according to the fact that the predictor function X is square-integrable, the Karhunen-Loeve expansion for the kernel operator k is given by

$$k(s,t) = \sum_{j=1}^{\infty} \lambda_j v_j(s) v_j(t) \tag{1.6}$$

Let us assume that the number of retained latent components is m. Then, in practice the FPCA method consists in projecting the response variable Y onto the space spanned by the eigenfunctions related to the first m largest eigenvalues, that is, $H_m = sp\{v_1, ..., v_m\}$. Therefore, by estimating the functional slope using such a method, we truncate the infinite sum given by Equation (1.6) to obtain a sum of m terms.

Then, the estimator of $\beta(t)$ is given by

$$\beta_m^{PCA}(t) = \sum_{j=1}^{m} \beta_j v_j(t) = \sum_{j=1}^{m} \frac{1}{\lambda_j} < v_j, C_{xy} > v_j(t). \tag{1.7}$$

Usually, the operator K and its eigensystem $(\lambda_j, v_j)_{j\geq 0}$ are not observed in practice. Let $\hat{\lambda}_j$ and $\hat{v}_j$ be the eigenvalues and eigenfunctions of $\hat{K}$, for $j \geq 1$. $\hat{K}$ can be written as :

$$\hat{K} = \sum_{j=1}^{n} \hat{\lambda}_j < \hat{v}_j, . > \hat{v}_j \tag{1.8}$$

and

$$\hat{k}(s,t) = \sum_{j=1}^{n} \hat{\lambda}_j \hat{v}_j(s) \hat{v}_j(t). \tag{1.9}$$

Assuming that $m \leq n$ is the number of functional components empirically selected, the generalized inverse of $\hat{K}$ by the FPCA method is given by

$$\hat{K}_m^{-1} = \sum_{j=1}^{m} \hat{\lambda}_j^{-1} < ., \hat{v}_j > \hat{v}_j. \tag{1.10}$$

By the same notion, a consistent estimator of C_{xy} is given by $\hat{C}_{xy}$, where

$$\hat{C}_{xy}(t) = \frac{1}{n} \sum_{i=1}^{n} X_i(t) Y_i \tag{1.11}$$

where $t \in S$ is the mean function of the sample of predictor functions and $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ is the mean of the response function. Therefore, $\beta(t)$ can be estimated by

$$\begin{aligned} \hat{\beta}_m^{PCA}(t) &= \hat{K}_m^{-1} \hat{C}_{xy}(t) \\ &= \sum_{j=1}^{m} \frac{1}{\hat{\lambda}_j} < \hat{C}_{xy}, \hat{v}_j > \hat{v}_j(t). \end{aligned} \tag{1.12}$$

1.2.3 The FPCA and factor models

The difference between the FPCA and the factor models is that in the factor models regressors X_i and the response Y_i are assumed to depend on a small number of factors whereas here no factor model is postulated. If the data are generated by a factor model, the FPCA would estimate the underlying factors up to a rotation (see Bai and Ng (2002)). In our model setting, the FPCA is considered as a dimension reduction technique used to estimate the slope function β. In addition, the uniqueness of the eigenfunctions is obtained by the orthonormalization such that

$$< v_j, v_l >= \begin{cases} 1 & if \quad j = l \\ 0 & if \quad j \neq l \end{cases}.$$

1.2.4 The functional spectral cut-off

A close alternative technique to the FPCA method is the functional spectral cut-off (FSC) technique. The idea is almost the same and the only difference is the approach used to select the components. Indeed, the FPCA method retains the first m components, while the FSC technique selects the components that are related to the eigenvalues greater or equal to a certain threshold $\alpha > 0$. Empirically, when using either FPCA or FSC usually leads to the same estimation result and therefore to the same prediction. We will use these results in order to make the comparison between the FPCA and FPLS much easier. Then the slope function $\beta(t)$ can be estimated via FSC method by the following formula

$$\hat{\beta}_{\alpha}^{SC}(t) = \sum_{\hat{\lambda}_j \geq \alpha} \frac{1}{\hat{\lambda}_j} < \hat{C}_{xy}, \hat{v}_j > \hat{v}_j(t). \tag{1.13}$$

Usually, in the empirical economic research, we are much more interested in predicting or forecasting variables. On the other hand, we can notice that when using the FPCA method, the latent components are selected in such a way that they reflect the maximal variance of the predictor variable $X(t)$, but those latent components might not contribute to a good prediction of the response variable. This problem was first documented by Jolliffe (1982). Moreover, in the theoretical framework of the FPCA, it is usually assumed that the Fourier coefficients of the slope function β are efficiently represented by the functional principal components. In other words, it means that, as the eigenvalues of the covariance operator K decrease, the Fourier coefficients of the slope function should decrease at the same rate (see Nie et al. (2018)). This is a strong assumption and it cannot be verified on real data as we do not observe the true slope function. Furthermore, in the context of functional regression, the eigenvalues of the covariance operator K are usually very small except for the first and second eigenvalue (see Kokoszka and Zhang (2010)). Indeed, in the context where we estimate the slope function, it is possible to select latent components that are irrelevant to get a good prediction of the target variable. The partial

least squares method has been introduced as an alternative technique to solve this issue by extracting factors that are closely related to the target variable.

1.2.5 The functional partial least squares regression

The partial least squares is a sequential procedure that consists of projecting the response variable onto a set of uncorrelated latent components. Those latent components are extracted in such a way that they capture not only the most important information concerning the variation observed in the predictor variables X but also present the best predicting performance of the response variable. This means that the selected latent factors capture simultaneously the information on the variation of X and on the relation between X and Y in the linear regression model. This dimension reduction method is widely appreciated in the multivariate linear regression model.

The theoretical properties concerning this method have been developed for the multivariate regression in statistics and chemometrics area by Wold et al. (1984), Naes and Martens (1985), Höskuldsson (1988), Helland (1990); Helland and Almøy (1994), De Jong (1993), Stoica and Söderström (1998), Naes and Martens (1985), Chun and Keleş (2010), Blanchard and Krämer (2010), and Blazere et al. (2014a). This method is quite new in the econometric literature, and we can identify some authors such as Kelly and Pruitt (2015) and Carrasco and Rossi (2016), who focused mainly on the context of prediction using the multivariate linear models. This method was recently adapted to the case of a functional regression model by Preda and Saporta (2005). However, only a few authors, such as Reiss and Ogden (2007), Aguilera et al. (2010) and Delaigle and Hall (2012), have tried to develop some theoretical results when using such method on a functional linear regression model with a scalar response. Delaigle and Hall (2012) are the only ones that have introduced consistency analysis by using another formulation of the PLS that they called the "Alternative PLS" (APLS).

Basically, the literature suggests two main algorithms that can be used in order to extract the latent components for this method : the Nonlinear Iterative Partial Least Squares (NIPALS) introduced by Wold et al. (1984) and the Statistically Inspired Modification of PLS (SIMPLS) developed by De Jong (1993). The two algorithms give the same results as was already proven by Ter Braak and de Jong (1998) for the multivariate linear regression, but the SIMPLS is much faster in terms of computation, more adapted to interpret the latent factors and the best way to go to develop theoretical results (see Ter Braak and de Jong (1998); Chun and Keleş (2010)).

In the following subsections, we will present the SIMPLS algorithm. Then, we will present the APLS formulation of the estimator as proposed by Delaigle and Hall (2012). A new formulation will be proposed by introducing the link between the PLS and the orthogonal polynomials, as suggested by Blazere et al. (2014a), for the case of a simple linear model.

The SIMPLS Algorithm

Let us denote by ω_j the weight function related to the j^{th} FPLS component ψ_j of the regression of Y on X. Let us consider the j^{th} FPLS latent component by

$$\psi_j = \int_S X(t)\omega_j(t)dt. \tag{1.14}$$

Then ω_j is the solution of a maximization problem of the covariance between the response variable Y and the latent factor ψ_j at each step j. We assume also that for the FPLS algorithm, we run m iterations.

The SIMPLS [1] adapted to the functional regression with scalar response is a sequential optimization procedure formulated at each step j as follows :

$$\begin{aligned} w_j = \underset{w\in L^2(S)}{\text{argmax}} \quad & cov^2\left(\int_S X(s)w(s)ds, Y - Y_{i-1}\right) \\ \text{subject to} \quad & ||w_j|| = 1, and \\ & \int_S\int_S w_j(s)k(s,t)w_l(t)dsdt = 0, 1 \leq l \leq j-1 \end{aligned} \tag{1.15}$$

where $Y_{i-1} = \int_S X(s)\beta_{j-1}(s)ds$ and $\beta_{j-1} = \sum_{l=1}^{j-1} q_l\omega_l(t)$. The first constraint of this optimization scheme ensures that the weight functions w_j is normalized to 1, while the second constraint ensures that the FPLS factors ψ_j are mutually orthogonal for each step $j = 1, 2, ...m$. β_{j-1} is the estimated slope function at the step $j-1$, with $\beta_0 = 0$.

Then, after m incrementations, the best predicted value of Y is given by

$$Y_m^{PLS} = q_1\psi_1 + q_2\psi_2 + ... + q_m\psi_m \tag{1.16}$$

Then

$$\begin{aligned} Y_m^{PLS} &= q_1\int_S X(t)\omega_1(t)dt + q_2\int_S X(t)\omega_2(t)dt + ... + q_m\int_S X(t)\omega_m(t)dt \\ &= \int_S X(t)\beta_m^{PLS}(t)dt \end{aligned} \tag{1.17}$$

where

$$\beta_m^{PLS}(t) = \sum_{j=1}^{m} q_j\omega_j(t) \tag{1.18}$$

where $q_1, ..., q_m \in \mathbb{R}$ are the factor loadings of the projection of Y on the FPLS factors.

Moreover, the $(\omega_j(t))_{j\geq 0}$ is an orthogonal basis as was proven by Preda and Saporta

1. For some references concerning this algorithm, we refer the reader to Ter Braak and de Jong (1998), Chun and Keleş (2010), and Delaigle and Hall (2012).

(2005), then the true functional slope β from the regression model can be expressed as

$$\beta(t) = \sum_{j=1}^{\infty} \beta_j \omega_j(t). \tag{1.19}$$

Because w_j and ψ_j are not observed, they are consistently estimated by $\hat{w}_j$ and $\hat{\psi}_j$ using the empirical compensation of the sequential optimization scheme.

The Alternative PLS of Delaigle and Hall

The Theorem 3.1 of Delaigle and Hall (2012) shows that, since the number m is finite, the estimated slope function can be expressed in the space spanned by the functions $K(\beta), K^2(\beta), ..., K^m(\beta)$ in the population model. This is what they called the Alternative PLS (APLS) formula of the estimated slope function.

$$\beta_m^{PLS}(t) = \sum_{j=1}^{m} \gamma_j K^j(\beta)(t) \tag{1.20}$$

where the vector γ is expressed as follows

$$\gamma = (\gamma_1, ..., \gamma_m)^T = H^{-1}(a_1, ..., a_m)^T$$

with $H = (h_{jl})_{1 \leq j,l \leq m}$ a $m \times m$ Hankel matrix such that

$$h_{jl} = \int_S K^{j+1}(\beta) K^l(\beta) \tag{1.21}$$

$$a_j = \int_S K(\beta) K^j(\beta).$$

According to the Theorem 3.2 of Delaigle and Hall (2012), under the condition that X is a square-integrable function and that all of the eigenvalues of the operator K are non-zero, the theoretical functional slope $\beta(t)$ can be expressed as

$$\beta(t) = \sum_{j=1}^{\infty} a_j K^j(\beta)(t). \tag{1.22}$$

FPLS representation based on the orthogonal polynomial theory

This section is devoted to the presentation of the FPLS estimator using the orthogonal polynomial theory. This approach was recently introduced by Blazère et al. (2014b) for a multivariate linear model with a fixed number of regressors in a fixed design context.

Let us define $\mathbb{R}_m[X]$ as the space of polynomials of degree at most m, mapping on the space of real numbers, and $\mathbb{R}_{m,1}[X]$ as the space of polynomials of degree at most m and a constant equal to 1.

Following the results derived by Delaigle and Hall (2012), we notice that the FPLS estimator can be written as a linear combination of the Krylov space, as presented on Equation (1.20). This means that estimating the functional slope using the FPLS approach is equivalent to running a constrained least squares model presented as follows

$$\beta_m^{PLS}(t) = \underset{\beta \in \mathscr{K}_m(K\beta, K)}{\operatorname{argmin}} \mathbb{E}\left[Y - \int_S X(t)\beta(t)dt\right]^2 \tag{1.23}$$

where β is constrained to belong to the Krylov basis $\mathscr{K}_m(K\beta, K)$, and

$$\mathscr{K}_m(K\beta, K) = \{K\beta, K \circ K\beta, ..., K^{m-1} \circ K\beta\}$$

and m is the number of selected components. It is worth noting that the constraint depends on the response variable, this means that we deal with stochastic constraints and the method is nonlinear in Y.

Theorem 1 : Corollary 7.4 of Engl et al. (1996)
The minimization problem subject to the Krylov space constraints in Equation (1.23) is identical to a minimization problem where one should find the optimal polynomial P_m in the space $\mathbb{R}_{m-1}[X]$, such that

$$P_m = \underset{P \in \mathbb{R}_{m,1}[X]}{\operatorname{argmin}} \mathbb{E}\left[Y - < X, P(K)K(\beta) > \right]^2$$

and $\beta_m^{PLS} = P_m(K)C_{xy}$.

Indeed, as β_m^{PLS} is the solution of (23), then $\beta_m^{PLS} \in \mathscr{K}_m(K, K(\beta))$ and $\beta_m^{PLS} = P_m(K)K(\beta) = P_m(K)C_{xy}$. Therefore,

$$\begin{aligned} P_m &= \underset{P \in \mathbb{R}_{m-1}[X]}{\operatorname{argmin}} \mathbb{E}\left[Y - \int_S X(t)P(K)K(\beta)(t)dt\right]^2 \\ &= \underset{P \in \mathbb{R}_{m-1}[X]}{\operatorname{argmin}} \mathbb{E}\left[Y - < X, P(K)K(\beta) > \right]^2 \\ &= \underset{P \in \mathbb{R}_{m-1}[X]}{\operatorname{argmin}} \mathbb{E}\left[f(x) - < X, P(K)K(\beta) > \right]^2 \end{aligned}$$

where $f(x) = \int \beta(s)x(s)ds$.

Let us define the operators T_x, T_x^*, W_x, respectively by

$$\begin{aligned} T_x : \ \mathbb{H} &\to \mathbb{R} \\ \phi &\to T_x(\phi) =< X, \phi > \end{aligned}$$

$$\begin{aligned} T_x^* : \quad \mathbb{R} &\to \mathbb{H} \\ Z &\to T_x^*(Z) = \mathbb{E}[XZ] \end{aligned}$$

$$\begin{aligned} W_x : \quad \mathbb{R} &\to \mathbb{R} \\ Z &\to W_x(Z) = \int_0^1 X(t)\mathbb{E}[X(t)Z]dt =< X(.), \mathbb{E}[X(.)Z] >_{\mathbb{H}} \end{aligned}$$

where Z is an arbitrary real random variable and, ϕ is an arbitrary random function. Moreover, T_x^* is the adjoint of T_x. W_x is called the Escoufier operator. It is straightforward to see that $K = T_x^* \circ T_x$, $W_x = T_x \circ T_x^*$ and $\int_0^1 X(t)\beta(t)dt = T_x\beta$. Note that T_x and T_x^* are the population versions of T_n and T_n^*. Using $X(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} u_j v_j$, we have

$$W_x(Z) = \sum_{j=1}^{n} \lambda_j \mathbb{E}[u_j Z] u_j$$

where $u_j = \frac{1}{\sqrt{\lambda_j}} \int_0^1 X(t)v_j(t)dt$, $\mathbb{E}[u_j u_k] = 1$ if $j = k$ and 0 otherwise. Then, the minimization problem (23) can be formalized in a different way as follows :

$$\begin{aligned} \mathbb{E}\Big[f(x) - < X; P_m(K)K(\beta) > \Big]^2 &= \mathbb{E}\Big[T_x(\beta) - T_x(P_m(K)K(\beta)) \Big]^2 \\ &= \mathbb{E}\Big[T_x\beta - T_x P_m T_x^* T_x) T_x^* T_x \beta \Big]^2 \\ &= \mathbb{E}\Big[T_x\beta - T_x T_x^* P_m(T_x T_x^*) T_x \beta \Big]^2 \\ &= \mathbb{E}\Big[T_x\beta - W_x P_m(W_x) T_x \beta \Big]^2 \\ &= \mathbb{E}\Big[\{1 - W_x P_m(W_x)\} T_x \beta \Big]^2 \\ &= \mathbb{E}\Big[Q_m(W_x) f(x) \Big]^2 \\ &= \min_{Q \in \mathbb{R}_{m,1}[X]} \mathbb{E}\Big[Q(W_x)(f(x)) \Big]^2 \end{aligned}$$

where $Q_m(x) = 1 - xP_m(x)$ is called the "residual polynomial". Indeed, it is a polynomial of degree at most m and with a constant equal to 1.

Let us define now the following measure on the space of real numbers :

$$d\mu(\lambda) = \sum_{j=1}^{\infty} \lambda_j \Big[\mathbb{E}\Big(Y u_j \Big) \Big]^2 \delta_{\lambda_j}$$

where δ_{λ_j} is the dirac measure that takes the value 1 if we are on the mass point related to the eigenvalue λ_j and 0 elsewhere. For each arbitrary polynomial function F and G that belongs to $\mathbb{R}_m[X]$, we introduce the inner product on the space $\mathbb{R}_m[X]$ associated with the measure μ, that is

$$\begin{aligned}[F,G] &= \int_0^{\lambda_1} F(t)G(t)d\mu(t) \\ &= \sum_{j=1}^{\infty} \lambda_j F(\lambda_j)G(\lambda_j)\left[\mathbb{E}\left(Yu_j\right)\right]^2.\end{aligned}$$

And the semi-norm associated with this inner product is given by

$$\begin{aligned}||F||^2 &= \int_0^{\lambda_1} F(t)^2 d\mu(t) \\ &= \sum_{j=1}^{\infty} \lambda_j F(\lambda_j)^2\left[\mathbb{E}\left(Yu_j\right)\right]^2.\end{aligned}$$

According to the definition of Q_m, we can directly observe that Q_m is the unique polynomial of degree at most m with a constant 1, which is orthogonal to the space $T_x(\mathbb{R}_{m-1}[X])$. Since, $Q_m \in \mathbb{R}_{m,1}[X] \subset \mathbb{R}_m[X]$, the polynomials $(Q_h)_{h\geq 1}$ are orthogonals with respect to the scalar product $[.,.]$. Also, Q_m is interpreted as an orthogonal projection of the origin onto the space $\mathbb{R}_{m,1}[X]$. This shows the unicity of the polynomial Q_m and therefore of P_m and β_m^{PLS}.

The detailed mathematical features of the residual polynomials on the PLS approach are developed by Engl et al. (1996) and for the multivariate linear regression in the paper of Blazere et al. (2014a). Moreover, the Partial Least Squares is very close to the Conjugate Gradient (CG) methods for the normal equation $K(\beta) = C_{xy}$ with a starting point $\beta_0 = 0$ (see Blanchard and Krämer (2010), Bro and Elden (2009)). Phatak and de Hoog (2002) have used the CG approach to derive some regularization properties of the PLS estimator. For more details on the mathematical results concerning the residual polynomials and the CG method in the context of ill-posed problems, see Engl et al. (1996), Plato (1998), Blanchard and Krämer (2016), and Hanke (2017). We will use some of these materials for the proofs of the results.

Proposition 1.

If m is the number of iterations of the FPLS, then the population version of the FPLS estimator related to the functional linear regression is presented as follows :

$$\beta_m^{PLS}(t) = \sum_{j=1}^{\infty} \frac{\overline{Q}_{m,j}}{\sqrt{\lambda_j}}\mathbb{E}[Yu_j]v_j(t) = \sum_{j=1}^{\infty} \overline{Q}_{m,j} < \beta, v_j > v_j(t)$$

where

$$\overline{Q}_{m,j} = 1 - Q_m(\lambda_j) = \sum_{(j_1,...,j_m)\in I_m^+} w_{j_1,...,j_m}\left[1 - \prod_{l=1}^{m}(1 - \frac{\lambda_j}{\lambda_{j_l}})\right]$$

$$w_{j_1,...,j_m} = \frac{g_{j_1}^2, ..., g_{j_m}^2\lambda_{j_1}^2...\lambda_{j_m}^2 V(\lambda_{j_1}...\lambda_{j_m})^2}{\sum_{(j_1,...,j_m)\in I_m^+} g_{j_1}^2, ..., g_{j_m}^2\lambda_{j_1}^2...\lambda_{j_m}^2 V(\lambda_{j_1}...\lambda_{j_m})^2}$$

$$g_j^2 = \lambda_j^2 < \beta, v_j >^2$$

$$I_m^+ = \{(j_1, ..., j_m) : j_1 > ... > j_m \geq 1\},$$

$$V(\lambda_{j_1}...\lambda_{j_m}) = \begin{vmatrix} 1 & \lambda_{j_1} & ... & \lambda_{j_1}^{m-1} \\ 1 & \lambda_{j_2} & ... & \lambda_{j_2}^{m-1} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & \lambda_{j_m} & ... & \lambda_{j_m}^{m-1} \end{vmatrix} = \prod_{1\leq i<j\leq m}(\lambda_j - \lambda_i)$$

and $0 \leq w_{j_1...j_m} \leq 1$ *is a stochastic weight which satisfy* $\sum_{I^+} w_{j_1...j_m} = 1$.

We can observe that the FPLS estimator depends on the eigensystem $(\lambda_j, v_j)_{j\geq 1}$ of the covariance operator K. It is worth noting that the main difference with the results of Blazere et al. (2014a) is that β_m^{PLS} is expressed as an infinite sum for the functional linear regression model, while in the multivariate linear regression β_m^{PLS} is a finite sum. Additionally, the authors consider a fixed design model.

If we recall the expression of the APLS estimator given by Delaigle and Hall (2012) and using the spectral decomposition of the covariance operator K, we have the following expression

$$\beta_m^{PLS}(t) = \sum_{j=1}^{m}\mu_j K^j(\beta)(t) = \sum_{j=1}^{m}\mu_j \sum_{l=1}^{\infty} < \beta, v_l >_{\mathbb{H}} \lambda_l^j v_l(t).$$

Then, we obtain the following equivalence

$$\beta_m^{PLS}(t) = \sum_{l=1}^{\infty}\left(\sum_{j=1}^{m}\mu_j\lambda_l^j\right) < \beta, v_l > v_l(t) = \sum_{l=1}^{\infty}\overline{Q}_{m,l} < \beta, v_l > v_l(t).$$

since $\overline{Q}_{m,l} = 1 - Q_m(\lambda_l) = \sum_{j=1}^{m}\mu_j\lambda_l^j$ where μ_j are derived from the minimization problem.

The values of $(\lambda_j, v_j)_{j\geq 1}$ are unknown and are consistently estimated by $(\hat{\lambda}_j, \hat{v}_j)_{j\geq 1}$, respectively. Similarly the empirical counterpart of the measure $d\mu$ is given by

$$d\hat{\mu} = \sum_{j=1}^{n} \hat{\lambda}_j < Y, \hat{u}_j >_n^2 \delta_{\hat{\lambda}_j}.$$

Then, the feasible estimator of the slope function is given by

$$\hat{\beta}_m^{PLS}(t) = \sum_{j=1}^{n} \frac{\hat{\overline{Q}}_{m,j}}{\sqrt{\hat{\lambda}_j}} < Y, \hat{u}_j >_n \hat{v}_j(t),$$

where $\hat{\overline{Q}}_{m,j}$ is the empirical counterpart of $\overline{Q}_{m,j}$.

$$\hat{\overline{Q}}_{m,j} = 1 - \hat{Q}_m(\hat{\lambda}_j) = \sum_{(j_1,...,j_m)\in I_m^+} \hat{w}_{j_1,...,j_m} \left[1 - \prod_{l=1}^{m}(1 - \frac{\hat{\lambda}_j}{\hat{\lambda}_{j_l}})\right]$$

$$\hat{w}_{j_1,...,j_m} = \frac{\hat{g}_{j_1}^2, ..., \hat{g}_{j_m}^2 \hat{\lambda}_{j_1}^2 ... \hat{\lambda}_{j_m}^2 V(\hat{\lambda}_{j_1} ... \hat{\lambda}_{j_m})^2}{\sum_{(j_1,...,j_m)\in I_m^+} \hat{g}_{j_1}^2, ..., \hat{g}_{j_m}^2 \hat{\lambda}_{j_1}^2 ... \hat{\lambda}_{j_m}^2 V(\hat{\lambda}_{j_1} ... \hat{\lambda}_{j_m})^2}$$

$$\hat{g}_j =< Y, \hat{u}_j >_n$$

$$\hat{I}_m^+ = \{(j_1, ..., j_m) : n \geq j_1 > ... > j_m \geq 1\},$$

$$V(\hat{\lambda}_{j_1} ... \hat{\lambda}_{j_m}) = \begin{vmatrix} 1 & \hat{\lambda}_{j_1} & ... & \hat{\lambda}_{j_1}^{m-1} \\ 1 & \hat{\lambda}_{j_2} & ... & \hat{\lambda}_{j_2}^{m-1} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & \hat{\lambda}_{j_m} & ... & \hat{\lambda}_{j_m}^{m-1} \end{vmatrix} = \prod_{1\leq i<j\leq m} (\hat{\lambda}_j - \hat{\lambda}_i)$$

and $0 \leq \hat{w}_{j_1...j_m} \leq 1$ is a stochastic weight that satisfies $\sum_{I^+} \hat{w}_{j_1...j_m} = 1$.

These new expressions of the FPLS slope estimator in population and sample are useful in order to derive some consistency properties of the estimator and to theoretically compare the FPLS and the FPCA. Those results are presented in the next section.

Remark 1.

An alternative setting of the estimation problem via FPLS in Equation (1.23) is presented as follows

$$\hat{\beta}_m^{PLS} = \underset{\beta\in\mathcal{K}_m(T_n^*\underline{Y}, T_n^*T_n)}{argmin} \{\|T_n^*\underline{Y} - T_n^*T_n\beta\|\}$$

where $\mathcal{K}_m\left(\hat{C}_{xy}, \hat{K}\right) = span\{\hat{C}_{xy}, \hat{K}\hat{C}_{xy}, \hat{K}^2\hat{C}_{xy}, ..., \hat{K}^{m-1}\hat{C}_{xy}\}$. The previous results are also valid for this setting. The only difference is the norm and the scalar products used

for the orthogonal polynomials. Indeed, we use the conjugate scalar product $[.,.]_{W_n}$ where the related matrix is $W_n = T_n T_n^*$. For the theoretical results we will use this alternative setting. In other words,

The standard PLS is obtained via the following regression

$$\min_{\beta} \|\underline{Y} - T_n\beta\|_n^2$$

s.t. to $\beta \in \mathcal{K}\left(\hat{C}_{xy}, \hat{K}\right) = span\left\{\hat{C}_{xy}, \hat{K}\hat{C}_{xy}, ..., \hat{K}^{m-1}\hat{C}_{xy}\right\}$.

The estimator is

$$\hat{\beta} = \sum_{j=0}^{m-1} \widehat{\gamma}_j \hat{K}^j \hat{C}_{xy}$$

where $\hat{\gamma} = H^{-1}a$ with $(H)_{ij} = \underline{Y}'(W_n)^{i+j}\underline{Y}$ and $a_i = \underline{Y}'(W_n)^i\underline{Y}$.

Our modified PLS is given by

$$\min_{\beta} \|T_n^*\underline{Y} - T_n^*T_n\beta\|^2$$

s.t. to $\beta \in \mathcal{K}\left(\hat{C}_{xy}, \hat{K}\right) = span\left\{\hat{C}_{xy}, \hat{K}\hat{C}_{xy}, ..., \hat{K}^{m-1}\hat{C}_{xy}\right\}$.

The estimator is

$$\hat{\beta} = \sum_{j=0}^{m-1} \widehat{\gamma}_j \hat{K}^j \hat{C}_{xy}$$

where $\hat{\gamma} = H^{-1}a$ with $(H)_{ij} = \underline{Y}'(W_n)^{i+j+1}\underline{Y}$ and $a_i = \underline{Y}'(W_n)^{i+1}\underline{Y}$.

1.3 Rate of convergence of the MSE

This section is dedicated to the analysis of the rate of convergence of the conditional MSE for the FPCA and FPLS approach, respectively. For such purpose, we need to set up some assumptions. At the first stage, we present the rate of convergence of the conditional MSE for the FPCA. The second stage derives the rate of convergence for the FPLS approach. The third part of this section is related to the comparison of the two techniques in terms of their respective conditional MSE.

Let us first specify the following assumption that would be useful for the consistency results.

Assumptions :

A1. (X_i, Y_i) are i.i.d with the same distribution law as (X,Y) with $E[||X||^2] < +\infty$.

A2. $\int_S \beta^2(t)dt < +\infty$, $E[\varepsilon_i|X] = 0$, $E[\varepsilon_i^2|X] = \sigma^2$, $E[\varepsilon_i^4|X] < +\infty$ and $E[||X||^4] < +\infty$.

A3. The eigenvalues of the covariance operator K and the estimated one $\hat{K}$ are distinct, i.e,

$\lambda_1 > \lambda_2 > ... > 0$ and $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_n$.

A4. We assume that for some $\mu \geq 0$, β satisfies

$$\sum_{j=1}^{\infty} \frac{< \beta, v_j >^2}{\lambda_j^{\mu}} < \infty.$$

Comments on the assumptions :

A1 is introduced to facilitate the consistency results of the sample covariance estimator $\hat{K}$. Moreover, $E[||X||^2] < +\infty$ guarantees that K is a nuclear operator, i.e $\sum_{j\geq 1} \lambda_j < +\infty$.

In **A2**, the condition $E[||X||^4] < +\infty$ is satisfied if X is a Gaussian process. For instance, A1 and A2 are sufficient conditions to ensure that $||\hat{K} - K||^2_{HS} = O_p(\frac{1}{n})$, see Proposition 5 of Dauxois et al. (1982).The condition $E[\varepsilon_i^4|X] < +\infty$ simplifies the proof of the consistency for the FPLS estimator. This assumption is not needed for the consistency of the FPCA estimator.

A3 implies that the eigenvalues λ_j and the eigenvectors v_j are consistently estimated by $\hat{\lambda}_j$ and $\hat{v}_j$ respectively, for each $j \geq 1$. This condition guarantees that the null space of the covariance operator K, $\mathcal{N}(K)$ is such that $\mathcal{N}(K) = 0$.

A4 is a source condition as discussed by Carrasco et al. (2007) in the context of ill-posed problems. In fact, this condition represents a smoothness class of the functional parameter β based on its Fourier coefficients. As K is a compact operator, the sequence of eigenvalues $(\lambda_j)_{j\geq 1}$ converges to zero as j goes to infinity. In other terms, this condition imposes that the Fourier coefficients $< \beta, v_j >$ decline faster than the eigenvalues. This condition also means that there exists a squared integrable function w such that $\beta = K^{\mu/2}(w)$ and $||w|| < +\infty$. In the inverse problem literature, this parameter characterizes the severity of the ill-posed problem. As μ becomes larger, the ill-posed problem becomes more severe, i.e the eigenvalues λ_j decay more faster (see proposition 3.13 of Engl et al. (1996)).

1.3.1 Rate of convergence of the MSE for the FPCA

The consistency results of the estimator $\hat{\beta}_m^{PCA}$ has been widely analysed in the literature of functional linear regression model with a scalar response. The main papers on this subject are the one of Cardot et al. (1999), Cardot et al. (2003), Hall and Hosseini-Nasab (2006), Hall and Horowitz (2007), and Cai et al. (2006). The mean-square prediction error of such a model in the case of finite sample is analyzed by Hall and Hosseini-Nasab (2006). They also derived the main conditions under which the estimator $\hat{\beta}_m^{PCA}$ converges to the theoretical value β. A particular condition considered by Hall and Horowitz (2007) is that the predictor variables and the slope function are enough smooth. They consider the following configuration : $\beta_j \leq Dj^{-\zeta}$ and $\lambda_j - \lambda_{j+1} \geq Dj^{-\delta-1}$, for $j \geq 1$ with $\zeta > \frac{1}{2}\delta + 1$, D a positive constant and $\delta > 1$. An interpretation of this result

is that the eigenvalues of K decrease at a polynomial rate. They proved that the rate of convergence of the MSE and MSPE depend essentially on the decreasing rate of the space between adjacent eigenvalues $(\lambda_j)_{j\geq 1}$ of the covariance operator K (see Cardot et al. (1999), Cai et al. (2006)). Also, they usually assume that the eigenvalues are distinct, well spaced, and decrease very slowly to zero as j increases.

Actually, we do not impose conditions on the decay rate of successive eigenvalues but introduced a source condition A4, which would make the results slightly different and more general compared to the ones obtained by the other authors when deriving the rate of convergence of the FPCA. Thus, we derive the convergence rate of

$$||\hat{\beta}_m^{PCA} - \beta||^2.$$

We have

$$\hat{\beta}_m^{PCA} - \beta = \left(\hat{\beta}_m^{PCA} - \beta_m^{PCA}\right) + \left(\beta_m^{PCA} - \beta\right)$$

where $\left(\beta_m^{PCA} - \beta\right)$ is the regularization bias and $\left(\hat{\beta}_m^{PCA} - \beta_m^{PCA}\right)$ is the estimation error term.

Therefore, we have the following inequality.

$$||\hat{\beta}_m^{PCA} - \beta|| \leq ||\hat{\beta}_m^{PCA} - \beta_m^{PCA}|| + ||\beta_m^{PCA} - \beta||.$$

Similarly, for the FSC method, we have

$$||\hat{\beta}_\alpha^{SC} - \beta|| \leq ||\hat{\beta}_\alpha^{SC} - \beta_\alpha^{SC}|| + ||\beta_\alpha^{SC} - \beta||$$

Then, we have the following results.

Theorem 2.
Under the model conditions, if the assumptions A1 - A4 hold, the rate of convergence of the conditional MSE is given by :

$$||\hat{\beta}_m^{PCA} - \beta||^2 = O_p\left(\lambda_{m+1}^{\mu}\right) + O_p\left(\frac{m}{n\lambda_m}\right) \tag{1.24}$$

$$||\hat{\beta}_\alpha^{SC} - \beta||^2 = O_p\left(\alpha^{\mu}\right) + O_p\left(\frac{1}{n\alpha^2}\right). \tag{1.25}$$

Remark 2.

— The rate of convergence with FPCA method depends on the number of selected components m and the configuration of the eigenvalues of K, while for the FSC method it depends on the threshold α.

— The convergence rate for the conditional MSPE is the same for each method.
— It should be mentioned that the consistency results with FPCA method are obtained without using an assumption on the decreasing rate of the eigenvalues in contrast to the results presented by Cardot et al. (1999). Therefore, our results are more general and are different from the one obtained by Hall and Horowitz (2007), Cai et al. (2006), and Hall and Hosseini-Nasab (2006), since we have more general assumptions and we consider a random design context.
— In contrast to Hall and Horowitz (2007), who considered a fixed design model, Crambes et al. (2013) considered a random design model and derived their results under the convexity assumptions of the eigenvalues of the covariance operator K, which englobe a larger class of eigenvalues configurations than the one proposed by Hall and Horowitz (2007) regarding the smoothness of the slope function combined with the decreasing rate of the eigenvalues of K, but their assumptions are still restrictive compared to the one considered in this paper.
— Concerning the FSC method, the optimal rate of convergence can be obtained by equating the bias and estimation error terms. Then by taking $\alpha \sim n^{-1/(\mu+2)}$, the optimal rate of the conditional MSE is given by

$$||\hat{\beta}_{\alpha}^{SC} - \beta||^2 = O_p\left(n^{-\mu/(\mu+2)}\right). \tag{1.26}$$

This result means that if we do not have any information about the configuration of the eigenvalues of the covariance operator K the convergence rate of the MSE is $n^{-\mu/(\mu+2)}$.

1.3.2 Rate of Convergence of the MSE for the FPLS

The rate of convergence of the conditional MSPE for the FPLS method has been derived recently by Blazere et al. (2014a) and Carrasco and Rossi (2016) for a multivariate linear regression. Blazere et al. (2014a) considered a fixed design model with a fixed number of covariates and some multicolinearity, while Carrasco and Rossi (2016) assumed that the number of predictor variables may increase with the number of observations. Delaigle and Hall (2012) derived the consistency results of the MSPE for the case of a functional linear regression with a scalar response by using an alternative formulation of the FPLS.

This section is focused in deriving the rate of convergence of the conditional MSE for the FPLS method. It should be mentioned that the FPLS is a nonlinear method since the latent factors at each step depend on the response variable, which therefore makes the derivation of the rate of convergence more difficult. In this section, we study the rate of convergence of

$$\left|\left|\hat{\beta}_{m}^{PLS} - \beta\right|\right|^2.$$

We can decompose the prediction error $\left(\hat{\beta}_m^{PLS} - \beta\right)$ as a sum of two terms, those are the estimation error for the first part and the regularization bias for the second part. Then, we have

$$\hat{\beta}_m^{PLS} - \beta = \left(\hat{\beta}_m^{PLS} - \beta_m^{PLS}\right) + \left(\beta_m^{PLS} - \beta\right).$$

Then,

$$\left|\left|\beta_m^{PLS} - \beta\right|\right| \leq \left|\left|\hat{\beta}_m^{PLS} - \beta_m^{PLS}\right|\right| + \left|\left|\beta_m^{PLS} - \beta\right|\right|.$$

The results are then mentioned in the next proposition and theorem.

Proposition 2.

Under the model condition, if the assumptions A1 - A4 hold, then for all $m \leq n$ we obtain the following results :

$$\left|\left|\beta_m^{PLS} - \beta\right|\right|^2 \leq \left|\left|\beta_m^{PCA} - \beta\right|\right|^2 \tag{1.27}$$

and

$$\left|\left|\beta_m^{PLS} - \beta\right|\right|^2 = O_p\left(\lambda_{m+1}^{\mu}\right). \tag{1.28}$$

Proposition 2 shows that the squared bias with FPLS is smaller than the one obtained when using FPCA and the upper bound rate depends of the eigenvalue λ_{m+1} of the covariance operator K.

Theorem 3.

Under the model condition, if the assumptions A1 - A4 hold, and given the stopping rule

$$||T_n^*(\underline{Y} - T_n\hat{\beta}_m^{PLS})|| \leq \frac{\tau\sqrt{A_0}}{\sqrt{n}} < ||T_n^*(\underline{Y} - T_n\hat{\beta}_{m-1}^{PLS})|| \tag{1.29}$$

where $\tau > 1$ is a given constant and $A_0 = \sigma^2 \int_S \mathbb{E}[X_i^2(s)]ds$, we obtain for $m \leq n$

$$||\hat{\beta}_m^{PLS} - \beta||^2 = O_p\left(\lambda_{m+1}^{\mu}\right) + O_p\left(\frac{m^2}{n\theta_m^2}\right). \tag{1.30}$$

Moreover, the optimal rate of convergence of the MSE is

$$||\hat{\beta}_m^{PLS} - \beta||^2 = O_p\left(n^{-\mu/(\mu+2)}\right). \tag{1.31}$$

Remark 3.

— In Equation (1.30), the first term is the convergence rate of the squared bias while the second one is the convergence rate of the variance. θ_m is the smallest root of the residual polynomial $\overline{Q}_{m,j}$. θ_m is also known as the smallest eigenvalue of the Hankel matrix H (see Equation (1.21)) defined by Delaigle and Hall (2012) or the Ritz value as defined by Lingjaerde and Christophersen (2000). Therefore, if m diverges in such a way that $\left(\frac{m^2}{\theta_m^2 n}\right) \to 0$, the conditional MSE converges to zero as n goes to infinity. It has been proven by Lingjaerde and Christophersen (2000) that, for j fixed, as m increases, $\overline{Q}_{m,j}^2$ converges to 1 and θ_m goes to 0 quickly; therefore, in such a situation, the estimation error can increase very rapidly as n goes to infinity. So, in order to have the variance go to zero, we should control the number m to increase at a very low rate compared to n. Furthermore, θ_m is smaller than the eigenvalue λ_m of the covariance operator K (see Lingjaerde and Christophersen (2000)).

— The convergence rate for the conditional MSPE is the same as that of MSE. The MSPE is defined as follows

$$MSPE = \mathbb{E}\left[\left(\int_S X(s)\hat{\beta}(s)ds - \int_S X(s)\beta(s)ds\right)^2\right]$$

Furthermore, to obtain the convergence rate of the conditional MSPE we condition on $X = x$ such that $||x||^2 = O_p(1)$.

$$\mathbb{E}\left[\left(\int_S X(t)(\hat{\beta}_m^{PLS} - \beta)(t)dt\right)^2 |X\right] \leq ||x||^2\mathbb{E}\left[\left|\left|\hat{\beta}_m^{PLS} - \beta\right|\right|^2 |X\right].$$

This result holds also for the FPCA and FSC methods.

Theorem 4 : the number of selected components.
Under the model conditions, if assumptions A1 - A4 and the stopping criterion from Equation (1.29) holds, the optimal number of selected components m_{pls} is such that $m_{pls} \leq cn^{1/(2\mu+4)}$ and this estimate is sharp in the sense that it cannot be replaced by a smaller one, for some positive constant $c > 0$.

This result means that if we do not have any information about the configuration of the eigenvalues of K, the best possible uniform bound for the optimal number of selected components m_{pls} is $n^{1/(2\mu+4)}$. By considering a specific decay rate on the eigenvalues of K, we are able to derive a better bound for both methods as described in the following corollary.

Corollary 1.
Assume that the model conditions, assumptions A1 - A4 and the stopping criterion hold.

— *If* $\lambda_j = O(j^{-2\gamma})$ *as* $j \to +\infty$, *with* $\gamma > 0$, *then* $\begin{cases} m_{pls} \leq cn^{1/(2\mu+4)(\gamma+1)} \\ m_{pca} \leq cn^{1/(2\mu(\gamma+1)+1)} \end{cases}$

— *If* $\lambda_j = O(d^j)$ *as* $j \to +\infty$ *with* $0 < d < 1$, *then* $\begin{cases} m_{pls} \leq c(1 + log^+(n)) \\ m_{pca} \leq c(1 + log^+(n)) \end{cases}$,

where $log^+(t) = log(t)$ *if* $t \geq 1$ *and* $log^+(t) = 0$ *otherwise and* $c > 0$ *is some positive constant.*

Corollary 1 shows that when the eigenvalues decrease at a polynomial rate, one can expect the bound of the optimal number of components obtained with FPLS to be smaller than the one obtained via FPCA, while when the eigenvalues decrease at an exponential rate, both methods tend to display the same bound of number of retained components. Additionally, the number of retained components is usually very small when the eigenvalues decrease at an exponential rate compared to the case where their decreasing rate is polynomial. Also, it is important to mention that for the FPLS method, as the covariance operator of the predictor function K is nuclear, the bound of the number of selected components is improved compared to the one obtained in Theorem 4.

The next proposition shows that the convergence rate obtained in Theorem 2 and 3 are sharp.

Proposition 3.
Let $\mathcal{B}$ be the set of all measurable functions of $(X_i, Y_i), i = 1, 2, ..., n$. There exists a positive constant c such that

$$\lim_{n\to\infty} \inf_{\bar{\beta}\in\mathcal{B}} \sup_{\beta\in\mathcal{R}\left(K^{\mu/2}\right)} n^{\frac{\mu}{\mu+2}} \mathbb{E} \left\|\bar{\beta} - \beta\right\|^2 \geq c.$$

where $\mathcal{R}(K^{\mu/2}) = \left\{ g \in \mathbb{H} : g = K^{\mu/2}(\phi), ||\phi|| < +\infty \right\}$. *This means that the rate of convergence of the MSE obtained from both methods is sharp. In other words, it is the best possible uniform rate in the set of estimators of* β *on the considered model.*

1.3.3 Comparison of the FPCA and FPLS

This section is devoted to the comparison of the FPCA and FPLS approach. Such analyses are operated by using the conditional MSPE. We first compare their regularization bias and at the second stage, we compare their estimation error.

Comparison of the regularization bias of FPCA and FPLS

Following Proposition 2, we can observe that given the same number of components m, the squared bias obtained with the FPLS method is in general lower than the one obtained by the FPCA method. Furthermore, both methods display the same upper bound rate for the squared bias, that is $O_p\left(\lambda_{m+1}^{\mu}\right)$.

Moreover, the regularization bias of FPLS does not necessarily decrease strictly as m increases, while it is the case for FPCA . This is due to the fact that $\overline{Q}_{m,j}$ oscillates around the value 1 (while for the FPCA, $\overline{Q}_{m,j} = 1$ for $1 \leq j \leq m$ and 0 otherwise). To observe this feature, we can consider a situation where we have $\lambda_j = \lambda_1$. Then, we obtain

$$\begin{cases} \prod_{l=1}^{m} \left(1 - \frac{\lambda_j}{\lambda_{j_l}}\right) \leq 0 & \text{if } m \text{ is odd} \\ \prod_{l=1}^{m} \left(1 - \frac{\lambda_j}{\lambda_{j_l}}\right) \geq 0 & \text{if } m \text{ is even} \end{cases}$$

where $j_l \in I_m^+ = \{(j_1, ..., j_m) : j_1 > ... > j_m \geq 1\}$. Therefore, we obtain the following result

$$\begin{cases} \overline{Q}_{m,j} \geq 1 & \text{if } m \text{ is odd} \\ \overline{Q}_{m,j} \leq 1 & \text{if } m \text{ is even} \end{cases}$$

Moreover, for m fixed, $\overline{Q}_{m,j}$ goes to zero as j increases and $j > m$ and for j fixed, $\overline{Q}_{m,j}$ goes to 1 as m diverges, see Theorem 3 of Lingjaerde and Christophersen (2000), section 4.2 of Blazere et al. (2014a), and section 2.1 of Carrasco and Rossi (2016). But this pattern of the filter factor induces the fact that the bias of the FPLS estimator is not more than the one obtained by the FPCA.

Comparison of the estimation error of FPCA and FPLS

Concerning the comparison of the estimation error of these methods, we already know that under the assumptions $A1 - A4$, the upper bound rate of the estimation error is

given by the following results

$$||\hat{\beta}_m^{PCA} - \beta_m^{PCA}||^2 = O_p\left(\frac{m}{n\lambda_m}\right)$$

and

$$||\hat{\beta}_m^{PLS} - \beta_m^{PLS}||^2 = O_p\left(\frac{m^2}{n\theta_m^2}\right).$$

Under the assumptions A1- A4, if m diverges at a slower rate than n such that $\left(\frac{m^2}{\theta_m^2 n}\right) \to 0$, then the estimation error converges in probability to zero. Another feature is that the convergence rate of the estimation error for the FPCA method depends on the decreasing rate of the smallest extracted eigenvalues λ_m of the operator K, while the one obtained with the FPLS method depends on the decreasing rate of the smallest root of the residual polynomial $\overline{Q}_{m,j}$, that is θ_m. Moreover, since $\theta_m < \lambda_m$ (see Lingjaerde and Christophersen (2000)), the upper bound of the estimation error with FPCA is asymptotically less than that of FPLS.

Furthermore, we can easily notice from the simulations in section 4 that the estimation error obtained from FPLS can diverge very quickly and may explode as m increases. This can present an overfitting issue if we intend to select a large number of components. In addition, for the same number of retained components, the estimation error with FPLS method tend to be larger than the one obtained with FPCA. A potential explanation is that the FPLS method is an iterative method whereby at each step, we should handle the residuals of predictor and the response variables obtained from the estimations of the previous step.

For both methods, as the number m increases, the bias rate decreases and the estimation error's rate increases. Thus, it is important to make a tradeoff when choosing the number for latent components m in order to obtain the minimal MSE. We use a data driven method mentioned in the next section in order to choose such number. The empirical simulation results will help us to highlight more results about it.

Optimal convergence rates

Concerning the optimal convergence rate for both methods, when there is no information about the eigenvalues.

$$||\hat{\beta}_\alpha^{SC} - \beta||^2 = O_p\left(n^{-\mu/(\mu+2)}\right) \tag{1.32}$$

$$||\hat{\beta}_m^{PLS} - \beta||^2 = O_p\left(n^{-\mu/(\mu+2)}\right) \tag{1.33}$$

and the number of selected components $m_{pls} \leq cn^{1/(2\mu+4)}$ for FPLS, when there is no information about the eigenvalues of K, with $c > 0$ and $\mu \geq 0$. Corollary 1 gives additional results concerning the number of components when we have additional informations about the eigenvalues of K. Based on these results, we can observe that FPLS method displays the same rate than FPCA and the number of selected components via FPLS is fewer than the one obtained by FPCA when the eigenvalues decrease at a polynomial rate, while both methods display the same number of selected components when the eigenvalues decrease at an exponential rate. Following Proposition 3, the convergence rate obtained from both methods is sharp and therefore the best possible uniform rate.

Remark 4.
The intuition behind the configuration considered by Hall and Horowitz (2007) is that the slope function is sufficiently smooth relative to the operator K (in terms of its eigenvalues), which in turn means that the slope function is efficiently represented by the m extracted eigenfunctions of K and therefore, this assumption is considered to guarantee the FPCA method to estimate properly the slope function. This configuration is very ambitious given that in the empirical analysis, we cannot verify such conditions. Moreover, even if it is possible to verify this, to the best of our knowledge, we usually do not obtain those patterns in economic or financial data in practice. If we consider a model where the two first Fourier coefficients of the slope function (in the eigensystem of K) vanish and the others are eventually different from zero, we will usually see that the FPCA will fail to estimate the first two eigenfunctions (and sometimes the first three eigenfunctions) that represent well the relation between the predictor variable and the response variable. Thus, more effort is required to estimate the slope function by considering more components (not the true ones that capture the relation between X and Y) according to the complexity of the linear relation between X and Y. Therefore, the FPLS tend to outperform the FPCA in terms of estimation. More details about this behavior are illustrated in the simulations.

Concerning the prediction error, given that after estimation, we have an additional smoothing step with the integration operator in order to get the prediction of Y, the difference between both methods is not very large in terms of prediction error.

1.3.4 Model selection

Since FPLS and FPCA methods involve a tuning parameter m, one of the main challenges is the choice of the optimal number of functional components m to consider in the model such that the MSPE is minimized. This question has been tackled by Stock and Watson (1998) in the situation of a multivariate linear regression model. They suggested to choose the number of latent components for which the mean squared forecast error (MSFE) is minimized. On the other hand, the BIC and AIC criteria have been proposed by Bai and Ng (2002) for the same type of model. We can also

identify other strategies such as Leave-one-out Cross-validation, K-Folds cross-validation, and Generalized Cross-validation. For this paper, the optimal number of components is chosen by using the K-Folds Cross-validation. The K-Folds cross-validation is given by the following formula. Let us split the initial sample in M subsamples denoted I_1,...,I_M.

$$m_{op} = \underset{m\in\mathcal{I}_0}{\operatorname{argmin}} \frac{1}{M}\sum_{\ell=1}^{M}\frac{1}{card(I_\ell)}\sum_{j\in I_\ell}\left(Y_j - \hat{Y}_j\right)^2. \tag{1.34}$$

For $\ell \in \{I_1,...,I_M\}$, we estimate the parameter β in the sample $\mathcal{I}_{-\ell}$ representing the observations not in $\mathcal{I}_\ell$. Then, we predict the response variable in $\mathcal{I}_\ell$ considered as the hold-out sample. $\hat{Y}_j$ is the prediction of the j^{th} observation in $\mathcal{I}_\ell$. Hence, we calculate the MSPE for each candidate m. $\mathcal{I}_0$ is the set of candidate m.

1.4 Simulation results

In this section, we run some simulations analysis of the estimator presented earlier. We highlight the difference between the FPLS and the FPCA results for $n = 1000$.

We consider 8 different models

$$Y = \int_S X(s)\beta(s)ds + \varepsilon,$$

where X and Y are both centered random variables. X maps on the space $S = [0, 1]$ and $\mathbb{H}$ is $L^2[0, 1]$. We consider that the errors ε are gaussian process with mean 0 and variance 1. We use the K-folds cross-validation in order to estimate the optimal number of components to be considered

The functional predictor is given by

$$X(t) = \sum_{j=1}^{\infty}\sqrt{\lambda_j}u_j v_j(t). \tag{1.35}$$

where $u_1, u_2, ...$ are i.i.d $\mathcal{N}(0, 1)$.

The following models are considered :

Model 1 : model considered by Cardot et al.(1999)

We have $v_j(t) = \sqrt{2}sin((j - 0.5)\pi t)$ and $\lambda_j = \frac{1}{(j-0.5)^2\pi^2}$, for $j = 1, 2, ...$ and $\beta(t) = 2sin(0.5\pi t) + 4sin(1.5\pi t) + 5sin(2.5\pi t)$. The slope function is a linear combination of exactly three eigenfunctions.

Model 2 : modified version of the model 1

We have $v_j(t) = \sqrt{2}sin((j - 0.5)\pi t)$ and $\lambda_j = \frac{1}{(j-0.5)^2\pi^2}$, for $j = 1, 2, ...$ and $\beta(t) = 6sin(3.5\pi t)$. The slope function is represented by the fourth eigenfunction of the covariance operator K.

Model 3 : model considered by Hall and Hosseini-Nasab (2006)

We have $v_j(t) = \sqrt{2}cos(j\pi t)$ and $\lambda_j = j^{-2}$, or $j = 1, 2, ...$ and $\beta(t) = \pi^2(t^2 - 1/3)$. Here, β is an infinite linear combination of the eigenfunctions with decreasing weights and the eigenvalues decrease at a polynomial rate.

Model 4 : modified version of model 3

We have $v_j(t) = \sqrt{2}cos(j\pi t)$ and $\lambda_j = (0.5)^{-2j}$, or $j = 1, 2, ...$ and $\beta(t) = \pi^2(t^2 - 1/3)$. Here, β is an infinite linear combination of the eigenfunctions with decreasing weights and the eigenvalues decrease at an exponential rate.

Model 5 : Model used by Hall and Horowitz (2007)

We have $v_1 = 1$, $\lambda_1 = 1$, $v_j(t) = \sqrt{2}cos((j-1)\pi t)$, and $\sqrt{\lambda_j} = j^{-1}$, for $j = 2, 3, ...$, and $\beta(t) = \sum_{j=1}^{\infty} \beta_j v_j(t)$, with $\beta_1 = 0.3$ and $\beta_j = 4(-1)^{j+1}j^{-2}$, for $j = 2, 3, ...$. In this case, the slope function β is an infinite linear combination of the eigenfunctions with decreasing weights and the eigenvalues $\lambda_1, \lambda_2, ...$ are well spaced.

Model 6 : modified version of the model 5

We have $v_1(t) = 1$, $\lambda_1 = 1$, $v_j = \sqrt{2}cos((j-1)\pi t)$, and $\sqrt{\lambda_j} = j^{-1}$, for $j = 2, 3, ...$, and $\beta(t) = \sum_{j=1}^{\infty} \beta_j v_j(t)$, with $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 3$, $\beta_4 = 5$, and $\beta_j = 4(-1)^{j+1}j^{-2}$, for $j = 5, 6, ...$. In this case, the slope function β does not depend on the first two eigenfunctions associated with the two largest eigenvalues of K.

Model 7 and 8 : smoothed and nonsmoothed predictor variables

We have $X_i = \pi(Z_i) + U_i$ where $\pi(s,t) = 1 - |s-t|^2$ is the kernel of an integral operator, $Z_i(t) = \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)+\Gamma(b_i)} + \eta_i$, $dU(t) = -U(t)dt + \sigma_u dG_u(t)$, G_u is a Wiener process, σ_u is the standard deviation of its incrementations dG_u, $a_i, b_i \sim$ iid $\mathcal{U}[2,5]$, $\eta_i \sim$ iid $\mathcal{N}(0,1)$ for $s, t \in [0,1]$, and $\beta(t) = \pi^2(t^2 - 1/3)$. Model 7 consider $\sigma_u = 0.2$, representing the case of smoothed predictor variables while for model 8 $\sigma_u = 2$ representing the nonsmoothed predictor variable.

To compute the integrals, we use the trapezoidal method in order to reduce the discretization bias. The simulation results are performed with the following procedure :

1. Write down a discretization of the interval $[0, 1]$ with 1000 equidistant points ;

2. Generate the true functional slope $\beta(t)$;

3. Generate n random variables ε_i from the standard normal distribution with zero mean and variance 1 ;

4. Generate n predictor function $X_i(t)$ for each t in the pseudo-interval constructed, with the infinite series truncated at $J = 50$;

5. Generate n response variables from the model $Y_i = \int_S X_i(s)\beta(s)ds + \varepsilon_i$, $i = 1, ..., 1000$;

6. Estimate the slope function β from the FPCA and FPLS approaches for the same number of latent factors selected and on a training sample representing 80% of the total number of observations and predict on a test sample, that is 20% of n ;

7. Choose the optimal number of components with the K-folds cross-validation ;

8. Repeat the steps 3 to 7 with 1000 iterations and calculate the MSE and MSPE on the test sample.

Model 1 and 2 consider the situation where the slope function is represented by a finite

number of eigenfunctions of the covariance operator K. For model 1, the slope function is well represented by the first three functional components ($v_1(t), v_2(t)$ and $v_3(t)$) that are the FPCA components while for model 2, the slope function is only represented by the fourth eigenfunction $v_4(t)$. We expect the FPCA method to estimate the slope function with the optimal number of components equal to three in the case of model 1, while for model 2 we expect to estimate the slope function with one component. We will see if FPLS method is able to estimate the slope function with three components for model 1 and one for model 2.

Model 3 and 4 consider the case where the slope function is smooth and the same for both models with the eigenvalues of the covariance operator K decreasing at a polynomial rate for model 3 and at an exponential rate for model 4. We compare the difference between the estimation techniques for the two different configurations of eigenvalues. These two models will contribute to analyse how well the considered methods behave in terms of selecting the number of components compared to what we obtained in corollary 1.

Model 5 and 6 compare the situation where the slope function an infinite linear combination of the eigenfunctions with decreasing weights and the eigenvalues $\lambda_1, \lambda_2, ...$ are distinct, to the one where slope function an infinite linear combination of the eigenfunctions with first four components not representing well the slope function.

Model 7 and 8 are used to see the difference between the estimation methods when the predictor variable is smooth and the case where the predictor variable is nonsmooth.

Figure 1.1 shows 10 observations of the functional predictor variables $X(t)$ for the 8 different models. The comparison of the FPCA and the FPLS is presented using the MSE and the MSPE. We will highlight how well the estimation results when the sample size is large.

Figure 1.3 presents the comparison of the MSPE of both methods using the 8 models for the different values with the number of observations $n = 1000$ and the standard deviation of the regression error $\sigma = 1$. We can observe that in overall, the optimal number of components selected via FPLS method tends to be less or equal to the one obtained via FPCA. This is observed on Table 1.7. Moreover, the predictive performance is almost the same for both methods.

Comparing model 1 and 2, we observe that the optimal number of selected components is the same for both methods, that is three for model 1 and one for model 2. This means that in the context where we know that the slope function is represented by a certain factor model, both the methods will estimate well the true number of factors. The point to highlight in this context is the fact that when the slope function is not efficiently represented by the first components which is the case of model 2, the optimal number of factor for the prediction purpose would be the optimal one while the number of components needed to estimate well the slope function should be the number representing the factor that efficiently present well the slope. For example in model 2, the slope function is efficiently represented by only the fourth component of K, but the optimal number of

components needed for a good prediction is one while the number of components for the estimation purpose is four.

Comparing model 3 and 4, we can observe that the number of selected components in model 3 is larger than in the case of model 4. This result means that in the context where the eigenvalues decrease at polynomial rate, we may need more components compared to case where the eigenvalues decrease at an exponential rate. This confirm the results obtained in corollary 1.

Concerning model 5 and 6, we can observe that when the predictor variables are smooth and the slope function is well represented by an infinite number of components, the optimal number of components is the same and equal to 3 (that is the case of model 5), while when the slope function is not well represented by the four first eigenfunctions, FPCA method needs 6 components to for a good prediction while we need only 3 components when using FPLS approach. This result hold since FPLS construct only factors that are related to the response variable and usually we obtain around 3 components.

When analyzing model 7 and 8, we can easily observe that in the context of smoothed predictor variables (model 7) we need more components for a good prediction compared to the situation where the predictor variables display more variability (model 8). These results are almost the same when looking at the median average prediction error (MAPE). For more details see Table 1.1 and Table 1.9.

When comparing the squared bias of the MSPE, we can observe that for a fixed number of components, the squared bias when using FPLS method is usually less or equal to the one obtained with FPCA. Table 1.5 displays more details about this result.

The results concerning the squared bias of the MSE are displayed in Table 1.3. We can observe that when m is less or equal to the optimal number of components, the squared bias with FPLS is less or equal to the one obtained via FPCA and when m is greater or equal to the optimal number of components, the squared bias with FPLS tends to be greater or equal to the one obtained via FPCA. This result hold for model 1 to 5. For model 7 and 8 we can observe that squared bias with FPLS is systematically less than the one obtained via FPCA.

Concerning the comparison of the variance term of the MSPE, we can observe that the one obtained when using FPLS is usually larger than the one obtained via FPCA (see Table 1.6). The results are similar when looking at the MSE (see Table 1.3 and 1.4). This confirms that the simulation results are in line with the theoretical one. The same results are obtained when the sample size is smaller but the quality of the estimations is not very good in this context. The results in the context of small sample are not reported.

Figure 1.2 presents the comparison of both estimation methods in terms of estimating the slope function. The blue red and yellow lines represent the true slope, the FPCA and FPLS estimates respectively. Concerning the comparison of the estimated slope functions with both methods using the optimal number of components, we can observe that for the 8 different models, both of the estimation methods are able to recover well the shape

of the true slope function as the sample size becomes large. Furthermore, the FPLS method seems to be generally closer to the true slope function than the FPCA estimation approach. When the true slope function is smoother than the predictor function (Model 3), the estimated slope functions are almost similar.

Moreover, when the slope function and the predictor variables are smooth (model 7), the slope function is well estimated, while when the slope function is smooth and the predictor variables are not smooth (model 8), the slope function is not well estimated and in particular at the beginning and the end of the interval of the function. Thus, if the practitioners have some prior information about the smoothness of the slope function, they can choose one of the methods accordingly. Moreover, if the practitioners do not need much variability in the estimation of the slope function, they can use the FPCA approach ; however, if they need a small bias, they should use the FPLS method. On the other hand, as we expect, according to Table 1.1, both methods tend to reach the same prediction error and this is due to the additional smoothing step with the integration operation in order to predict $Y = \int_0^1 X(t)\beta(t)dt$.

1.5 Real data application : stock market return prediction

A simple application of the functional linear regression model is presented in this section. We compare the FPCA and the FPLS method based on real data. We consider a problem of predicting the return of the S&P 500 in the New York Stock Exchange.

1.5.1 The literature

The New York Stock Exchange is endowed by a large set of stocks that represents the shares of corporations. We consider a simple prediction model where we use the previous day cumulative return to predict the next day instant return. The cumulative returns are observed on the one-minute frequency. The idea is motivated by the fact that market participants are likely to take advantage of the additional information coming from the intraday price values in order to improve the next day return prediction. Furthermore, one can consider price series as the discretization of a curve that contains enough information to predict the next week instant return. This idea holds in the sense that one of the most popular ways to analyze asset prices is to look at the price evolution during a certain period for prediction and not at the price at a certain time point. This means that market participants take the whole price structure in a given window as an information to account in their price prediction procedure.

1.5.2 The model

We assume that the whole price evolution of the S&P 500 within a day can contribute to predict the next day instant return. We consider that we observe the whole price evolution within a day on the one-minute frequency and we predict the next daily return. Since the New York Stock Market session is opened between 09 :30 AM and 04 :00 PM, every single trading day is represented by 390 minutes. If $P_t(s)$ is the price of a stock at the minute s of the day t (with $s \in (\frac{j-1}{390}, \frac{j}{390}]$ for $j = 1, ..., 390$), then the cumulative return between the minute s and the first minute of the day t is given by

$$X_t(s) = 100 * \Big[ln(P_t(s)) - ln(P_t(1)) \Big].$$

Therefore, the basic model is presented as follows

$$r_{t+1} = \sum_{s=1}^{390} X_t(s)\beta_s + \varepsilon_{t+1},$$

where r_{t+1} is the market return between two consecutive days and is given by

$$r_{t+1} = 100 * \Big[ln(P_{t+1}(390)) - ln(P_t(390)) \Big],$$

and ε_{t+1} is the error term. Then, at each day t one observes the whole price evolution from 09 :30 AM to 04 :00 PM as a unique curve and predicts the return of the next day.

Since the cumulative intraday returns are assumed to be continuous functions observed within a day, the model can be written as

$$r_{t+1} = \int_0^1 X_t(s)\beta(s)ds + \varepsilon_{t+1},$$

where we have normalized the integration interval to $[0, 1]$. The discretized cumulative intraday returns are interpolated on a B-spline basis to ensure the continuity of the functions and to complete the missing data within each day. The cumulative intraday return curve $X(s)$ is proportional to the price and will display the same pattern as the original price curve. Figure 1.4 presents the price evolution of 100 selected days on the discretized level (at the bottom) and the functional cumulative return curves (at the top). The price evolution is presented on the one-minute timeframe for each day. We shall specify here that the data are all centered and scaled. r_{t+1} represents the target return for the next day to be predicted for the considered stock, that is, the one day ahead return. Our purpose is to estimate the slope function based on the techniques presented in this paper, predict the returns, and evaluate both the techniques based on the predictive performance in-sample and out-of-sample.

1.5.3 Data Preprocessing

The sample data of the S&P 500 are collected in the historical data of the website www.backtestmarket.com. Due to the limited access, we will focus only on data from 1 January 2014 to 31^{th} December 2017, which is 1007 daily price values. After the pretreatment of the data to get the variables $X_t(s)$ and r_t, we obtain a sample of 1007 observations. For the construction of the cumulative intraday returns $X(t)$, we preliminarily group the 1-minute discrete observations obtained between 09 :30 AM to 04 :00 PM for the 1007 daily curves. To ensure each daily curve is of the same length, we use interpolation to complete the missing data at the end of each day. This operation ends up with $N = 392730$ minutes points that are divided into 1007 equal-spaced curves. We split our sample into 3 sub-samples, that are training, validation, and test sample (70%, 15%, and 15% of the whole sample respectively). The training and validation sample are used to train the model and select the optimal number of components, while the test sample is used to track the out-of-sample performance of the methods. The 10-folds cross-validation is used in the selection process of the tuning parameter and we compare the results obtained based on the out-of-sample (oos) $MSPE$, R^2_{is} and R^2_{oos}, where R^2_{is} and R^2_{oos} are respectively the in-sample and out-of-sample R-squared. The R^2_{oos} belongs to the interval $(-\infty, 1]$, where a negative value indicates a less accurate forecast than the target's historical mean and a value of R^2_{oos} closer to 1 represents a better predictive performance.

$$R^2_{is} = 1 - \frac{\sum_{i=1}^{N}(r_{tri} - \hat{r}_{tri})^2}{\sum_{i=1}^{N}(r_{tri} - \bar{r}_{tr})^2},$$

$$R^2_{oos} = 1 - \frac{\sum_{i=1}^{N}(r_{osi} - \hat{r}_{osi})^2}{\sum_{i=1}^{N}(r_{osi} - \bar{r}_{os})^2},$$

where r_{tri} is the observation i of the returns in the training sample and r_{osi} is the observation i of the return out of sample.

1.5.4 Estimation results

Figure 1.6 shows the estimated slope function obtained by the FPCA and FPLS methods, respectively. The estimated slope function $\hat{\beta}(t)$ presents almost the same shape with slightly more variability observed when using the FPLS method. The estimated values of the functional slope are read horizontally. For example, $\hat{\beta}(04:00PM) = \hat{\beta}(390)$ represents how the end of previous day's cumulative return is correlated with the next day's return, and $\hat{\beta}(09:30AM) = \hat{\beta}(0)$ represents the correlation between the beginning of the previous day's cumulative return and the next day's instant return. We can observe a positive value of the slope function from 03 :00 PM to 04 :00 PM with FPCA, while it can be observed earlier when using FPLS, that is 01 :00 PM to 04 :00 PM. Additionally

we have a less significant or negative correlation in the period 09 :30 AM to 10 :30 AM. The high correlation between the end of the previous day's return and the next day's return can be explained by the fact that at the end of the previous day we have more information about the market and market participants will react accordingly during the next day. For example it could be the forced sales by hedge funds and also news announcements. Furthermore, the FPLS approach seems to present larger magnitudes of the slope function than the FPCA method.

We also address the question concerning the predictive power of the concerned methods. To do so, we plot the predicted out-of-sample daily return for the FPCA and the FPLS estimation methods. Figure 1.7 displays the predicted daily returns from May 2017 to December 2017. We can observe that both methods tend to capture well the overall return variations on the hold-out-sample. Moreover, Figure 1.8 suggests that the residuals are normally distributed.

Table 1.8 presents the MSPE, R^2_{is}, R^2_{oos} for both estimation methods. We can observe that the root mean squared prediction error (RMSPE) for the FPLS approach is almost similar to the one obtained by the FPCA approach. This was expected given that the eigenvalues decrease very quickly as in model 8 of our simulations, which shows that, when the predictor function presents more variability, the FPLS and FPCA tend to display the same predictive performance. Furthermore, the FPLS method displays a slightly larger R^2_{is} than the FPCA, that is 23.1% (compared to 22.4% for FPCA). We observe the same result for the out-of-sample R^2. Indeed, it is evaluated at 2.2% for FPLS and 1.75% for FPCA. The OLS model tends to present an overfitting problem and this result was expected. In fact, it can be observed that it displays an R^2_{is} of 77.8% while the R^2_{oos} is -127.7%. Thus, it is highly outperformed by the other methods.

1.6 Conclusion

This paper investigates and compares the theoretical features of two dimension reduction techniques for a functional linear regression model with a scalar response. The first one is the Functional Principal Component Analysis (FPCA) and the second one is the Functional Partial Least Squares (FPLS). The theoretical results and the empirical simulations suggest that the squared bias of the conditional MSE when using the FPLS method is usually smaller than the one obtained with the FPCA approach when the same number of latent components is considered. Moreover, the variance obtained when using the FPLS method is usually larger than the one obtained by FPCA. Furthermore, the estimation error obtained with FPLS could increase very quickly and may explode as the number of retained components increases. A theoretical upper bound rate is derived for both methods under a certain number of regularity conditions and we find that both methods display the same convergence rate of the MSE. Also we show that this convergence rate is sharp. Additionally, we show that FPLS tends to display the

same convergence rate of the MSE as FPCA and it selects usually a fewer number of components than FPCA.

We also find that when the slope function and the predictor variables are smooth, FPCA and FPLS tend to display almost the same estimation and prediction performance, while when the predictor variables are not smooth enough, both methods are not able to estimate well the slope function and the results with FPCA are worse than the one with FPLS. Those results are more observable in the context where the sample size is large. Additionally, the convergence rate of the FPCA method depends essentially on the decreasing rate of the eigenvalues, while the one of the FPLS method relies on the smallest roots of the residual polynomial. Based on the real data application, we can notice that the MSPE for the FPLS tends to be smaller than that of the FPCA.

Moreover, the MSE for the FPLS is usually smaller than the one obtained by the FPCA and especially when the number of chosen latent components is lower or equal to the optimal number of latent components (selected via cross-validation). Ultimately, FPLS performs well when we choose the optimal number of latent components for each method. This suggests to the practitioner that if they care more about the estimation, they can use the FPLS method ; however, if the emphasis is on the prediction of the response, they could choose either methods.

1.7 Appendix.

1.7.1 Mathematical tools on orthogonal polynomials theory

Let us define
$\underline{X} = \Big[X_1, X_2, ..., X_n\Big]'$
$\underline{Y} = \Big[Y_1, Y_2, ..., Y_n\Big]'$
and
$\underline{\varepsilon} = \Big[\varepsilon_1, \varepsilon_2, ..., \varepsilon_n\Big]'$

Let us denote by $Z : [a, b] \to \mathbb{R}$ a bounded and mesurable function with a and b as positive numbers ($a < b$). Then,

$$Z(K) = \sum_{j=1}^{+\infty} Z(\lambda_j) v_j \otimes v_j,$$

$$Z(\hat{K}) = \sum_{j=1}^{n} Z(\hat{\lambda}_j) \hat{v}_j \otimes \hat{v}_j$$

and $TZ(W_x) = Z(K)T_x^*$ since $K = T_x^* T_x$ and $W_x = T_x T_x^*$. For an integer ν, let us define the empirical version measure

$$d\mu_n^{\nu} := \sum_{j=1}^{n} \hat{\lambda}_j^{\nu} < T_n^* \underline{Y}, \hat{v}_j >^2 \delta_{\hat{\lambda}_j}$$

where $\delta_{\hat{\lambda}_j} = 1$ if $t = \hat{\lambda}_j$ and 0 otherwise. Then, for each function Z, we have

$$||Z(\hat{K})T_n^*\underline{Y}||^2 = \sum_{j=1}^{n} \hat{\lambda}_j Z(\hat{\lambda}_j)^2 < T_n^*\underline{Y}, \hat{v}_j >^2 \delta_{\hat{\lambda}_j} = \int_0^{\hat{\lambda}_1} Z^2(t) d\mu_n^{(0)}(t).$$

Let us define by $\mathcal{R}_{m-1}$, the vector space of polynomials of at most $m-1$ degrees and $\mathcal{R}_m^0$ the space of real polynomials of at most m degrees with the constant term equal to 1. Then, $\hat{\beta}_m^{PLS} = \hat{P}_m(W_n)T_n^*\underline{Y}$ where $P_m \in \mathcal{R}_{m-1}$ and the residual polynomial related to the functional regression model is given by

$$\hat{Q}_m(t) = 1 - t\hat{P}_m(t)$$

and $\hat{Q}_m \in \mathcal{R}_m^0$. Moreover, we have $||T_n^*(\underline{Y} - T_n(\hat{\beta}_m^{PLS}))|| = ||\hat{Q}(\hat{K})T_n^*Y||$.

Let us define the scalar product related to the space $\mathcal{R}_n$ as

$$\begin{aligned}
[\phi, \psi]_{(\nu)} &= \int_0^{\hat{\lambda}_1} \phi(t)\psi(t) d\mu_n^{(\nu)}(t) \\
&= \left\langle \phi(W_n)\underline{Y}, W_n^{\nu}\psi(W_n)\underline{Y} \right\rangle \\
&= \sum_{j=1}^{n} \hat{\lambda}_j^{\nu}\phi(\hat{\lambda}_j)\psi(\hat{\lambda}_j) < T_n^*\underline{Y}, \hat{v}_j >^2 .
\end{aligned}$$

See page 9 of Hanke (1995). Then, in our model $[\phi, \psi]_{(1)} = [\phi, \psi]$ for the modified PLS. Therefore, following Theorem 1, the estimation via FPLS method using the Krylov space approach is equivalent to identifying the polynomial $\hat{Q}_m(t)$ of degree at most m such that

$$\hat{Q}_m(t) = 1 - t\hat{P}_m(t)$$

where

$$\hat{P}_m = \underset{\phi \in \mathcal{R}_{m-1}}{argmin} \left|\left|(I - \hat{K}\phi(\hat{K}))T_n^*\underline{Y}\right|\right|^2 .$$

This means that $\hat{Q}_m$ is the minimizer of $\left|\left|Q(\hat{K})T_n^*\underline{Y}\right|\right|^2 = [Q, Q]$ with $Q \in \mathcal{R}_m^0$. Moreover, $[\hat{Q}_m, \mathcal{T}P] = [\hat{Q}_m, P]_{(1)} = 0$ for each $P \in \mathcal{R}_{m-1}$, where $\mathcal{T}P(t) = tP(t)$. Then, the sequence $\hat{Q}_0, ..., \hat{Q}_{n-1}$ is orthogonal with respect to $[.,.]$ and we can have at most n orthogonal polynomials for the FPLS empirically (reflecting the maximum possible number of functional components).

Proof of Proposition 1.
The proof of the Proposition 1 is similar to the proof of Proposition 7.1 of Blazère et al. (2014b) and hence omitted.

Proof of Theorem 2.
Upper bound of the squared bias.
Under the model condition, if the assumptions A1 - A4 hold, then we obtain the following result :

$$||\beta_m^{PCA} - \beta||^2 = O\Big(\lambda_{m+1}^{\mu}\Big).$$

and

$$||\beta_\alpha^{SC} - \beta||^2 = O\Big(\alpha^{\mu}\Big).$$

Proof.
We have

$$\begin{aligned}
||\beta_m^{PCA} - \beta||^2 &= ||\sum_{j=1}^{m} < \beta, v_j > v_j - \sum_{j=1}^{\infty} < \beta, v_j > v_j||^2 \\
&= \sum_{j=m+1}^{\infty} < \beta, v_j >^2 \\
&= \sum_{j=m+1}^{\infty} \lambda_j^{\mu} \frac{< \beta, v_j >^2}{\lambda_j^{\mu}} \\
&\leq \underset{j \geq m+1}{sup} \{\lambda_j^{\mu}\} \sum_{j=m+1}^{\infty} \frac{< \beta, v_j >^2}{\lambda_j^{\mu}} \\
&\leq \lambda_{m+1}^{\mu} \sum_{j=m+1}^{\infty} \frac{< \beta, v_j >^2}{\lambda_j^{\mu}}.
\end{aligned}$$

Since, assumption A4 holds, we obtain the result.

Following the same logic with FSC method, we have

$$
\begin{aligned}
||\beta_\alpha^{SC} - \beta||^2 &\leq || \sum_{\lambda_j \geq \alpha} < \beta, v_j > v_j - \sum_{j=1}^{\infty} < \beta, v_j > v_j||^2 \\
&\leq \sum_{\lambda_j < \alpha} < \beta, v_j >^2 \\
&\leq \sum_{\lambda_j < \alpha} \lambda_j^\mu \frac{< \beta, v_j >^2}{\lambda_j^\mu} \\
&\leq \underset{\lambda_j < \alpha}{sup}\{\lambda_j^\mu\} \sum_{\lambda_j < \alpha} \frac{< \beta, v_j >^2}{\lambda_j^\mu} \\
&\leq \alpha^\mu \sum_{\lambda_j < \alpha} \frac{< \beta, v_j >^2}{\lambda_j^\mu}.
\end{aligned}
$$

since, $||X||^2 = O_p(1)$ and assumption A4 hold, we obtain the result.

Upper bound of the estimation error.

Under the model condition, if assumptions A1 - A4 hold, we obtain the following result :

$$
\mathbb{E}\left[||\hat{\beta}_m^{PCA} - \beta_m^{PCA}||^2 | X\right] = O_p\left(\frac{m}{n\lambda_m}\right) \tag{1.36}
$$

and

$$
\mathbb{E}\left[||\hat{\beta}_\alpha^{SC} - \beta_\alpha^{SC}||^2 | X\right] = O_p\left(\frac{1}{n\alpha^2}\right) \tag{1.37}
$$

Proof for FPCA.

Note that $\hat{C}_{xy} = \frac{1}{n}\sum_{i=1}^n X_i(t)Y_i = \hat{K}\beta + \hat{C}_{x\varepsilon}$ where, $\hat{C}_{x\varepsilon} = \frac{1}{n}\sum_{i=1}^n X_i(t)\varepsilon_i$ and $C_{xy} = K\beta$. Hence, we have

$$
\begin{aligned}
\hat{\beta}_m^{PCA} - \beta_m^{PCA} &= \sum_{j=1}^m \frac{1}{\hat{\lambda}_j} < \hat{C}_{xy}, \hat{v}_j > \hat{v}_j - \sum_{j=1}^m \frac{1}{\lambda_j} < C_{xy}, v_j > v_j \\
&= \sum_{j=1}^m \frac{1}{\hat{\lambda}_j} < \hat{K}(\beta), \hat{v}_j > \hat{v}_j - \sum_{j=1}^m \frac{1}{\lambda_j} < K(\beta), v_j > v_j + \sum_{j=1}^m \frac{1}{\hat{\lambda}_j} < \hat{C}_{x\varepsilon}, \hat{v}_j > \hat{v}_j \\
&= \sum_{j=1}^m < \beta, \hat{v}_j > \hat{v}_j - \sum_{j=1}^m < \beta, v_j > v_j + \sum_{j=1}^m \frac{1}{\hat{\lambda}_j} < \hat{C}_{x\varepsilon}, \hat{v}_j > \hat{v}_j.
\end{aligned}
$$

Then, we have

$$||\hat{\beta}_m^{PCA} - \beta_m^{PCA}||^2 \leq 2||\sum_{j=1}^{m} < \beta, \hat{v}_j > \hat{v}_j - \sum_{j=1}^{m} < \beta, v_j > v_j||^2 + 2||\sum_{j=1}^{m} \frac{1}{\hat{\lambda}_j} < \hat{C}_{x\varepsilon}, \hat{v}_j > \hat{v}_j||^2$$
$$\leq 2||\sum_{j=1}^{m} < \beta, \hat{v}_j > \hat{v}_j - \sum_{j=1}^{m} < \beta, v_j > v_j||^2 + 2\sum_{j=1}^{m} \frac{1}{\hat{\lambda}_j^2} < \hat{C}_{x\varepsilon}, \hat{v}_j >^2$$

Let us define $B_1 = ||\sum_{j=1}^{m} < \beta, \hat{v}_j > \hat{v}_j - \sum_{j=1}^{m} < \beta, v_j > v_j||^2$ and $B_2 = \mathbb{E}\left[\sum_{j=1}^{m} \frac{1}{\hat{\lambda}_j^2} < \hat{C}_{x\varepsilon}, \hat{v}_j >^2\right]$. Then, under the conditions B1 and B2, we have

$$B_1 = ||\sum_{j=1}^{m} < \beta, \hat{v}_j > \hat{v}_j - \sum_{j=1}^{m} < \beta, v_j > v_j||^2 = O_p\left(\frac{1}{n}\right)$$

by Propositions 3 and 7 of Dauxois et al. (1982). Moreover, we have

$$\mathbb{E}\left[\sum_{j=1}^{m} \frac{1}{\hat{\lambda}_j^2} < \hat{C}_{x\varepsilon}, \hat{v}_j >^2 |X\right] = \sum_{j=1}^{m} \mathbb{E}\left[\frac{1}{\hat{\lambda}_j^2} < \hat{C}_{x\varepsilon}, \hat{v}_j >^2 |X\right]$$
$$= \sum_{j=1}^{m} \frac{1}{\hat{\lambda}_j^2} \mathbb{E}\left[< \hat{C}_{x\varepsilon}, \hat{v}_j >^2 |X\right].$$

Furthermore,

$$\begin{aligned}
\mathbb{E}[< \hat{C}_{x\varepsilon}, \hat{v}_j >^2 |X] &= \mathbb{E}\left[\int_S (\hat{v}_j(t)\frac{1}{n}\sum_{i=1}^{n}(X_i(t) - \bar{X}(t))\varepsilon_i)dt \int_S (\hat{v}_j(s)\frac{1}{n}\sum_{l=1}^{n}(X_l(s) - \bar{X}(s))\varepsilon_l)ds|X\right] \\
&= \int_S \int_S \hat{v}_j(t)\mathbb{E}\left[(\frac{1}{n}\sum_{i=1}^{n}(X_i(t) - \bar{X}(t))\varepsilon_i)(\frac{1}{n}\sum_{l=1}^{n}(X_l(s) - \bar{X}(s))\varepsilon_l)\right]\hat{v}_j(s)dsdt \\
&= \int_S \int_S \hat{v}_j(t)\left[\frac{\sigma^2}{n}\sum_{i=1}^{n} X_i(t)X_l(s)\right]\hat{v}_j(s)dsdt \\
&= \frac{\sigma^2}{n} < \hat{v}_j, \hat{K}(\hat{v}_j) > \\
&= \frac{\sigma^2\hat{\lambda}_j}{n}.
\end{aligned} \tag{1.38}$$

Therefore,

$$\mathbb{E}\Big[\sum_{j=1}^{m}\frac{1}{\hat{\lambda}_j}<\hat{C}_{x\varepsilon},\hat{v}_j>^2|X\Big]=\sum_{j=1}^{m}\frac{1}{\hat{\lambda}_j}\mathbb{E}\Big[<\hat{C}_{x\varepsilon},\hat{v}_j>^2|X\Big]$$
$$=\frac{\sigma^2}{n}\sum_{j=1}^{n}\frac{1}{\hat{\lambda}_j}$$
$$\leq\frac{\sigma^2}{n}\frac{m}{\hat{\lambda}_m}.$$

In addition, $\hat{\lambda}_m$ is a consistent estimator of λ_m, we have

$$\frac{m\sigma^2}{n\hat{\lambda}_m}=O_p\Big(\frac{m\sigma^2}{n\lambda_m}\Big).$$

Then, the estimation error has the following rate of convergence.

$$||\hat{\beta}_m^{PCA}-\beta_m^{PCA}||^2=O_p\Big(\frac{m\sigma^2}{n\lambda_m}\Big). \tag{1.39}$$

Proof for FSC.

Using the same logic as with FPCA method, we have

$$\hat{\beta}_\alpha^{SC}-\beta_\alpha^{SC}=\sum_{\hat{\lambda}_j\geq\alpha}\frac{1}{\hat{\lambda}_j}<\hat{C}_{xy},\hat{v}_j>\hat{v}_j-\sum_{\lambda_j\geq\alpha}\frac{1}{\lambda_j}<C_{xy},v_j>v_j$$
$$=\sum_{\hat{\lambda}_j\geq\alpha}<\beta,\hat{v}_j>\hat{v}_j-\sum_{\lambda_j\geq\alpha}<\beta,v_j>v_j+\sum_{\hat{\lambda}_j\geq\alpha}\frac{1}{\hat{\lambda}_j}<\hat{C}_{x\varepsilon},\hat{v}_j>\hat{v}_j.$$

Then, we have

$$||\hat{\beta}_\alpha^{SC}-\beta_\alpha^{SC}||^2\leq 2||\sum_{\hat{\lambda}_j\geq\alpha}<\beta,\hat{v}_j>\hat{v}_j-\sum_{\lambda_j\geq\alpha}<\beta,v_j>v_j||^2+2||\sum_{\hat{\lambda}_j\geq\alpha}\frac{1}{\hat{\lambda}_j}<\hat{C}_{x\varepsilon},\hat{v}_j>\hat{v}_j||^2$$
$$\leq 2||\sum_{\hat{\lambda}_j\geq\alpha}<\beta,\hat{v}_j>\hat{v}_j-\sum_{\lambda_j\geq\alpha}<\beta,v_j>v_j||^2+2\sum_{\hat{\lambda}_j\geq\alpha}\frac{1}{\hat{\lambda}_j^2}<\hat{C}_{x\varepsilon},\hat{v}_j>^2.$$

Let us define D_1 and D_2 by $D_1=||\sum_{\hat{\lambda}_j\geq\alpha}<\beta,\hat{v}_j>\hat{v}_j-\sum_{\lambda_j\geq\alpha}<\beta,v_j>v_j||^2$ and $D_2=\mathbb{E}\Big[\sum_{\hat{\lambda}_j\geq\alpha}\frac{1}{\hat{\lambda}_j^2}<\hat{C}_{x\varepsilon},\hat{v}_j>^2|X\Big]$.

Then,

$$
\begin{aligned}
D_2 &= \mathbb{E}\Big[\sum_{\hat{\lambda}_j \geq \alpha} \frac{1}{\hat{\lambda}_j^2} < \hat{C}_{x\varepsilon}, \hat{v}_j >^2 |X\Big] \\
&\leq \frac{1}{\alpha^2} \sum_{\hat{\lambda}_j \geq \alpha} \mathbb{E}\Big[< \hat{C}_{x\varepsilon}, \hat{v}_j >^2 |X\Big] \\
&\leq \frac{\sigma^2}{n\alpha^2} \sum_{j=1}^{n} \hat{\lambda}_j \\
&\leq \frac{\sigma^2}{n\alpha^2} \\
&= O_p\Big(\frac{1}{n\alpha^2}\Big).
\end{aligned}
$$

Then, $D_2 = O_p\Big(\frac{1}{n\alpha^2}\Big)$. The third line holds following Equation (1.38). The last line holds since $\sum_{j=1}^{n} \hat{\lambda}_j < \infty$.

Now we will prove that

$$
D_1 = || \sum_{\hat{\lambda}_j \geq \alpha} < \beta, \hat{v}_j > \hat{v}_j - \sum_{\lambda_j \geq \alpha} < \beta, v_j > v_j||^2 = O_p\Big(\frac{1}{\alpha^2 n}\Big).
$$

This term is treated using an approach similar to that of Proof of Proposition 2 by Carrasco and Rossi (2016). Let us define $I_\alpha = \{j : \lambda_j \geq \alpha\}$ and $\hat{I}_\alpha = \{j : \hat{\lambda}_j \geq \alpha\}$. Without loss of generality, let us assume that $I_\alpha \subset \hat{I}_\alpha$. Let us define $D_3 = || \sum_{\hat{\lambda}_j \geq \alpha} < \beta, \hat{v}_j > \hat{v}_j - \sum_{\lambda_j \geq \alpha} < \beta, v_j > v_j||^2$. Then,

$$
\begin{aligned}
D_3 &= \Big|\Big| \sum_{j \in I_\alpha} < \beta, \hat{v}_j > \hat{v}_j + \sum_{j \in \hat{I}_\alpha - I_\alpha} < \beta, \hat{v}_j > \hat{v}_j - \sum_{j \in I_\alpha} < \beta, v_j > v_j \Big|\Big|^2 \\
&\leq 2 \Big|\Big| \sum_{j \in I_\alpha} < \beta, \hat{v}_j > \hat{v}_j - \sum_{j \in I_\alpha} < \beta, v_j > v_j \Big|\Big|^2 + 2 \Big|\Big| \sum_{j \in \hat{I}_\alpha - I_\alpha} < \beta, \hat{v}_j > \hat{v}_j \Big|\Big|^2.
\end{aligned}
$$

According to Propositions 3 and 7 of Dauxois et al. (1982), the projection operator on the eigenspace associated with the same eigenvalue λ_j, $\sum_{j \in I_i} < \hat{\phi}_j, . > \hat{\phi}_j$ converges to $\sum_{j \in I_i} < \phi_j, . > \phi_j$ at the $\sqrt{n}$ rate.

Furthermore,

$$
\begin{aligned}
|I_\alpha| &= \sum_j I(\lambda_j \geq \alpha) \\
&= \sum_j \frac{\lambda_j}{\lambda_j} I(\lambda_j \geq \alpha) \\
&\leq \frac{1}{\alpha} \sum_j \lambda_j I(\lambda_j \geq \alpha) \\
&\leq \frac{1}{\alpha} \sum_{j=1}^{\infty} \lambda_j.
\end{aligned}
$$

As $\sum_{j=1}^{\infty} \lambda_j < \infty$, we have $|I_\alpha| = O(\frac{1}{\alpha})$ and therefore

$$
\begin{aligned}
\left|\left| \sum_{j \in I_\alpha} < ., \hat{v}_j > \hat{v}_j - \sum_{j \in I_\alpha} < ., v_j > v_j \right|\right|^2 &= O_p\left(\frac{|I_\alpha|^2}{n}\right) \\
&= O_p\left(\frac{1}{\alpha^2 n}\right).
\end{aligned}
$$

Remark. For this proof, the eigenvalues are not necessarily distinct.
Moreover,

$$
\begin{aligned}
\left|\left| \sum_{j \in \hat{I}_\alpha - I_\alpha} < \beta, \hat{v}_j > \hat{v}_j \right|\right|^2 &= \sum_{j \in \hat{I}_\alpha - I_\alpha} < \beta, \hat{v}_j >^2 \\
&\leq |\hat{I}_\alpha - I_\alpha| \underset{j}{Sup} \{< \beta, \hat{v}_j >^2\}
\end{aligned}
$$

and $|\hat{I}_\alpha - I_\alpha| = O_p\left(\frac{1}{\sqrt{n}}\right)$ since $|\hat{\lambda}_j - \lambda_j| = O_p\left(\frac{1}{\sqrt{n}}\right)$. Then,

$$
|| \sum_{j \in \hat{I}_\alpha - I_\alpha} < \beta, \hat{v}_j > \hat{v}_j ||^2 = O_p\left(\frac{1}{n}\right).
$$

and therefore $D_1 = O_p\left(\frac{1}{\alpha^2 n}\right) + O_p\left(\frac{1}{n}\right) = O_p\left(\frac{1}{\alpha^2 n}\right)$.
We conclude that

$$
\left|\left| \hat{\beta}_\alpha^{SC} - \beta_\alpha^{SC} \right|\right|^2 = O_p\left(\frac{1}{n\alpha^2}\right). \tag{1.40}
$$

By combining the upper bound of the squared bias and the estimation error, we obtain the convergence rate of the conditional MSE for both methods.

Proof of Proposition 2.
We have

$$\begin{aligned}
\left\|\beta_m^{PLS}-\beta\right\|^2 &= \sum_{j=1}^{\infty}(1-\overline{Q}_{m,j})^2 < \beta, v_j >^2 \\
&\leq \sum_{j=1}^{\infty} Q_m^2(\lambda_j) < \beta, v_j >^2 \\
&\leq \sum_{j=1}^{\infty}\Big[\sum_{(j_1,...,j_m)\in I^+} w_{j_1,...,j_m}\prod_{l=1}^{m}(1-\frac{\lambda_j}{\lambda_{j_l}})\Big]^2 < \beta, v_j >^2 \\
&\leq \sum_{j=1}^{\infty} \underset{(j_1,...,j_m)\in I^+}{sup}\prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_{j_l}}\right)^2 < \beta, v_j >^2 \\
&\leq \sum_{j=1}^{\infty}\prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_l}\right)^2 < \beta, v_j >^2
\end{aligned}$$

because for $j = 1, ..., m$, $\prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_l}\right)^2 = 0$, while for $j \geq m+1$, we have $0 \leq \prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_l}\right)^2 \leq 1$, because all of the eigenvalues are nonzero ordered in such a way that $\lambda_1 > \lambda_2 > .... > \lambda_m > \lambda_{m+1} > ... > 0$. We should also recall that $0 \leq w_{j_1,...,j_m} \leq 1$, with $\sum_{(j_1,...,j_m)\in I_m^+} w_{j_1,...,j_m} = 1$. The last inequality follow from the fact that the eigenvalues are all distinct and

$$\underset{(j_1,...,j_m)\in I^+}{sup}\prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_{j_l}}\right)^2 = \prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_l}\right)^2 \leq 1.$$

Therefore, we have

$$\begin{aligned}
\left\|\beta_m^{PLS}-\beta\right\|^2 &= \sum_{j=1}^{\infty}(1-\overline{Q}_{m,j})^2 < \beta, v_j >^2 \\
&\leq \sum_{j=1}^{\infty}\prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_l}\right)^2 < \beta, v_j >^2 \\
&\leq \sum_{j=m+1}^{\infty} < \beta, v_j >^2 \\
&= \left\|\beta_m^{PCA}-\beta\right\|^2.
\end{aligned}$$

Then, $\left|\left|\beta_m^{PLS}-\beta\right|\right|^2 \leq \left|\left|\beta_m^{PCA}-\beta\right|\right|^2$ and

$$\left|\left|\beta_m^{PLS}-\beta\right|\right|^2 = O_p\left(\lambda_{m+1}^{\mu}\right),$$

We should point out that this is an adaptation of the results of Carrasco and Rossi (2016) for the functional regression model.

Lemma 1.

Let m $(1 \leq m \leq n)$ be the number of selected components for the FPLS method. Then, we have :
(i) The residual polynomial $\hat{Q}_m$ has exactly m distinct roots denoted by $(\hat{\theta}_l)_{l=1,...m}$ with $\hat{\theta}_1 > \hat{\theta}_2 > ... > \hat{\theta}_m > 0$.
(ii) Q_m is positive, decreasing, and convex on the interval $[0, \hat{\theta}_m)$.
(iii) Let us define the function φ_m on the interval $[0, \hat{\theta}_m)$ as

$$\varphi_m(t) = \hat{Q}_m(t)\left(\frac{\hat{\theta}_m}{\hat{\theta}_m - t}\right)^{1/2}.$$

Then, we have $\hat{Q}_m(0) = 1$,

$$||Q_m(\hat{K})(T_n^*(\underline{Y}))|| \leq ||(I - \hat{\Pi}_{\hat{\theta}_m})\{\varphi_m(\hat{K})(T_n^*(\underline{Y}))\}||,$$

and for any nonnegative real number $\gamma \geq 0$, we have

$$\underset{t\in[0,\hat{\theta}_m]}{Sup}\ |t^\gamma \varphi_m^2(t)| \leq \gamma^\gamma |\hat{Q}_m'(0)|^{-\gamma},$$

$$\underset{t\in[0,\hat{\theta}_m]}{Sup}\ |\varphi_m(t)| \leq 1,$$

where $\hat{\Pi}_{\hat{\theta}_m}\phi = \sum_{j=1,\hat{\lambda}_j<\hat{\theta}_m}^n < \phi, \hat{v}_j > \hat{v}_j$ is the orthogonal projection on the eigenfunctions $\hat{v}_j$ of $\hat{K}$ related to the eigenvalues $0 < \hat{\lambda}_j < \hat{\theta}_m$. Then $(I - \hat{\Pi}_{\hat{\theta}_m})$ is orthogonal projection onto the space spanned by the eigenfunctions $\hat{v}_j$ of $\hat{K}$ related to the eigenvalues $\hat{\lambda}_j > \hat{\theta}_m$.
(iv) $\hat{\lambda}_1 > \hat{\theta}_1 \geq ... \geq \hat{\theta}_m > 0$
(v) $\hat{\theta}_j < \hat{\lambda}_j$, $j = 1, ..., m$
(vi) Each interval $(\hat{\theta}_{j+1}, \hat{\theta}_j)$ contains at least one eigenvalue $\hat{\lambda}_{j+1}$, $j = 1, ..., m$.
(vii) Let us denote $\hat{Q}_0^{(2)}$, $\hat{Q}_1^{(2)}$, ..., $\hat{Q}_n^{(2)}$ the unique sequence of orthogonal polynomials with respect to the polynomial norm $[.,.]_{(1)}$ and with constant term equal to 1. This sequence of polynomials satisfies properties (i) and (ii) above with $(\hat{\theta}_j^{(2)})_{1\leq j\leq m}$ denoting the roots of $\hat{Q}_m^{(2)}$ such that $\hat{\theta}_1^{(2)} > \hat{\theta}_2^{(2)} > ... > \hat{\theta}_m^{(1)}$. Then it holds that $\hat{\theta}_m \leq \hat{\theta}_m^{(2)}$ and the Christoffel - Darboux identity is given by

$$0 \leq \hat{Q}'_{m-1}(0) - \hat{Q}'_m(0) = \frac{[\hat{Q}_{m-1}, \hat{Q}_{m-1}] - [\hat{Q}_m, \hat{Q}_m]}{[\hat{Q}^{(2)}_{m-1}, \hat{Q}^{(2)}_{m-1}]_{(1)}} \leq \frac{[\hat{Q}_{m-1}, \hat{Q}_{m-1}]}{[\hat{Q}^{(2)}_{m-1}, \hat{Q}^{(2)}_{m-1}]_{(1)}}. \tag{1.41}$$

(viii) $|\widehat{Q}'_m(0)|^{-1} \leq \hat{\theta}_m$, where $\hat{\theta}_m$ is the smallest root of $\hat{Q}_m$.

Proof of Lemma 1.

For the proofs of results (i), (ii), (iii), (vii), see (i), (ii), (iii) and (iv) in Lemma 5.2 of Blanchard and Krämer (2016). For the proofs of (iv), (v), and (vi), see the Theorem 2 of Lingjaerde and Christophersen (2000).

Proof of (viii) : Given $\hat{Q}_m(0) = 1$, we can write $\hat{Q}_m(t)$ as

$$\hat{Q}_m(t) = \prod_{j=1}^{m} \left(1 - \frac{t}{\hat{\theta}_{j,m}}\right),$$

where t is an arbitrary constant and $\hat{\theta}_{j,m}$ are the roots of $\hat{Q}_m(t)$ ordered in decreasing order $\hat{\theta}_{m,m} < \hat{\theta}_{m-1,m} < ... < \hat{\theta}_{1,m}$. Then, we can see that

$$|\widehat{Q}'_m(0)| = \prod_{j=1}^{m} \frac{1}{\hat{\theta}_{j,m}} \geq \frac{1}{\hat{\theta}_{m,m}}.$$

Hence $\widehat{Q}'_m(0)^{-1} \leq \hat{\theta}_{m,m}$. The result follows by denoting $\hat{\theta}_{m,m}$ as $\hat{\theta}_m$ to simplify the notations.

Lemma 2.

If $(X_i, Y_i)_{i=1...n}$ *are i.i.d with* $\int_0^1 E[X^2(s)]ds < +\infty$ *and* $(\varepsilon_i)_{i=1...n}$ *respect the homoskedasticity and the exogeneity conditions, then*

$$E\left[||T^*_n(\underline{\varepsilon})||^2\right] = \frac{A_0}{n},$$

where $A_0 = \sigma^2 \int_S \mathbb{E}[X_i^2(t)]dt$.

Proof of Lemma 2.

$$
\begin{aligned}
E\left[||T_n^*(\underline{\varepsilon})||^2\right] &= E\left[||\frac{1}{n}\sum_{i=1}^{n} X_i\varepsilon_i||^2\right] \\
&= E\left[\frac{1}{n^2}\int_S \left(\sum_{i=1}^{n} X_i(s)\varepsilon_i\right)^2 ds\right] \\
&= E\left[\frac{1}{n^2}\int_S \left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i(s)X_j(s)\varepsilon_i\varepsilon_j\right) ds\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_S E\left[X_i(s)X_j(s)\varepsilon_i\varepsilon_j\right] ds \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\int_S E\left[X_i^2(s)\varepsilon_i^2\right] ds \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2\int_S E[X_i^2(s)] ds \\
&= \frac{\sigma^2}{n^2}\sum_{i=1}^{n}\int_S E[X_i^2(s)] ds \\
&= \frac{A_0}{n},
\end{aligned}
$$

where $A_0 = \sigma^2 \int_S E[X_i^2(s)]ds < +\infty$. Then,

$$
E\left[||T_n^*(\underline{\varepsilon})||^2\right] = \frac{A_0}{n}.
$$

Passing from the third to the fourth line is possible since we have $\int_0^1 E[X^2(s)]ds < +\infty$. The fifth line is possible because of the exogeneity condition. The sixth line is true following the homoskedasticity condition.

Lemma 3. For $m > 0$, we have

$$
\left\|T_n^*\underline{Y} - T_n^*T_n\widehat{\beta}_m^{PLS}\right\| \leq \frac{\sqrt{A_0}}{\sqrt{n}} + C\widehat{Q}'_m(0)^{-\frac{\mu+2}{2}}.
$$

Proof of Lemma 3.
We have

$$
\begin{aligned}
T_n^*\underline{Y} - T_n^*T_n\widehat{\beta}_m^{PLS} &= T_n^*\underline{Y} - T_n^*T_nP_m\left(T_n^*T_n\right)T_n^*\underline{Y} \\
&= \widehat{Q}_m\left(T_n^*T_n\right)T_n^*\underline{Y}.
\end{aligned}
$$

Following Lemma 1(iii) where the function φ_m has been defined (see also Blanchard and

Krämer (2016), Lemma 5.2), we have

$$\begin{aligned}\left\|\widehat{Q}_m\left(T_n^*T_n\right)T_n^*\underline{Y}\right\| &\leq \left\|\Pi_{\widehat{\theta}_m}\varphi_m\left(\widehat{K}\right)T_n^*\underline{Y}\right\| \\ &\leq \left\|\Pi_{\widehat{\theta}_m}\varphi_m\left(\widehat{K}\right)T_n^*T_n\beta\right\| + \left\|\Pi_{\widehat{\theta}_m}\varphi_m\left(\widehat{K}\right)T_n^*\underline{\varepsilon}\right\|.\end{aligned}$$

As $0<\varphi(\lambda)\leq 1$ for $0<\lambda<\widehat{\theta}_m$, we have

$$\left\|\Pi_{\widehat{\theta}_m}\varphi_m\left(\widehat{K}\right)T_n^*\underline{\varepsilon}\right\| \leq \|T_n^*\underline{\varepsilon}\| = \frac{\sqrt{A_0}}{\sqrt{n}}.$$

On the other hand,

$$\begin{aligned}\left\|\Pi_{\widehat{\theta}_m}\varphi_m\left(\widehat{K}\right)T_n^*T_n\beta\right\| &= \left\|\Pi_{\widehat{\theta}_m}\varphi_m\left(\widehat{K}\right)T_n^*T_n\left(T^*T\right)^{\mu/2}w\right\| \\ &= \left\|\Pi_{\widehat{\theta}_m}\varphi_m\left(\widehat{K}\right)T_n^*T_n\left(T_n^*T_n\right)^{\mu/2}w\right\| + \frac{\sqrt{A_0}}{\sqrt{n}}.\end{aligned}$$

Moreover, by Lemma 1(iii), we have

$$\begin{aligned}\left\|\Pi_{\widehat{\theta}_m}\varphi_m\left(\widehat{K}\right)T_n^*T_n\left(T_n^*T_n\right)^{\mu/2}w\right\| &\leq \sup_{t<\widehat{\theta}_m} t^{\frac{\mu}{2}+1}\varphi_m(t)\,\|w\| \\ &\leq C\widehat{Q}'_m(0)^{-\frac{\mu+2}{2}}.\end{aligned}$$

Where C is a positive constant. Therefore, the result follows.

Lemma 4.

Under the conditions A1-A4, if the stopping rule holds, then

$$|\hat{Q}'_m(0)| \leq Cn^{1/(\mu+2)}. \tag{1.42}$$

Proof of lemma 4.

We can set without loss of generality, $m\geq 1$, since $\hat{Q}'_0(x)=0$. By the stopping rule and Lemma 3, we have

$$\frac{\tau\sqrt{A_0}}{\sqrt{n}} < ||T_n^*(\underline{Y}-T_n(\hat{\beta}_{m-1}))|| \leq \frac{\sqrt{A_0}}{\sqrt{n}} + C|\hat{Q}'_{m-1}(0)|^{-\frac{\mu+2}{2}}, \tag{1.43}$$

where C is a positive constant. For $m=1$, the inequality (1.42) always holds since the right hand side of the previous equation is $+\infty$.
For $m>1$, Equation (1.43) implies

$$|\hat{Q}'_{m-1}(0)| \leq Cn^{1/(\mu+2)}$$

for some positive constant C.

We have

$$\hat{Q}_m^{'}(0) = \hat{Q}_m^{'}(0) - \hat{Q}_{m-1}^{'}(0) + \hat{Q}_{m-1}^{'}(0),$$

then

$$|\hat{Q}_m^{'}(0)| \leq |\hat{Q}_m^{'}(0) - \hat{Q}_{m-1}^{'}(0)| + |\hat{Q}_{m-1}^{'}(0)|.$$

Therefore, it is sufficient to prove that $|\hat{Q}_m^{'}(0) - \hat{Q}_{m-1}^{'}(0)| \leq Cn^{1/(\mu+2)}$ for some positive constant C.

According to Equation (22) of Lemma 1 of Blanchard and Krämer (2016), we have

$$|\hat{Q}_m^{'}(0) - \hat{Q}_{m-1}^{'}(0)| \leq \frac{[\hat{Q}_{m-1}, \hat{Q}_{m-1}]}{[\hat{Q}_{m-1}^{(2)}, \hat{Q}_{m-1}^{(2)}]_{(1)}}$$

Let a be such that $0 < a < \hat{\theta}_{m-1}$, then we have

$$\begin{aligned}
[\hat{Q}_{m-1}, \hat{Q}_{m-1}] &= ||\hat{Q}_{m-1}(\hat{K})T_n^*\underline{Y}||^2 \\
&\leq ||\hat{Q}_{m-1}^{(2)}(\hat{K})T_n^*\underline{Y}||^2 \\
&\leq 2||\hat{\Pi}_a\hat{Q}_{m-1}^{(2)}(\hat{K})T_n^*\underline{Y}||^2 + 2||(I-\hat{\Pi}_a)\hat{K}^{-1/2}\hat{K}^{1/2}\hat{Q}_{m-1}^{(2)}(\hat{K})T_n^*\underline{Y}||^2 \\
&\leq 2 \underset{t\in[0,a]}{sup}\{\hat{Q}_{m-1}^{(2)}(t)\}||\hat{\Pi}_a T_n^*\underline{Y}||^2 + 2\frac{1}{a}||(I-\hat{\Pi}_a)\hat{K}^{1/2}\hat{Q}_{m-1}^{(2)}(\hat{K})T_n^*\underline{Y}||^2 \\
&\leq 2||\hat{\Pi}_a T_n^*\underline{Y}||^2 + \frac{1}{a}[\hat{Q}_{m-1}^{(2)}, \hat{Q}_{m-1}^{(2)}]_{(1)}.
\end{aligned}$$

The second line holds since that $\hat{Q}_{m-1}$ is the unique polynomial of at most $m-1$ degrees with constant 1 that minimizes the residual in the class of all the polynomials of the same degree. For the fourth and fifth lines, we have used the fact $|\hat{Q}_{m-1}^{(2)}(t)| < 1$ for $t \in [0, a]$ and a is such that $0 < a < \hat{\theta}_{m-1}$. The first term on the right hand side of the last inequality can be bounded as follows

$$
\begin{aligned}
||\hat{\Pi}_a T_n^* \underline{Y}|| &\leq ||\hat{\Pi}_a T_n^* (\underline{Y} - T_n \beta)|| + ||\hat{\Pi}_a T_n^* T_n \beta|| \\
&\leq ||\hat{\Pi}_a T_n^* \underline{\varepsilon}|| + ||\hat{\Pi}_a \hat{K} K^{\mu/2} w|| \\
&\leq ||\hat{\Pi}_a T_n^* \underline{\varepsilon}|| + ||\hat{\Pi}_a \hat{K} \hat{K}^{\mu/2} w|| + ||\hat{\Pi}_a \hat{K} (K^{\mu/2} - \hat{K}^{\mu/2}) w|| \\
&\leq \frac{\sqrt{A_0}}{\sqrt{n}} + \left\{ \underset{t \in [0,a]}{Sup} \{ |t^{\frac{\mu}{2}+1}| \} \right\} ||w|| + \left\{ \underset{t \in [0,a]}{Sup} \{ |t| \} \right\} ||K^{\mu/2} - \hat{K}^{\mu/2}||_{op} ||w|| \\
&\leq \frac{\sqrt{A_0}}{\sqrt{n}} + a^{\frac{\mu}{2}+1} ||w|| + \frac{C_3 a}{\sqrt{n}} ||w|| \\
&\leq \frac{\sqrt{A_0}}{\sqrt{n}} + a^{\frac{\mu}{2}+1} ||w||,
\end{aligned}
$$

where C_3 is some positive constant. Let us choose a such that $a = \left(\frac{\sqrt{A_0} C_a}{||w|| \sqrt{n}} \right)^{2/(\mu+2)}$ with C_a sufficiently small such that $0 < C_a < \tau - 1$ to ensure that $0 < a < |\hat{Q}'_{m-1}(0)|^{-1} < \hat{\theta}_{m-1}$. By the stopping rule in Equation (1.29), we have $\left(\frac{\tau \sqrt{A_0}}{\sqrt{n}} \right)^2 \leq [\hat{Q}_{m-1}, \hat{Q}_{m-1}]$. Then we obtain

$$
\begin{aligned}
||\hat{\Pi}_a T_n^* \underline{Y}|| &\leq \frac{\sqrt{A_0}}{\sqrt{n}} + \left(\frac{C_a \sqrt{A_0}}{||w|| \sqrt{n}} \right) ||w|| \\
&\leq \frac{1}{\tau} \Big(1 + C_a \Big) [\hat{Q}_{m-1}, \hat{Q}_{m-1}]^{1/2}.
\end{aligned}
$$

Therefore,

$$
[\hat{Q}_{m-1}, \hat{Q}_{m-1}] \leq \frac{1}{\tau^2} \Big(1 + C_a \Big)^2 [\hat{Q}_{m-1}, \hat{Q}_{m-1}] + a^{-1} [\hat{Q}^{(2)}_{m-1}, \hat{Q}^{(2)}_{m-1}]_{(1)}.
$$

Then,

$$
\left(1 - \frac{(1 + C_a)^2}{\tau^2} \right) [\hat{Q}_{m-1}, \hat{Q}_{m-1}] \leq a^{-1} [\hat{Q}^{(2)}_{m-1}, \hat{Q}^{(2)}_{m-1}]_{(1)}.
$$

This implies with our choice of a,

$$
\begin{aligned}
\frac{[\hat{Q}_{m-1}, \hat{Q}_{m-1}]}{[\hat{Q}^{(2)}_{m-1}, \hat{Q}^{(2)}_{m-1}]_{(1)}} &\leq C_1 a^{-1}. \\
&\leq C_0 n^{1/(\mu+2)}.
\end{aligned}
$$

where $C_0 > 0$ is some constant. Then we obtain

$$|\hat{Q}'_m(0) - \hat{Q}'_{m-1}(0)| \leq C_0 n^{1/(\mu+2)}.$$

This concludes the proof.

Proof of Theorem 3.

Proof of the second result of Theorem 3.

Let us denote by

$$\Pi_a \phi = \sum_{j=1, \lambda_j < a}^{\infty} < \phi, v_j > v_j,$$

and

$$\hat{\Pi}_a \phi = \sum_{j=1, \hat{\lambda}_j < a}^{\infty} < \phi, \hat{v}_j > \hat{v}_j$$

the orthogonal projection onto the eigenvectors of the covariance operator K (respectively $\hat{K}$) for which the corresponding eigenvalues λ_j ($\hat{\lambda}_j$) are smaller than a, where a is a positive number. In the sequel, we choose a such that $0 < a \leq |\hat{Q}'_m(0)|^{-1}$. Let us consider the following function $\tilde{\beta}_m^{PLS}$ defined by

$$\tilde{\beta}_m^{PLS} = \hat{P}_m(\hat{K})\hat{K}\beta$$

We have

$$\begin{aligned} ||\hat{\beta}_m^{PLS} - \beta|| &\leq ||\hat{\Pi}_a(\hat{\beta}_m^{PLS} - \beta)|| + ||(I - \hat{\Pi}_a)(\hat{\beta}_m^{PLS} - \beta)|| \\ &\leq ||\hat{\Pi}_a(\hat{\beta}_m^{PLS} - \tilde{\beta}_m^{PLS})|| + ||\hat{\Pi}_a(\tilde{\beta}_m^{PLS} - \beta)|| + ||(I - \hat{\Pi}_a)(\hat{\beta}_m^{PLS} - \beta)||. \end{aligned}$$

where $(I - \hat{\Pi}_a)$ is the orthogonal projection on the space spanned by $\{\hat{v}_j; \hat{\lambda}_j > a\}$. Let us define (I), (II) and (III) by $(I) = ||\hat{\Pi}_a(\hat{\beta}_{PLS} - \tilde{\beta}_{PLS})||$, $(II) = ||\hat{\Pi}_a(\tilde{\beta}_{PLS} - \beta)||$ and $(III) = ||(I - \hat{\Pi}_a)(\hat{\beta}_{PLS} - \beta)||$ respectively.

We will derive the upper bound rate of the three terms (I), (II), and (III).

Upper bound rate of (I) :

We have

$$\begin{aligned} (I) &= ||\hat{\Pi}_a(\hat{\beta}_m^{PLS} - \tilde{\beta}_m^{PLS})|| \\ &= ||\hat{\Pi}_a \hat{P}_m \hat{K} T_n^* \underline{Y} - \hat{P}_m(\hat{K})\hat{K}\beta|| \\ &= ||\hat{\Pi}_a \{\hat{P}_m(\hat{K}) T_n^* \underline{\varepsilon}\}|| \\ &\leq \left|\left|\hat{\Pi}_a \hat{P}_m(\hat{K})\right|\right| ||T_n^* \underline{\varepsilon}||. \end{aligned}$$

with $\|T_n^* \underline{\varepsilon}\| = O_p\left(\frac{\sqrt{A_0}}{\sqrt{n}}\right)$ and

$$\left\|\hat{\Pi}_a \hat{P}_m(\hat{K})\right\| \leq \sup_{t\in[0,a]} \left|\hat{P}_m(t)\right|.$$

Since $|\widehat{Q}_m'(0)|^{-1} \leq \hat{\theta}_m$, we have $a \leq \hat{\theta}_m$. As $\widehat{Q}_m$ is convex on [0,a], we have

$$\hat{P}_m(t) = \frac{1-\widehat{Q}_m(t)}{t} \leq -\widehat{Q}_m'(0).$$

On the other hand, we have

$$|\hat{Q}_m'(0)| = |Q_m'(0)| + o_p\left(\frac{1}{\sqrt{n}}\right)$$

which follows from the mean-value theorem since $\hat{Q}_m'$ is a continuously differentiable function of $\hat{K}$ and $T_n^* \underline{Y}$, which are consistent estimators of K and $K\beta$. Then,

$$(I) = O_p\left(\frac{1}{\sqrt{n}}|Q_m'(0)|\right).$$

Upper bound rate of (II) :
We have

$$\begin{aligned}
(II) &= ||\hat{\Pi}_a(\tilde{\beta}_m^{PLS} - \beta)|| \\
&= ||\hat{\Pi}_a\{\hat{P}_m(\hat{K})\hat{K}\beta - \beta\}|| \\
&= ||\hat{\Pi}_a\hat{Q}_m(\hat{K})\beta|| \\
&= ||\hat{\Pi}_a\hat{Q}_m(\hat{K})K^{\mu/2}w|| \\
&= ||\hat{\Pi}_a\hat{Q}_m(\hat{K})[K^{\mu/2} - \hat{K}^{\mu/2} + \hat{K}^{\mu/2}](w)|| \\
&\leq ||\hat{\Pi}_a\hat{Q}_m(\hat{K})\hat{K}^{\mu/2}w|| + ||\hat{\Pi}_a\hat{Q}_m(\hat{K})[K^{\mu/2} - \hat{K}^{\mu/2}]w|| \\
&\leq \left\{\underset{t\in[0,a]}{Sup}|t^{\mu/2}\hat{Q}_m(t)|\right\}||w|| + \left\{\underset{t\in[0,a]}{Sup}|\hat{Q}_m(t)|\right\}||K^{\mu/2} - \hat{K}^{\mu/2}||_{op}||w|| \\
&= O_p\left(a^{\mu/2}\right) + O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

This is possible given that $0 \leq \underset{t\in[0,a]}{Sup}|t^{\mu/2}\hat{Q}_m(t)| \leq a^{\mu/2}$ see (ii) of Lemma 1. Moreover, $||w|| < +\infty$, and $||K^{\mu/2} - \hat{K}^{\mu/2}||_{op} = O_p\left(\frac{1}{\sqrt{n}}\right)$, then

$$(II) = O_p\left(a^{\mu/2}\right) + O_p\left(\frac{1}{\sqrt{n}}\right).$$

Upper bound rate of (III) :

We have

$$\begin{aligned}
(III) &= ||(I-\hat{\Pi}_a)(\hat{\beta}_m^{PLS}-\beta)|| \\
&= ||(I-\hat{\Pi}_a)\hat{K}^+\hat{K}\hat{P}_m(\hat{K})(T_n^*\underline{Y}-\beta)|| \\
&\leq ||(I-\hat{\Pi}_a)\hat{K}^+||||(I-\hat{\Pi}_a)\hat{K}\hat{P}_m(\hat{K})(T_n^*\underline{Y}-\beta)|| \\
&\leq \frac{1}{a}||(I-\hat{\Pi}_a)\hat{K}\hat{P}_m(\hat{K})(T_n^*\underline{Y}-\hat{K}\beta)|| \\
&\leq \frac{1}{a}||T_n^*\hat{W}_n\hat{P}_m(W_n)(\underline{Y}-T_n\beta)|| \\
&\leq \frac{1}{a}||T_n^*\hat{W}_n(\hat{P}_m(W_n)\underline{Y}-T_n\beta)|| \\
&\leq \frac{1}{a}||T_n^*(\hat{W}_n\hat{P}_m(W_n)\underline{Y}-\underline{Y}+\underline{Y}-T_n\beta)|| \\
&\leq \frac{1}{a}||T_n^*\hat{Q}_m(W_n)\underline{Y}||+\frac{1}{a}||T_n^*(\underline{Y}-T_n\beta)|| \\
&\leq \frac{1}{a}\left|\left|T_n^*(\underline{Y}-T_n\hat{\beta}_m^{PLS})\right|\right|+\frac{1}{a}||T_n^*\underline{\varepsilon}|| \\
&\leq \frac{2C}{a}\Lambda_m.
\end{aligned}$$

where $\hat{K}^+$ is the generalized inverse of $\hat{K}$ using the PCA regularization, $\Lambda_m = max\{||T_n^*(\underline{Y}-T_n\hat{\beta}_m)||, \frac{\sqrt{A_0}}{\sqrt{n}}\}$ and $\hat{K} = T_n^*T_n$ and $W_n = T_nT_n^*$. The fourth and fifth line are possible given that $\hat{K} = T_n^*T_n$ and $W_n = T_nT_n^*$. The eighth line holds since $\hat{Q}_m(W_n)\underline{Y} = \hat{W}_n\hat{P}_mW_n\underline{Y}-\underline{Y}$ and $\underline{\varepsilon} = \underline{Y}-T_n\beta$. The last line comes from the fact that $||T_n^*\underline{\varepsilon}|| = \frac{\sqrt{A_0}}{\sqrt{n}}$.

Combining the upper bound from (I), (II), and (III) leads to the following result

$$||\hat{\beta}_m^{PLS}-\beta|| = O_p\left(a^{\mu/2}\right)+O_p\left(\frac{1}{\sqrt{n}}|\hat{Q}_m'(0)|\right)+O_p\left(\frac{2}{a}\Lambda_m\right).$$

Now, let us consider the function g defined as

$$g(a) = a^{\mu/2}+\frac{2}{a}\Lambda_m+\frac{1}{\sqrt{n}}|\hat{Q}_m'(0)|$$

Then, g is differentiable and admit a minimum at a value a_* such that $g'(a_*) = 0$. This leads to ${a_*}^{\frac{\mu+2}{2}} = \frac{\Lambda_m}{\mu}$. Then $a_* = \left(4\frac{\Lambda_m}{\mu}\right)^{2/(\mu+2)}$.

Since $a \in (0, |\hat{Q}_m'(0)|^{-1}]$, we need to consider 2 cases : $a_* < |\hat{Q}_m'(0)|^{-1}$ and $a_* > |\hat{Q}_m'(0)|^{-1}$.

Case 1 : let $a_* < |\hat{Q}_m'(0)|^{-1}$

Then, the minimum is attained at a_* and we have

$$||\hat{\beta}_m^{PLS} - \beta|| \leq C\left(a_*^{\mu/2} + \frac{2\Lambda_m}{a_*} + \frac{1}{\sqrt{n}}|\hat{Q}'_m(0)|\right)$$
$$\leq C\left(\Lambda_m^{\mu/(\mu+2)} + \Lambda_m|\hat{Q}'_m(0)|\right).$$

Case 2 : let $a_* > |\hat{Q}'_m(0)|^{-1}$
Then, the minimum is attained at $|\hat{Q}'_m(0)|^{-1}$ and we have

$$||\hat{\beta}_m^{PLS} - \beta|| \leq C\left[|\hat{Q}'_m(0)|^{-\mu/2} + 2|\hat{Q}'_m(0)|\Lambda_m + \frac{1}{\sqrt{n}}|\hat{Q}'_m(0)|\right]$$
$$\leq C\left[a_*^{\mu/2} + 2|\hat{Q}'_m(0)|\Lambda_m + \Lambda_m|\hat{Q}'_m(0)|\right]$$
$$\leq C\left[a_*^{\mu/2} + 3|\hat{Q}'_m(0)|\Lambda_m\right]$$
$$\leq C\left[\Lambda_m^{\mu/(\mu+2)} + 3|\hat{Q}'_m(0)|\Lambda_m\right].$$

Then, considering both cases, we obtain

$$||\hat{\beta}_m^{PLS} - \beta|| \leq C\left[\Lambda_m^{\mu/(\mu+2)} + 3|\hat{Q}'_m(0)|\Lambda_m\right]$$

Till now, we have not used the stopping rule. Let us denote $m\,(n)$ the value of m corresponding to
$\left\|T_n^*\left(\underline{Y} - T_n\widehat{\beta}_{m(n)}^{PLS}\right)\right\| \leq \frac{\tau\sqrt{A_0}}{\sqrt{n}} \leq \left\|T_n^*\left(\underline{Y} - T_n\widehat{\beta}_{m(n)-1}^{PLS}\right)\right\|$. By the definition of Λ_m, we have $\Lambda_{m(n)} \leq \frac{\tau\sqrt{A_0}}{\sqrt{n}}$.

Moreover by Lemma 4, we have

$$\left|Q'_{m(n)}\,(0)\right| \leq \sqrt{n}^{\frac{2}{\mu+2}}.$$

Then, we obtain the result

$$||\hat{\beta}_m^{PLS} - \beta|| = O_p\left(\sqrt{n}^{-\frac{\mu}{\mu+2}}\right).$$

Proof of the first result of Theorem 3.
For each a such that $0 < a \leq |Q'_m(0)|^{-1} \leq \theta_m$, and using the alternative upper bound of (II) we obtain

$$||\hat{\beta}_m^{PLS} - \beta||^2 = O_p\left(||\hat{\Pi}_a(\beta)||^2\right) + O_p\left(\frac{1}{n}|Q'_m(0)|^2\right) + O_p\left(\frac{1}{a^2}\Lambda_m^2\right).$$

by taking $a = |Q_m^{'}(0)|^{-1}$, we have

$$\begin{aligned}
||\hat{\beta}_m^{PLS} - \beta||^2 &= O_p\left(||\hat{\Pi}_{|Q_m^{'}(0)|^{-1}}(\beta)||^2\right) + O_p\left(\frac{1}{n}|Q_m^{'}(0)|^2\right) + O_p\left(\frac{1}{|Q_m^{'}(0)|^{-2}}\Lambda_m^2\right) \\
&= O_p\left(||\hat{\Pi}_{|Q_m^{'}(0)|^{-1}}(\beta)||^2\right) + O_p\left(\frac{1}{n}|Q_m^{'}(0)|^2\right) + O_p\left(\frac{1}{n}|Q_m^{'}(0)|^2\right) \\
&= O_p\left(\sum_{\lambda_j<\theta_m}^{\infty} < \beta, v_j >^2\right) + O_p\left(\frac{1}{n}\left(\sum_{l=1}^{m}\frac{1}{\theta_l}\right)^2\right) \\
&= O_p\left(\sum_{j=m+1}^{\infty} < \beta, v_j >^2\right) + O_p\left(\frac{m^2}{n\theta_m^2}\right) \\
&= O_p\left(\lambda_{m+1}^{\gamma}\right) + O_p\left(\frac{m^2}{n\theta_m^2}\right).
\end{aligned}$$

The first term is the bias term and the second one is the variance. The second line hold since $0 < a \leq |Q_m^{'}(0)|^{-1} \leq \theta_m$ and $|Q_m^{'}(0)| = \sum_{l=1}^{m}\frac{1}{\theta_l}$. Also, the third term of the second line holds since following the stopping rule, we have $\Lambda_m^2 \leq \frac{A_0\tau}{n}$. The third line hold following that $\hat{\lambda}_{m+1} \leq \hat{\theta}_m$. The last line is possible given that $\sum_{j=m+1}^{\infty} < \beta, v_j >^2 = ||\beta_m^{PCA} - \beta||^2 = O_p\left(\lambda_{m+1}^{\gamma}\right)$, combined with the fact that λ_{m+1} is the largest eigenvalue of K that is smaller than θ_m (see (iii) of Lemma 1). Therefore, $||\beta_m^{PLS} - \beta||^2 = O_p\left(\lambda_{m+1}^{\gamma}\right)$.

Proof of Theorem 4.
We have $\beta = K^{\mu/2}w$ with $||w|| < \infty$. Let us consider the Jacobi polynomial defined as

$$\psi_m(t) = Q_m^{(\frac{\mu}{2}+1)}\left(\frac{t}{\hat{\lambda}_1}\right)$$

where $Q_m^{(\frac{\mu}{2}+1)}$ is the residual polynomial related to the considered functional linear model with the norm $[.,.]_{(\frac{\mu}{2}+1)}$. Following theorem 6.12 of Engl et al. (1996), ψ_m is the m-th residual polynomial with the norm $[.,.]_{(\frac{\mu}{2}+1)}$ and we have $|\psi_m(t)| < 1$ and $|t^{\frac{\mu}{2}+1}\psi_m(t)| \leq cm^{-\mu-2}$.

For each t such that $0 < t < \hat{\lambda}_1$,

$$
\begin{aligned}
||T_n^*(\underline{Y} - T_n(\hat{\beta}_m))|| &= ||\hat{Q}_m(\hat{K})T_n^*\underline{Y}|| \\
&\leq ||\psi_m(\hat{K})T_n^*\underline{Y}|| \\
&\leq ||\psi_m(\hat{K})T_n^*(\underline{Y} - T_n^*\beta)|| + ||\psi_m(\hat{K})T_n^*T_n\beta|| \\
&\leq \left\{ \underset{t}{Sup}|\psi_m(t)| \right\} ||T_n^*\underline{\varepsilon}|| + ||\psi_m(\hat{K})\hat{K}K^{\mu/2}w|| \\
&\leq \left\{ \underset{t}{Sup}|\psi_m(t)| \right\} ||T_n^*\underline{\varepsilon}|| + ||\psi_m(\hat{K})\hat{K}(\hat{K}^{\mu/2}(w))|| + ||\psi_m(\hat{K})\hat{K}[\hat{K}^{\mu/2} - K^{\mu/2}]w|| \\
&\leq \left\{ \underset{t}{Sup}|\psi_m(t)| \right\} ||T_n^*\underline{\varepsilon}|| + \left\{ \underset{t}{Sup}|t^{\frac{\mu}{2}+1}\psi_m(t)| \right\} ||w|| \\
&+ \left\{ \underset{t}{Sup}|t\psi_m(t)| \right\} ||\hat{K}^{\mu/2} - K^{\mu/2}||_{op}||w||.
\end{aligned}
$$

We have $||T_n^*\underline{\varepsilon}|| = \frac{\sqrt{A_0}}{\sqrt{n}}$, $\underset{t}{Sup}|t^{\frac{\mu}{2}+1}\psi_m(t)| < c_1 m^{-\mu-2}$, $\underset{t}{Sup}|t\psi_m(t)| < c_2 m^{-2}$ and $|\psi_m(t)| < 1$ with $c_1, c_2 > 0$. Then,

$$
||T_n^*(\underline{Y} - T_n\hat{\beta}_m)|| \leq \left[\frac{\sqrt{A_0}}{\sqrt{n}} + c_1 m^{-\mu-2} + c_2 \frac{m^{-2}}{\sqrt{n}} \right].
$$

Then $||T_n^*(\underline{Y} - T_n\hat{\beta}_m)|| \leq \frac{\tau\sqrt{A_0}}{\sqrt{n}}$ if $\left[\frac{1}{\sqrt{n}} + c_1 m^{-\mu-2} + c_2 \frac{m^{-2}}{\sqrt{n}} \right] \leq \frac{\tau\sqrt{A_0}}{\sqrt{n}}$, which leads to $c_1 m^{-\mu-2} \leq \frac{(\tau-1)\sqrt{A_0}}{\sqrt{n}}$ and hence, $m_{PLS} \leq cn^{1/(2\mu+4)}$ for some $c > 0$.

Now we want to show the inequality $m_{PLS} \geq cn^{1/(2\mu+4)}$ with $c > 0$. Let us consider $\nu > 0$ be an arbitrary real value such that $\beta = K^{(\mu+\nu)/2}(w)$ for all $\nu > 0$. Assume that $\underline{Y}_\nu$ and X are chosen such that

$$
\begin{aligned}
||T_n^*(\underline{Y} - T_n\hat{\beta}_m^{PLS})||_w &= [\hat{Q}_m, \hat{Q}_m]_{(\frac{\mu}{2}+\frac{1}{2}+\frac{\nu}{2})}. \\
&= \int_S \hat{Q}_m^2(t) w_{\frac{\mu}{2}+\frac{1}{2}+\frac{\nu}{2}}(t)dt.
\end{aligned}
$$

is minimized among all possible residual polynomials of degree m with $w_{\frac{\mu}{2}+\frac{1}{2}+\frac{\nu}{2}}(t) = t^{2(\frac{\mu}{2}+\frac{\nu}{2})}(1-t)^{-1/2}$. Then $||T_n^*(\underline{Y} - T_n(\hat{\beta}_m^{PLS}))||^2$ is given by the m-th Christoffel function associated with the Jacobi weight $w_{\frac{\mu}{2}+\frac{1}{2}+\frac{\nu}{2}}(t)$ (see Equation (A.20) by Engl et al. (1996)). Following Equation (A.11) by Engl et al. (1996), we have

$$
||T_n^*(\underline{Y} - T_n(\hat{\beta}_m^{PLS}))||_{w_{\frac{\mu}{2}+\frac{1}{2}+\frac{\nu}{2}}} = \Lambda_m(0, w_{\frac{\mu}{2}+\frac{1}{2}+\frac{\nu}{2}}) \sim m^{-(\mu+\nu)-2}.
$$

Then, the stopping rule is satisfied if

$$||T_n^*(\underline{Y} - T_n(\hat{\beta}_m^{PLS}))||_{w_{\frac{\mu}{2}+\frac{1}{2}+\frac{\nu}{2}}} \leq \frac{\tau\sqrt{A_0}}{\sqrt{n}}.$$

This implies that $m^{-(\mu+\nu)-2} \sim \frac{\tau\sqrt{A_0}}{\sqrt{n}}$ and therefore $m \sim n^{1/(2(\mu+\nu)+4)}$ for all $\nu > 0$. Then, for $\nu \to 0$ we obtain $m \sim Cn^{1/(2\mu+4)}$, with $C > 0$ being some constant. Hence, the best possible uniform bound for m is $n^{1/(2\mu+4)}$ if there is no extra information about the configuration of the eigensystem of the covariance operator K.

Proof of Corollary 1.

Concerning FPCA :

We have

$$\mathbb{E}\left[\left(\hat{f}_m^{PCA} - f\right)^2 |X\right] = O_p\left(\lambda_{m+1}^{\mu}\right) + O_p\left(\frac{m}{n\lambda_m}\right).$$

If $\lambda_j = O(j^{-2\gamma})$ with $\gamma > 0$:

Then, when equating the squared bias and the variance term, we obtain

$$m^{-2\mu\gamma} \geq (m+1)^{-2\mu\gamma} = \frac{m}{nm^{-2\gamma}}.$$

This leads to

$$m_{pca} \leq n^{1/(2\gamma(\mu+1)+1)}.$$

If $\lambda_j = O(d^j)$ with $0 < d < 1$:

Then, when equating the squared bias and the variance term, we obtain

$$d^{-\mu m} \leq d^{-\mu(m+1)} = \frac{m}{nd^{-m}}.$$

By taking the log on both sides of the previous equation, we obtain

$$m_{pca} \leq c(1 + log(n)).$$

where $c > 0$ is an arbitrary positive number.

Concerning FPLS :

$\hat{Q}_{j\leq m}$ are the residual polynomial for the FPLS for the defined functional linear model. Let us choose $k \in \mathbb{N}$ such that $0 \leq k \leq m$ and define

$$\Psi_k(\lambda) = \prod_{l=1}^{k}\left(1 - \frac{\lambda}{\lambda_l}\right)$$

and take ψ_{m-k} to be the shifted Jacobi polynomial as

$$\psi_m(\lambda) = G_m^{\frac{\mu}{2}+1}\left(\frac{\lambda}{\hat{\lambda}_l}\right)$$

as

$$G_m(t) = \Psi_k(\lambda)\psi_{m-k}\left(\frac{\hat{\lambda}_1 t}{\hat{\lambda}_{k+1}}\right)$$

G_m is a polynomial of degree m and $G_m(0) = 1$. Then,

$$||T_n^*(\underline{Y} - T_n(\hat{\beta}_m^{PLS}))|| = ||\hat{Q}_m(\hat{K})T_n^*\underline{Y}|| \leq ||G_m(\hat{K})T_n^*\underline{Y}||.$$

As $G_m(t)$ vanishes for all $\hat{\lambda}_{k+1} < t \leq \hat{\lambda}_1$ and $G_m(t) \leq \psi_{m-k}\left(\frac{\hat{\lambda}_1 t}{\hat{\lambda}_{k+1}}\right)$ for $0 \leq t \leq \hat{\lambda}_{k+1}$, we have.

$$\begin{aligned}
||T_n^*(\underline{Y} - T_n\hat{\beta}_m^{PLS})|| &= ||\hat{\Pi}_{\hat{\lambda}_m}G_m(\hat{K})T_n^*\underline{Y}|| \\
&\leq ||\hat{\Pi}_{\hat{\lambda}_{k+1}}G_m(\hat{K})T_n^*\underline{Y}|| \\
&\leq ||\Pi_{\hat{\lambda}_{k+1}}G_m(\hat{K})T_n^*(\underline{Y} - T_n\beta)|| + ||\Pi_{\hat{\lambda}_{k+1}}G_m(\hat{K})(T_n^*T_n\beta)|| \\
&\leq ||\Pi_{\hat{\lambda}_{k+1}}G_m(\hat{K})T_n^*\underline{\varepsilon}|| + ||\Pi_{\hat{\lambda}_{k+1}}G_m(\hat{K})\hat{K}K^{\mu/2}w|| \\
&\leq ||\Pi_{\hat{\lambda}_{k+1}}G_m(\hat{K})T_n^*\underline{\varepsilon}|| + ||\Pi_{\hat{\lambda}_{k+1}}G_m(\hat{K})\hat{K}K^{\mu/2}w|| \\
&+ \left|\left|\Pi_{\hat{\lambda}_{k+1}}G_m(\hat{K})\hat{K}\left[\hat{K}^{\mu/2} - K^{\mu/2}\right]w\right|\right| \\
&\leq \left\{Sup|G_m(t)|\right\}||T_n^*\underline{\varepsilon}|| + \left\{Sup|t^{\frac{\mu}{2}+1}G_m(t)|\right\}||w|| \\
&+ \left\{Sup|t^{\frac{\mu}{2}+1}G_m(t)|\right\}\left|\left|\hat{K}^{\mu/2} - K^{\mu/2}\right|\right|_{op}||w|| \\
&\leq ||T_n^*(\underline{\varepsilon})|| + \underset{t\leq\hat{\lambda}_{k+1}}{Sup}\left|t^{(\frac{\mu}{2}+1)}\psi_{m-k}\left(\frac{\hat{\lambda}_1 t}{\hat{\lambda}_{k+1}}\right)\right| \\
&+ \underset{t\leq\hat{\lambda}_{k+1}}{Sup}\left|t\psi_{m-k}\left(\frac{\hat{\lambda}_1 t}{\hat{\lambda}_{k+1}}\right)\right|\left|\left|\hat{K}^{\mu/2} - K^{\mu/2}\right|\right|_{op}||w|| \\
&\leq \frac{\sqrt{A_0}}{\sqrt{n}} + C_1\hat{\lambda}_{k+1}^{(\frac{\mu+2}{2})}(m-k)^{-\mu-2} + C_2\hat{\lambda}_{k+1}(m-k)^{-2}\frac{1}{\sqrt{n}}.
\end{aligned}$$

where $C_1, C_2 > 0$

If $\lambda_j = O(j^{-2\gamma})$:

We take $k \sim \frac{m}{2}$ and obtain

$$||T_n^*(\underline{Y} - T_n\hat{\beta}_m^{PLS})|| \leq \frac{\sqrt{A_0}}{\sqrt{n}} + C_1 m^{-(\mu+2)(\gamma+1)} + C_2 \frac{m^{-2(\gamma+1)}}{\sqrt{n}}.$$

Then, the stopping rule hold if $C_1 m^{-(2\mu+1)(\gamma+1)} \leq \frac{\tau\sqrt{A_0}}{\sqrt{n}}$ and therefore $m_{pls} \leq cn^{1/(2\mu+4)(\gamma+1)}$, for some $c > 0$.

If $\lambda_j = O(d^j)$ with $0 < d < 1$:

We take $k \sim m$ and obtain

$$||T_n^*(\underline{Y} - T_n(\hat{\beta}_m^{PLS}))|| \leq \frac{\sqrt{A_0}}{\sqrt{n}} + C_1 d^{(m+1)\frac{(\mu+2)}{2}} m^{-(\mu+2)} + C_2 d^{(m+1)} m^{-2} \frac{1}{\sqrt{n}}.$$

Then, using the stopping rule, we obtain $\frac{\sqrt{A_0}}{\sqrt{n}} + C_1 d^{(m+1)\frac{(\mu+2)}{2}} m^{-(\mu+2)} + C_2 d^{(m+1)} m^{-2} \frac{1}{\sqrt{n}} \leq \frac{\tau\sqrt{A_0}}{\sqrt{n}}$. Therefore, $(m+1)log(d) \leq Clog(n)$ and we obtain the result. C, C_0 are arbitrary positive constant.

Proof of proposition 3.

We follow a similar approach to the proof of Theorem 1 in Cai et al. (2012). Remark that any lower bound for a special case is also a lower bound for the general case. Therefore, we consider the special case where $\varepsilon_i|X_i$ is iid $\mathcal{N}(0, \sigma^2)$ and the operator K has eigenvalues $\lambda_j = j^{-1}$. Note that as X is square integrable, the operator K is necessarily trace-class which means that the eigenvalues $\lambda_j = O(1/j)$ when j goes to infinity. By taking $\lambda_j = j^{-1}$, we are taking a limit-case where K is not nuclear any more. It can be thought as the limit of the case $\lambda_j = j^{-1-\epsilon}$ for $\epsilon > 0$ arbitrarily small. Let M be the smallest integer greater than $c_0 n^{\frac{1}{(\mu+2)}}$ where the constant $c_0 > 0$ will be characterized later on. Let γ be a positive constant to be specified below. We construct candidates for β in the following manner. Let $\theta = (\theta_{M+1}, ..., \theta_{2M}) \in \Omega \equiv \{0,1\}^M$ and

$$\begin{aligned} \beta_\theta &= \gamma M^{-1/2} \sum_{j=M+1}^{2M} \theta_j K^{\mu/2} v_j \\ &= \gamma M^{-1/2} \sum_{j=M+1}^{2M} \theta_j \lambda_j^{\mu/2} v_j. \end{aligned}$$

It is easy to see that β_θ satisfies the source condition A4, indeed

$$\begin{aligned} \left\|K^{-\mu/2}\beta_\theta\right\|^2 &= \left\|\gamma M^{-1/2} \sum_{j=M+1}^{2M} \theta_j v_j\right\|^2 \\ &= \frac{\gamma^2}{M} \sum_{j=M+1}^{2M} \theta_j^2 \leq \gamma^2 < \infty \end{aligned}$$

because the eigenfunctions v_j are orthonormal and θ_j are either 1 or 0.

For $\theta, \omega \in \mathbb{R}^M$, we denote $H(\theta, \omega) = \sum_{j=1}^{M} I\{\theta_j \neq \omega_j\}$ the Hamming distance between θ and ω. Note that when $\theta, \omega \in \Omega$, then $H(\theta, \omega) = \sum_{j=1}^{M} (\theta_j - \omega_j)^2$. An important element of the proof is the application of Varshamov-Gilbert bound (see **?**, Lemma 2.9). This lemma states that for any $M \geq 8$, there exists a subset $\Theta = \left\{\theta^{(0)}, ..., \theta^{(N)}\right\}$ of elements of Ω such that (a) $\theta^{(0)} = (0, ...0)$, (b) $H\left(\theta^{(j)}, \theta^{(l)}\right) \geq \frac{M}{8}$, for all $0 \leq j < l \leq N$ and (c) $N > 2^{M/8}$. In the sequel, we will consider only candidates β_θ where $\theta \in \Theta$.

Now, we need to check Conditions (i) and (ii) of Theorem 2.5 of **?**.

(i) The first condition we need to check is the following

$$\|\beta_{\theta'} - \beta_\theta\|^2 \geq 4n^{-\frac{\mu}{\mu+2}} \tag{1.44}$$

for all couples $(\theta', \theta) \in \Theta^2$ such that $\theta \neq \theta'$.

We have

$$\begin{aligned}
\|\beta_{\theta'} - \beta_\theta\|^2 &= \left\|\gamma M^{-1/2} \sum_{j=M+1}^{2M} \left(\theta'_j - \theta_j\right) \lambda_j^{\mu/2} v_j\right\|^2 \\
&= \frac{\gamma^2}{M} \sum_{j=M+1}^{2M} \left(\theta'_j - \theta_j\right)^2 \lambda_j^{\mu} \\
&\geq \gamma^2 \lambda_{2M}^{\mu} \frac{1}{M} H(\theta', \theta) \\
&\geq \frac{\gamma^2}{8} \lambda_{2M}^{\mu} \\
&= \frac{\gamma^2 (2c_0)^{-\mu}}{8} n^{-\frac{\mu}{\mu+2}}.
\end{aligned}$$

Theorefore, by taking $c_0 \leq \frac{1}{2}\left(\frac{\gamma^2}{32}\right)^{\frac{1}{\mu}}$, the inequality (1.44) is satisfied.

(ii) Let P_θ be the joint distribution of (X_i, Y_i), $i = 1, 2, ..., n$ with $\beta = \beta_\theta$. Condition (ii) of Theorem 2.5 of Tsybakov (2009) stipulates that $P_{\theta^{(j)}} \ll P_{\theta^{(o)}}$ for all $j = 1, ..., N$ and

$$\frac{1}{N} \sum_{j=1}^{N} K\left(P_{\theta^{(j)}}, P_{\theta^{(0)}}\right) \leq \alpha \ln N \tag{1.45}$$

where $0 < \alpha < 1/8$ and K is the Kullback divergence.

For $\theta, \theta' \in \Theta$, we have

$$\ln \frac{P_\theta}{P_{\theta'}} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(Y_i - \langle X_i, \beta_\theta \rangle\right) \langle X_i, \beta_\theta - \beta_{\theta'} \rangle - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \langle X_i, \beta_\theta - \beta_{\theta'} \rangle^2.$$

The Kullback divergence between P_θ and $P_{\theta'}$ equals

$$\begin{aligned}
K\left(P_\theta, P_{\theta'}\right) &= \frac{n}{2\sigma^2}\left\|K^{1/2}\left(\beta_\theta - \beta_{\theta'}\right)\right\|^2 \\
&= \frac{n}{2\sigma^2}\left\|\gamma M^{-1/2}\sum_{j=M+1}^{2M}\left(\theta'_j - \theta_j\right)\lambda_j^{(\mu+1)/2} v_j\right\|^2 \\
&= \frac{n}{2\sigma^2}\frac{\gamma^2}{M}\sum_{j=M+1}^{2M}\left(\theta'_j - \theta_j\right)^2\lambda_j^{(\mu+1)} \\
&\leq \frac{n}{2\sigma^2}\gamma^2\lambda_M^{(\mu+1)}\frac{1}{M}\sum_{j=M+1}^{2M}\left(\theta'_j - \theta_j\right)^2 \\
&\leq \frac{n}{2\sigma^2}\gamma^2\lambda_M^{(\mu+1)} \\
&= \frac{n}{2\sigma^2}\gamma^2 M^{-(\mu+1)} \\
&\leq \alpha\ln\left(2^{M/8}\right)
\end{aligned}$$

where the last inequality is obtained by fixing α to an arbitrary value in $\left(0, \frac{1}{8}\right)$ and choosing $c_0 \geq \left(\frac{4\gamma^2}{\alpha\sigma^2\ln 2}\right)^{\frac{1}{2+\mu}}$. Indeed, we have

$$\frac{n}{2\sigma^2}\gamma^2 M^{-(\mu+1)} = \frac{n}{2\sigma^2}\gamma^2 n^{-\frac{\mu+1}{\mu+2}}c_0^{-(\mu+1)} = \frac{\gamma^2}{2\sigma^2}n^{\frac{1}{\mu+2}}c_0^{-(\mu+1)}$$

and

$$\alpha\ln\left(2^{M/8}\right) = \alpha\frac{M}{8}\ln 2 = \frac{\alpha\ln 2}{8}c_0 n^{\frac{1}{\mu+2}}.$$

Moreover, by Varshamov-Gilbert bound, $\ln\left(2^{M/8}\right) \leq \ln N$.

We choose γ^2 so that both bounds for c_0 coincide, i.e.

$$\frac{1}{2}\left(\frac{\gamma^2}{32}\right)^{\frac{1}{\mu}} = \left(\frac{4\gamma^2}{\alpha\sigma^2\ln 2}\right)^{\frac{1}{2+\mu}}.$$

Hence, we set

$$\begin{aligned}
\gamma^2 &= 2^{\frac{\mu(\mu+2)}{2}}32^{\frac{\mu+2}{2}}\left(\frac{4}{\alpha\sigma^2\ln 2}\right)^{\frac{\mu}{2}}, \\
c_0 &= \frac{1}{2}\left(\frac{\gamma^2}{32}\right)^{\frac{1}{\mu}} = \frac{8\times 2^{\frac{\mu+1}{2}}}{\left(\alpha\sigma^2\ln 2\right)^{1/2}}.
\end{aligned}$$

For this choice of c_0, both inequalities (1.44) and (1.45) hold and therefore the convergence result follows from Theorem 2.5 of **?**.

1.7.2 Algorithm of FPLS

Algorithm 1: algorithm for the estimation of β via FPLS with m iterations

Result: $\hat{\beta}_m^{PLS}(t) = \sum_{j=1}^{m} q_j w_j(t)$

Initialisation : $\hat{\beta}_0^{PLS} = 0$, $X^{(0)} = \underline{X}$, $Y^{(0)} = \underline{Y} - T_n \hat{\beta}_0^{PLS}$, $S_0 = T_n^* \underline{Y}$;

$\hat{C}_{xy}^{(1)} = \frac{1}{n} \sum_{i=1}^{n} X_i^{(0)} Y_i^{(0)}$, $\widehat{\underline{Y}}_0 = 0$;

for $j \leftarrow 1$ *to* m **do**

1. Select the first eigenfunction r_j;

$r_j \leftarrow \hat{C}_{xy}^{(1)}$;

2. Calculate the score of X, called ψ_j;

$\psi_j \leftarrow [\int_S X_1^{(j)}(t) r_j(t) dt, ..., \int_S X_n^{(j)}(t) r_j(t) dt]'$;

3. Calculate the Fourier coefficient of β on the basis r_j;

$q_j \leftarrow \frac{||S_{j-1}||^2}{||\psi_j||^2}$;

4. Update the calculation of the estimated β;

$\hat{\beta}_j^{PLS}(t) \leftarrow \hat{\beta}_{j-1}^{PLS}(t) + q_j r_j(t)$;

5. Update the new values of Y and the predicted $\widehat{\underline{Y}}_i$;

$Y^{(j)} \leftarrow Y^{(j-1)} - q_j \psi_j$;

$\widehat{\underline{Y}}_i \leftarrow \widehat{\underline{Y}}_{i-1} + q_j \psi_j$;

6. Update the new covariance S_j;

$S_j \leftarrow T_n^* Y^{(j)}$;

7. Update the new eigenfunctions;

$\eta_j \leftarrow \frac{||S_j||^2}{||S_{j-1}||^2}$;

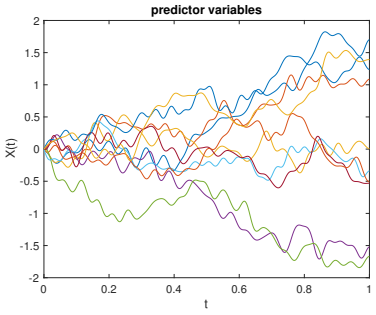
$r_{j+1} \leftarrow S_j + \eta_j r_j$;

$j \leftarrow j + 1$;

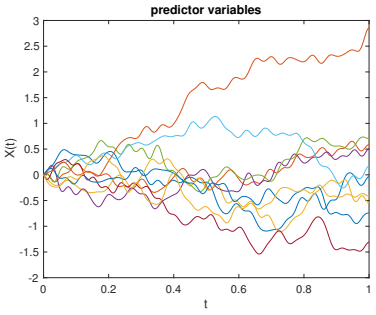
end

1.8 Graphics and Tables

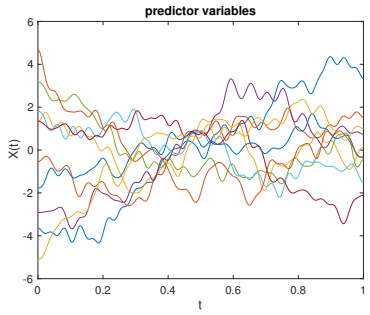
FIGURE 1.1: Sample of 10 observations of the functional predictor variable X(t) for the 8 different models



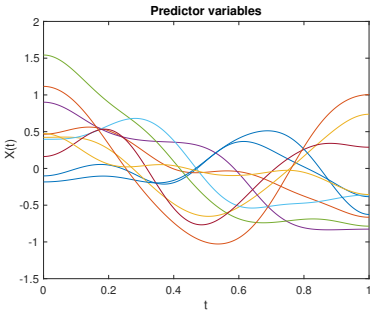
(a) Model 1



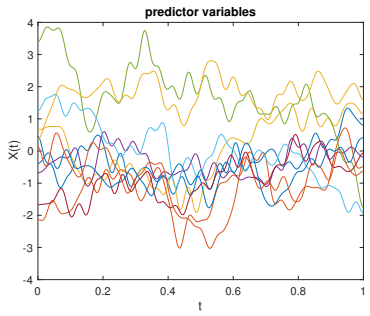
(b) Model 2



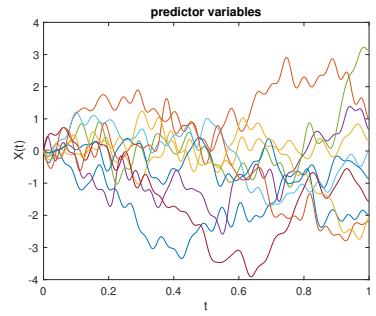
(c) Model 3



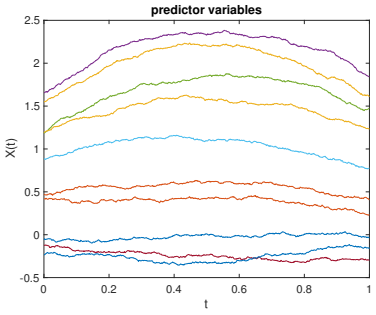
(d) Model 4



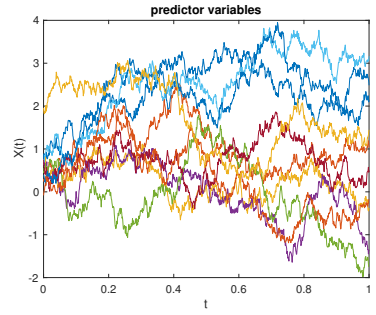
(e) Model 5



(f) Model 6

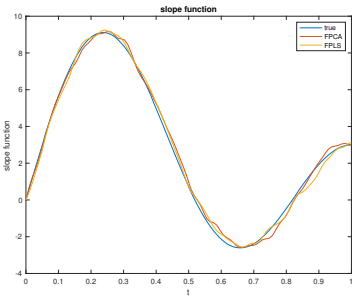


(g) Model 7

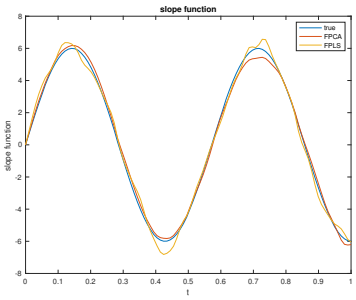


(h) Model 8

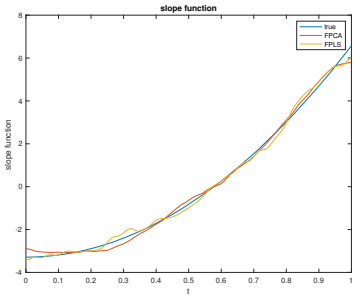
FIGURE 1.2: Estimation slope function for the 8 different models



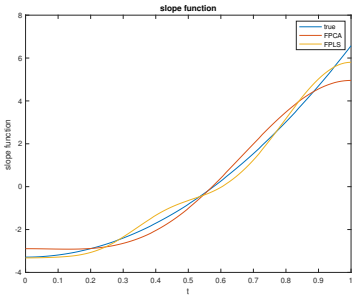
(a) Model 1



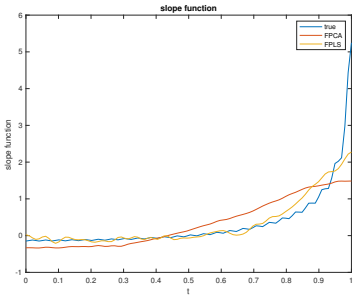
(b) Model 2



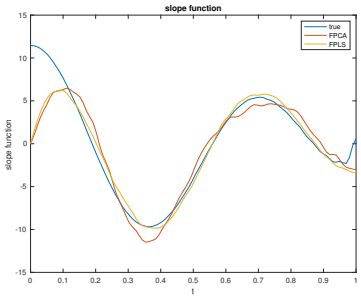
(c) Model 3



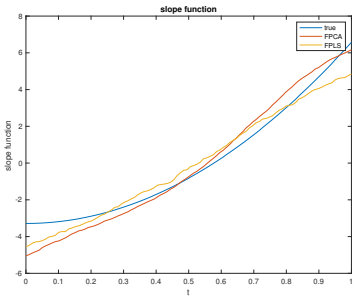
(d) Model 4



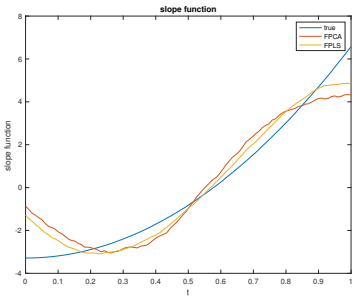
(e) Model 5



(f) Model 6

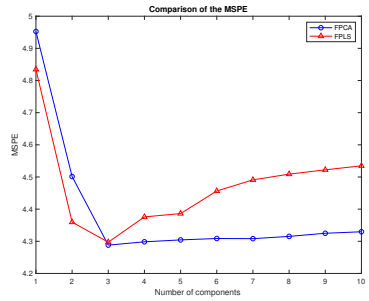


(g) Model 7

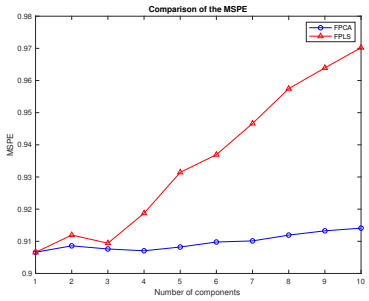


(h) Model 8

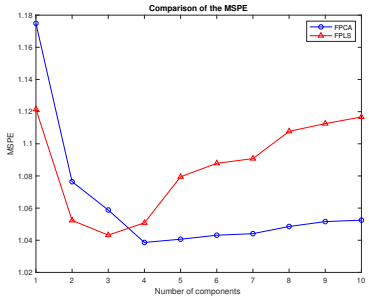
FIGURE 1.3: Comparison of the MSPE using FPCA and FPLS



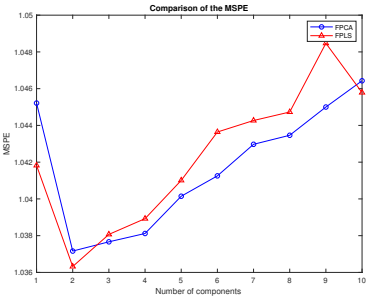
(a) Model 1



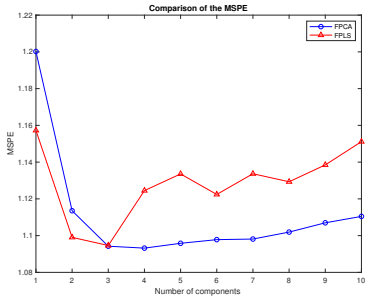
(b) Model 2



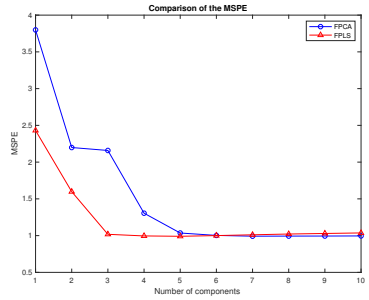
(c) Model 3



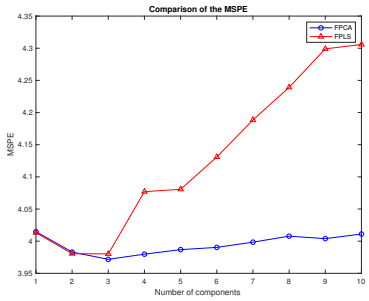
(d) Model 4



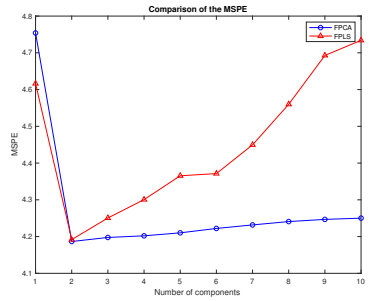
(e) Model 5



(f) Model 6



(g) Model 7



(h) Model 8

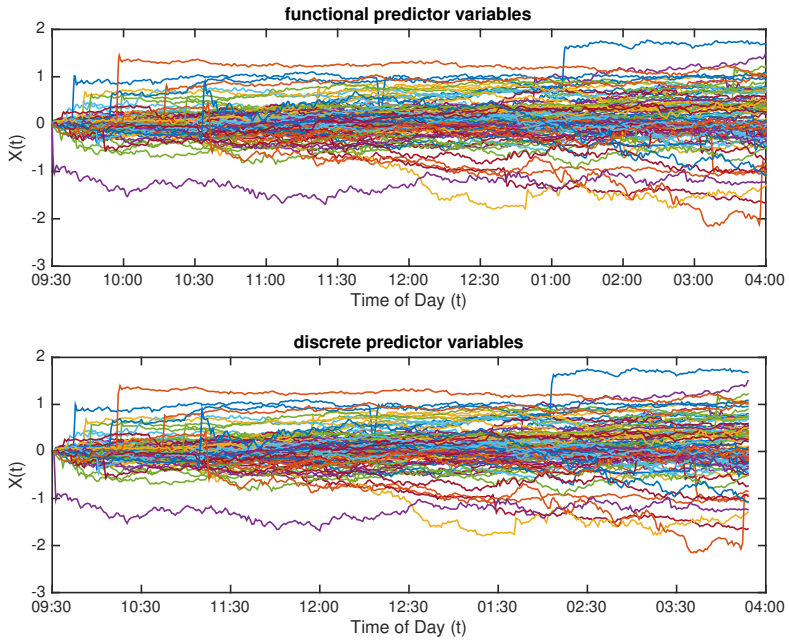
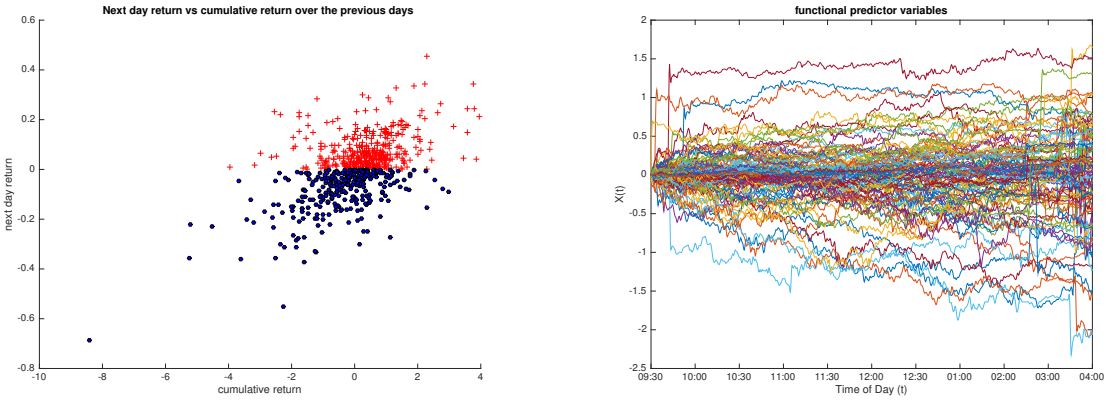


FIGURE 1.4: Daily price evolution of 100 observations of the return curves within a day from January 2014 to December 2017



(a) Next day returns vs end of the day cumulative return over the previous day

(b) 100 observations of the cumulative return curves within a trading day

FIGURE 1.5: Target daily return and cumulative returns

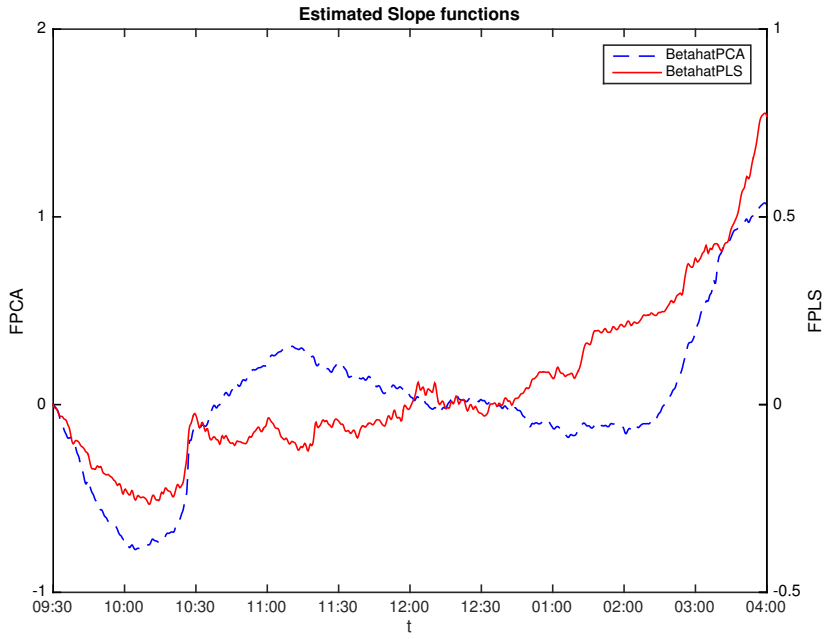


FIGURE 1.6: Estimated slope functions

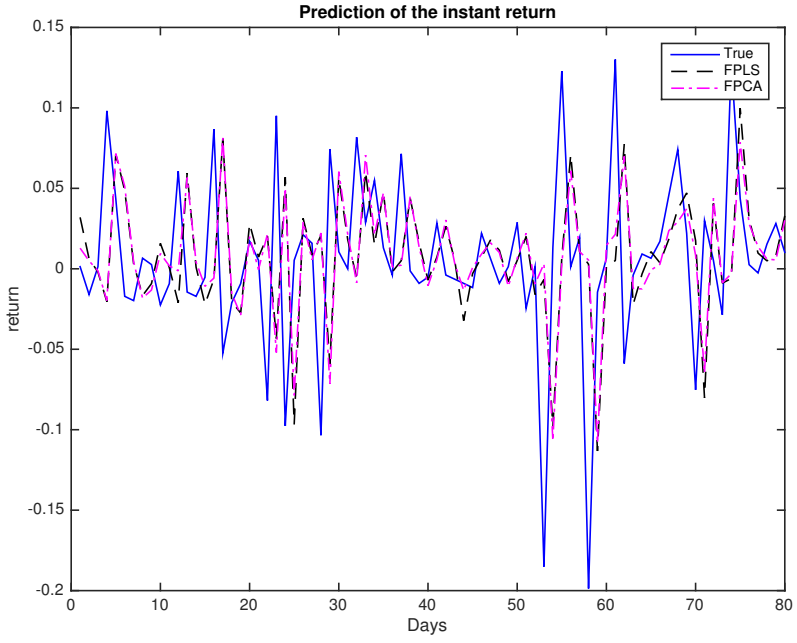
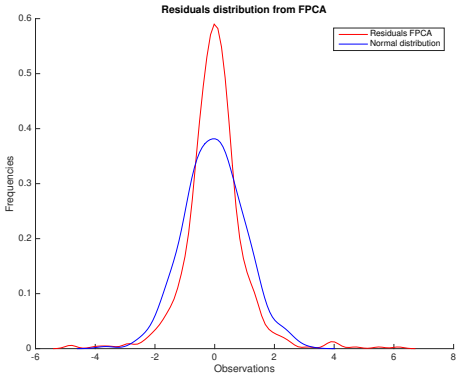
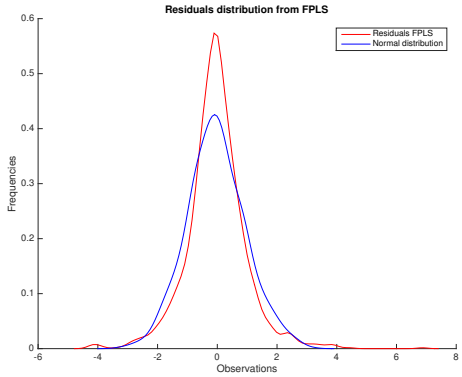


FIGURE 1.7: Realized return versus predicted



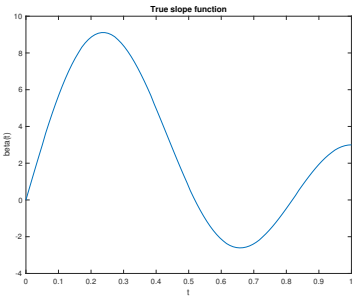
(a) Residuals distribution via FPCA



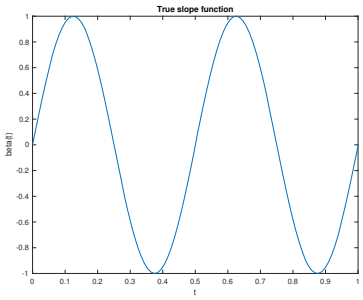
(b) Residuals distribution via FPLS

FIGURE 1.8: Residuals distributions

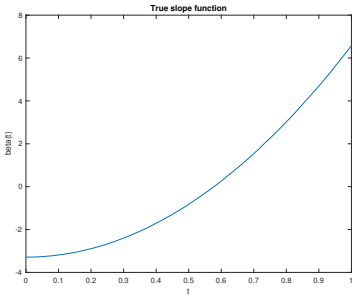
FIGURE 1.9: True slope function for the 8 different models



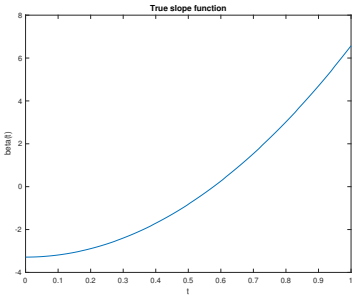
(a) Model 1



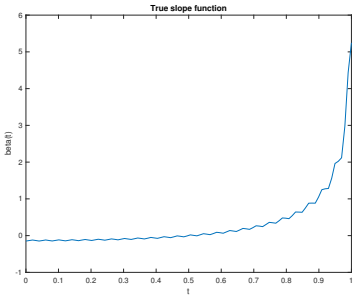
(b) Model 2



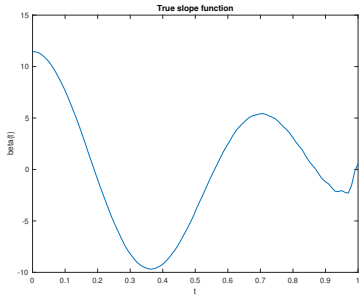
(c) Model 3



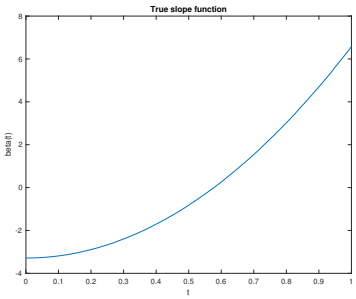
(d) Model 4



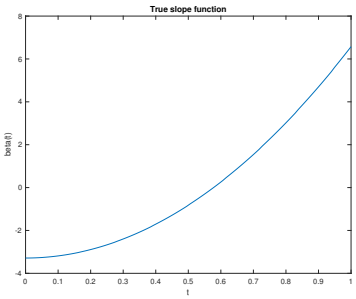
(e) Model 5



(f) Model 6



(g) Model 7



(h) Model 8

TABLE 1.1: Comparison of the MSPE using the 8 models, $n = 1000$, $m = 1, ..., 10$, $\sigma = 1$, $J = 50$ and $M = 1000$ simulations.

m	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS
1	1.60	1.50	1.02	1.02	1.17	1.11	1.15	1.04	1.09	1.06	3.77	2.61	3.77	2.61	3.75	2.58
2	1.16	1.05	0.99	1.00	1.01	1.00	1.01	1.00	1.01	1.00	2.40	1.66	2.40	1.66	2.38	1.66
3	1.00	1.00	1.00	1.01	0.98	0.98	1.00	1.00	1.01	1.01	2.33	1.07	2.33	1.07	2.32	1.07
4	1.00	1.01	1.01	1.02	0.99	1.01	1.00	1.00	1.00	1.01	1.26	1.04	1.26	1.04	1.26	1.04
5	1.00	1.02	0.99	1.02	1.02	1.05	1.01	1.01	0.99	1.04	1.11	1.03	1.11	1.03	1.10	1.03
6	1.03	1.05	1.01	1.04	1.01	1.05	1.00	1.00	1.01	1.06	1.07	1.03	1.07	1.03	1.07	1.03
7	0.97	1.01	1.01	1.05	1.00	1.04	1.01	1.02	1.03	1.04	1.03	1.03	1.03	1.03	1.05	1.05
8	1.01	1.05	1.01	1.06	0.98	1.03	1.01	1.01	1.00	1.04	1.02	1.04	1.02	1.04	1.03	1.04
9	0.98	1.05	1.01	1.05	1.00	1.06	1.02	1.02	1.00	1.04	1.02	1.05	1.02	1.05	1.03	1.05
10	1.03	1.08	1.02	1.07	0.99	1.04	1.00	1.00	1.02	1.08	1.02	1.06	1.02	1.06	1.02	1.06

TABLE 1.2: Comparison of the MSE using the 8 models, $n = 1000$, $m = 1, ..., 10$, $\sigma = 1$, $J = 50$ and $M = 1000$ simulations.

m	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS
1	20.51	17.61	0.49	0.47	0.67	0.43	0.68	0.24	0.61	0.45	34.02	24.30	34.02	24.30	34.03	24.18
2	12.33	4.15	0.50	0.57	0.17	0.08	0.16	0.07	0.37	0.23	28.54	15.50	28.54	15.50	28.55	15.51
3	0.17	0.40	0.53	2.38	0.07	0.16	0.08	0.08	0.27	0.25	27.81	4.67	27.81	4.67	27.84	4.65
4	0.29	8.49	0.46	9.23	0.06	1.42	0.07	0.24	0.23	0.90	11.07	3.45	11.07	3.45	10.94	3.47
5	0.44	31.53	0.51	26.08	0.09	3.75	0.10	0.69	0.23	3.34	6.98	2.99	6.98	2.99	6.87	2.97
6	0.60	60.33	1.02	60.32	0.13	7.89	0.17	1.43	0.24	8.11	5.73	3.89	5.73	3.89	5.73	3.85
7	1.28	111.52	1.40	114.28	0.19	16.31	0.33	2.66	0.28	14.21	4.58	8.05	4.58	8.05	4.51	7.81
8	2.38	171.16	1.87	182.45	0.26	23.37	0.67	5.64	0.32	21.60	4.14	15.17	4.14	15.17	4.17	15.18
9	3.25	269.53	2.95	253.07	0.34	31.77	1.28	12.05	0.43	28.17	3.64	23.72	3.64	23.72	3.63	23.55
10	3.72	381.58	4.02	328.47	0.50	39.69	2.73	25.77	0.55	38.03	3.54	33.12	3.54	33.12	3.53	32.82

TABLE 1.3: Comparison of the squared bias for the MSE using the 8 models, $n = 1000$, $m = 1, ..., 10$, $\sigma = 1$, $J = 50$ and $M = 1000$ simulations.

m	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS
1	20.50	17.60	0.49	0.45	0.66	0.42	0.65	0.23	0.61	0.45	34.01	24.23	34.02	24.30	34.02	24.10
2	12.28	4.06	0.47	0.14	0.15	0.06	0.15	0.05	0.36	0.22	28.50	15.43	28.51	15.43	28.51	15.43
3	0.02	0.04	0.43	0.08	0.05	0.02	0.06	0.02	0.25	0.11	27.71	4.42	27.70	4.41	27.72	4.39
4	0.01	1.39	0.22	0.17	0.03	0.03	0.03	0.01	0.19	0.08	10.77	3.04	10.77	3.04	10.94	3.47
5	0.04	4.13	0.03	0.34	0.01	0.07	0.01	0.01	0.15	0.09	6.71	2.18	6.72	2.17	6.61	2.15
6	0.06	8.38	0.03	0.82	0.01	0.12	0.01	0.01	0.12	0.08	5.45	1.53	5.45	1.53	5.44	1.55
7	0.04	9.18	0.01	1.77	0.01	0.34	0.01	0.01	0.10	0.15	4.24	1.27	4.26	1.25	4.20	1.23
8	0.20	19.57	0.03	1.54	0.01	0.53	0.01	0.05	0.09	0.27	3.77	1.12	3.78	1.14	3.77	1.11
9	0.10	25.29	0.02	2.93	0.01	0.81	0.01	0.03	0.08	0.20	3.17	0.98	3.17	0.98	3.16	1.00
10	0.34	41.60	0.01	3.71	0.01	1.16	0.01	0.04	0.07	0.32	2.92	1.01	2.95	1.03	2.92	1.01

TABLE 1.4: Comparison of the variance term for the MSE using the 8 models, $n = 1000$, $m = 1, ..., 10$, $\sigma = 1$, $J = 50$ and $M = 1000$ simulations.

m	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS
1	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.07	0.01	0.08	0.01	0.08
2	0.05	0.09	0.03	0.43	0.01	0.02	0.02	0.02	0.01	0.02	0.04	0.07	0.05	0.08	0.04	0.07
3	0.15	0.36	0.10	2.31	0.02	0.14	0.02	0.06	0.02	0.14	0.09	0.25	0.09	0.25	0.11	0.26
4	0.28	7.10	0.24	9.07	0.04	1.39	0.04	0.23	0.04	0.82	0.30	0.41	0.33	0.42	0.28	0.40
5	0.40	27.41	0.48	25.74	0.08	3.68	0.08	0.68	0.08	3.25	0.27	0.81	0.27	0.85	0.26	0.82
6	0.54	51.95	0.99	59.50	0.12	7.77	0.16	1.42	0.12	8.03	0.28	2.35	0.26	2.38	0.29	2.30
7	1.24	102.34	1.39	112.51	0.18	15.97	0.33	2.65	0.18	14.06	0.34	6.78	0.36	6.68	0.32	6.58
8	2.19	151.59	1.84	180.91	0.25	22.83	0.66	5.59	0.23	21.34	0.37	14.05	0.33	14.07	0.41	14.07
9	3.14	244.24	2.92	250.14	0.32	30.96	1.28	12.02	0.35	27.97	0.48	22.75	0.43	22.71	0.48	22.55
10	3.38	339.98	4.01	324.76	0.49	38.53	2.71	25.73	0.48	37.71	0.62	32.12	0.58	32.17	0.60	31.81

TABLE 1.5: Comparison of the squared bias for the MSPE using the 8 models, $n = 1000$, $m = 1, ..., 10$, $\sigma = 1$, $J = 50$ and $M = 1000$ simulations.

m	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS
1	1.31	1.30	0.01	0.01	7.69	7.69	4.04	4.04	0.19	0.19	2.72	2.72	2.72	2.72	2.85	2.86
2	1.33	1.32	0.01	0.01	9.93	9.92	4.26	4.26	0.20	0.20	3.22	3.21	3.22	3.21	2.66	2.65
3	1.57	1.57	0.01	0.01	8.61	8.61	3.56	3.56	0.17	0.17	2.86	2.86	2.86	2.86	3.21	3.21
4	1.16	1.16	0.01	0.01	8.71	8.72	4.73	4.73	0.15	0.15	3.13	3.13	3.13	3.13	2.83	2.83
5	1.38	1.38	0.01	0.01	9.50	9.50	3.71	3.71	0.15	0.15	3.13	3.13	3.13	3.13	3.14	3.14
6	1.27	1.26	0.01	0.01	8.39	8.39	3.84	3.84	0.14	0.14	3.12	3.12	3.12	3.12	2.77	2.77
7	1.08	1.10	0.01	0.01	7.71	7.74	3.73	3.73	0.16	0.15	3.25	3.25	3.25	3.25	2.37	2.36
8	1.14	1.13	0.01	0.01	7.44	7.45	3.52	3.53	0.15	0.15	2.66	2.66	2.66	2.66	3.09	3.08
9	1.30	1.30	0.01	0.01	8.11	8.11	3.59	3.59	0.18	0.17	2.88	2.88	2.88	2.88	2.78	2.79
10	1.28	1.28	0.01	0.01	9.41	9.43	4.62	4.61	0.17	0.17	3.14	3.15	3.14	3.15	2.47	2.48

TABLE 1.6: Comparison of the variance term for the MSPE using the 8 models, $n = 1000$, $m = 1, ..., 10$, $\sigma = 1$, $J = 50$ and $M = 1000$ simulations.

m	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS
1	0.75	0.85	0.01	0.01	7.96	8.03	3.95	4.06	0.09	0.13	0.21	1.40	0.21	1.40	0.19	1.37
2	1.18	1.30	0.01	0.01	8.01	8.03	4.13	4.15	0.16	0.17	1.57	2.30	1.57	2.30	1.56	2.29
3	1.41	1.41	0.01	0.02	8.07	8.09	4.11	4.12	0.16	0.18	1.66	2.89	1.66	2.89	1.65	2.91
4	1.33	1.36	0.01	0.03	8.22	8.25	4.11	4.12	0.18	0.19	2.67	2.92	2.67	2.92	2.71	2.96
5	1.45	1.47	0.02	0.04	8.20	8.23	4.14	4.15	0.18	0.20	2.86	2.97	2.86	2.97	2.88	2.98
6	1.44	1.48	0.02	0.05	8.26	8.29	4.11	4.11	0.18	0.21	2.91	2.99	2.91	2.99	2.93	3.01
7	1.41	1.45	0.02	0.05	8.00	8.07	4.15	4.15	0.18	0.21	2.90	2.96	2.90	2.96	2.94	3.01
8	1.44	1.48	0.02	0.06	8.01	8.05	4.17	4.18	0.18	0.22	2.96	3.03	2.96	3.03	2.94	3.01
9	1.35	1.39	0.02	0.07	8.25	8.30	4.14	4.14	0.18	0.22	2.94	3.01	2.94	3.01	2.94	3.00
10	1.36	1.42	0.02	0.07	8.36	8.41	4.13	4.14	0.18	0.23	2.95	3.02	2.95	3.02	2.95	3.01

TABLE 1.7: Number of selected components via cross-validation for FPCA and FPLS.

Models	Components	
	FPCA	FPLS
1	3	3
2	1	1
3	4	3
4	2	2
5	3	3
6	6	3
7	3	3
8	2	2

TABLE 1.8: Comparison of the RMSPE and R-squared

	FPCA	FPLS	OLS
m_*	6	3	-
RMSPE	0.093	0.092	0.14
R^2_{is}	22.4%	23.1%	77.8%
R^2_{oos}	1.75%	2.2%	−127.7%

TABLE 1.9: Comparison of the MAPE using the 8 models, $n = 1000$, $m = 1, ..., 10$, $\sigma = 1$, $J = 50$ and $M = 1000$ simulations.

m	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS	FPCA	FPLS
1	1.61	1.51	1.01	1.01	1.11	1.11	1.15	1.05	1.08	1.05	3.77	2.53	3.76	2.60	3.74	2.57
2	1.13	1.02	0.99	0.99	0.99	1.00	1.00	0.99	1.01	0.99	2.38	1.68	2.40	1.67	2.36	1.64
3	1.00	1.00	1.00	1.01	1.01	0.98	1.00	1.00	1.02	1.01	2.29	1.05	2.31	1.07	2.29	1.06
4	0.98	0.99	1.00	1.02	1.04	1.01	1.00	1.00	0.99	1.00	1.22	1.01	1.25	1.03	1.26	1.03
5	1.02	1.06	0.99	1.02	1.07	1.04	1.01	1.01	0.99	1.02	1.11	1.02	1.10	1.02	1.09	1.02
6	1.03	1.05	1.00	1.03	1.04	1.07	1.00	1.00	0.99	1.03	1.06	1.03	1.06	1.03	1.07	1.02
7	0.95	0.97	1.01	1.05	1.02	1.04	1.01	1.02	1.02	1.06	1.03	1.03	1.02	1.02	1.05	1.04
8	0.98	1.02	1.01	1.06	0.98	1.02	1.01	1.01	0.99	1.04	1.04	1.07	1.02	1.04	1.03	1.04
9	0.97	1.00	1.01	1.04	1.00	1.06	1.02	1.02	0.99	1.03	1.01	1.04	1.01	1.05	1.02	1.05
10	1.04	1.08	1.02	1.06	0.99	1.03	1.00	1.00	1.02	1.06	1.04	1.07	1.02	1.06	1.02	1.06

Chapitre 2

Intraday Stock Market Forecasting via Functional Time Series *

2.1 Introduction

Times series models are commonly used in financial econometrics for return prediction, asset pricing, sentiment analysis, and asset allocation. These models usually consider each observation as a scalar observed sequentially (for example the daily frequency) and a standard autoregressive model is used for forecasting purpose. Alternatively, lower frequency data is exploited in an extension of the Autoregressive Conditional Heteroskedasticity (GARCH) model for prediction. However, when using this standard approach, each daily observation is considered as a scalar and the information about the dynamics between day $t-1$ and t is ignored. This leads to a potential loss of the additional insights that could have been discovered (see Ramsay and Silverman (2007)). Moreover, even if standard approaches have been proposed, the functional data analysis (FDA) framework tends to provide a more natural description of the data with more accurate inference and prediction (see Horváth et al. (2010)). However, the FDA is a less explored approach to financial data in a context where high-frequency data become available and high-frequency trading is gaining in popularity. In fact, functional time series (FTS) usually arise when very dense data $\{X(t),\ t \in [0,T]\}$ in which t is a continuous real variable can be naturally split into equal-length segments observed sequentially over time. Then, $X_n(t) = X(n-1+t)$, $t \in [0,1]$, $n = 1,2,...,N$.

Following the approach proposed by Bosq (2000), this paper attempts to forecast the S&P 500 intraday return via functional time series. Thus, the S&P 500 price

∗. I am greatly indebted to Marine Carrasco for her invaluable guidance. I am grateful to Benoit Perron, William McCausland, Lucienne Talba, Zerbo Souleymane, and the members of the atelier de discussion des doctorants en économie de l'Université de Montréal, CIREQ PhD Students 2018 in Montréal, the Marcel Dagenais Econometrics Seminar 2019, the SCSE 2019 and CEA conferences for their helpful comments.

values observed at the 1-minute frequency are used to construct the daily curves of cumulative intraday returns (CIDRs) as suggested by Gabrys et al. (2010) to obtain daily stationary return curves. Additionally, an autoregressive model of order 1 on functions, called FAR(1) is used for estimation and forecasting. Since a usual trading day happens between 09 :30 AM and 04 :00 PM, that is 390 minutes, each CIDR is considered as one observation [1] containing 390 discretization points. I consider the fully functional model

$$X_{n+1}(t) = \int_0^1 \psi(t,s)X_n(s)ds + \varepsilon_{n+1}(t) \;\; n \in \mathbb{Z} \tag{2.1}$$

where $X_n(s)$ is the curve of the cumulative intraday return at the minute s of day n, $\psi(t,s)$ is the kernel of the autoregressive operator, and $\varepsilon_{n+1}(t)$ is the innovation function of day $n+1$. This approach is practically important because market participants can use the forecast results to tactically adjust their market timing or portfolio rebalancing strategy within a trading day. Furthermore, from an econometric point of view, using FDA is interesting since it makes it possible to exploit additional information on the price dynamics within a day to improve the return forecast. Moreover, this approach offers room for developing new tools to analyze the returns predictability, such as the functional out-of-sample $R^2_{oos}(t)$ and others. This model is also considered as the generalization of the simple AR(1) or VAR(1) model when there is a very large number of parameters to estimate. In this context, one can exploit the additional information and the interpretation of the results can be done in a convenient way.

One of the most important challenges of this model is to estimate the autoregressive operator [2]. Indeed, with the high dimensionality of the space, the estimation of this model leads to an ill-posed inverse problem and there is a high probability of obtaining unstable estimators of the autoregressive operator. To overcome that issue, the literature usually suggest to use the FPCA to reduce the dimensionality and obtain the estimator via the estimation of the scores (see Bosq (2000), Kokoszka and Zhang (2012), Crambes et al. (2013), Aue et al. (2015),Shang (2017), Imaizumi and Kato (2018), Shang et al. (2019)). The problem with this approach is the fact that the estimation is usually limited by the decay rate of the eigenvalues of the covariance operator of the predictor function. This means that one tends to overfit if the eigenvalues decay very rapidly. Moreover, the factors extracted by the FPCA approach are not necessarily the ones that contribute optimally to predicting the response variable and are not usually interpretable. There is also a literature suggesting nonparametric methods in order to analyze the functional data models (see Besse et al. (2000), Ramsay and Silverman (2007), Ferraty and Vieu (2006), and Hörmann and Kokoszka (2012)), but using this approach assumes that one project the data on some fixed basis that are not data related.

1. The 1 minute is considered here just for illustration purpose. It is possible to use other timeframes such as the 5-minutes, or tick frequency.

2. The autoregressive operator is similar to the slope parameter in the context of a simple AR(1) model.

The contribution of this paper is to exploit the ill-posed problem literature and develop a comparative analysis of 4 different regularization methods that endeavour to avoid such drawbacks. The suggested methods are the functional Tikhonov (FT), the functional Landweber-Fridman (FLF), the functional spectral cut-off (FSC) approach, and the functional partial least squares (FPLS). The functional principal component analysis (FPCA) is also considered for comparative purposes. These methods depend on a tuning parameter. The convergence rate of the mean square error (MSE) and asymptotic normality of the estimator are derived for a given tuning parameter for the suggested methods. Additionally, a test based strategy is proposed to identify the number of lag needed on a general context of functional autoregressive of order p ($FAR(p)$). The advantage of the proposed test strategy is that the procedure is not necessarily PCA-based as proposed in prior papers (see Kokoszka and Reimherr (2013), Aue et al. (2015), and Liu et al. (2016)), but can be considered for different regularization methods that are linear in terms of the response variable.

Some Monte Carlo simulations have been developed to support the relevance of the theoretical results and compare the 4 methods. The comparison is based on the estimation criteria, which are the Mean Squared Error (MSE) and the Mean Absolute Deviation (AD), used to measure the quality of the estimation of the autoregressive operator. Furthermore, the predictive performance of the different methods is also compared based on the Mean Squared Prediction Error (MSPE), the Mean Absolute Prediction Error (MAPE), and the out-of-sample R^2 (R^2_{oos}). Based on a majority of the model settings considered, the simulation results show that the FPLS tends to outperform the other methods in terms of estimation accuracy of the autoregressive operator. This estimation performance of the FPLS is due to the fact that it is a supervised method in the sense that the estimated parameter is a combination of factors that are relevant to predict the response variable. In terms of prediction, all the estimation methods tend to present almost the same predictive performance for most of the data generating processes.

An overview of the real data findings shows the evidence that the cumulative intraday return curve of the current trading day contributes significantly to predicting the next day's cumulative return curve. All the considered estimation methods present approximately the same estimation of the autoregressive operator. In terms of prediction, all the methods display the same predictive performance when considering the MSPE and MAPE as criteria. Additionally, when we consider the R^2_{oos}, the FPLS method tends to outperform the other methods with an R^2_{oos} of 3.9%. Furthermore, the most predictable period of the next day's return based on current day's return are the periods 11 :00 AM - 11 :30 AM and 02 :00 PM - 4 :00 PM as they display maximum values of the functional $R^2_{oos}(t)$, that is 3% for FPCA and FPLS and 2% for FT and FLF. On the other hand, the less predictable period of the day is the period 09 :30 AM - 09 :45 AM as the $R^2_{oos}(t)$ is closed to zero.

The literature on FDA is gaining more attention, but the theoretical concepts and tools for functional times series are still nascent. The idea of using cumulative return

is inspired by the paper of Gabrys et al. (2010) and have been used in some of the preceding papers (Kokoszka and Zhang (2012) and Shang (2017)). Indeed, Kokoszka and Zhang (2012) use individual assets and their main purpose is comparing simple functional and fully functional model settings for a simple version of capital asset pricing model (CAPM). Kargin and Onatski (2008) suggested a predictive factor method for predicting the next day curve with FAR(1) model. The consistency results of the proposed estimation methods are derived and a comparison of these methods is made based on simulation and empirical analysis. This paper is also related to the paper by Benatia et al. (2017), Imaizumi and Kato (2018), Crambes et al. (2013), and Shang (2017). This paper is different from the one by Benatia et al. (2017) in the sense that I introduce a dynamics in the functional observations. It is also different to the paper by Carrasco and Tsafack (2020) as I consider a more general model where the response variable is a function.

The rest of the paper is organized as follows. Section 2.2 is dedicated to presenting the related literature. Section 2.3 details the functional econometric model. In section 2.4, I explain how to estimate the model using the four aforementioned methods. Section 2.5 analyzes the convergence rate of the estimated autoregressive operator. Section 2.6 examines the asymptotic normality of the parameter. Section 2.7 address the selection of the optimal number of lags for a generalized functional autoregressive model. Section 2.8 discusses on data driven approach to select the tuning parameter. Section 2.9 presents the comparison of the four methods based on Monte Carlo simulations. Section 2.10 develops the real data application. Finally, Section 2.11 concludes. The proofs of the main theoretical results are presented in the appendix.

2.2 Related literature

This paper is related to three key pieces of literature : the functional data analysis, the functional autoregressive model, and the intraday return predictability in the financial market.

The literature on FDA has attracted a lot of attention in the statistical field during the last decade. Some of the pioneers are Ramsay and Silverman (2007), Kokoszka and Zhang (2012), and Ferraty and Vieu (2006), all of which developed a general context. One of the main challenges is to be able to estimate the slope function (in the context where the response is a scalar and the predictor is a function) or the operator (if the predictor and the response variable are both functions) due to the high dimensionality issue. More recently, authors such as Benatia et al. (2017), Crambes et al. (2013), and Imaizumi and Kato (2018) have analyzed the convergence rate and the asymptotic distribution of the estimated parameter for the i.i.d model where the predictor and the response variables are both functions. They respectively used the FT and FPCA estimation methods.

This paper is also related to the functional autoregressive model literature. Bosq (2000) and Hörmann and Kokoszka (2012) (among others) considered a parametric model

and used the Yule-Walker technique for estimation. On the same line, Kargin and Onatski (2008) proposed a predictive factor approach to estimate the autoregressive operator and developed the related consistency and convergence rate results. The idea of their approach is to project the response variable on a set of factors that ensures the minimization of the prediction error. Didericksen et al. (2012) compared the FPCA method proposed by Bosq (2000) and the predictive factor technique of Kargin and Onatski (2008) based on some simulation data ; they subsequently showed that, in an overview of the comparison, the FPCA outperforms.

Authors such as Besse et al. (2000) and Hays et al. (2012) adopted a nonparametric approach to estimating the autoregressive operator. Didericksen et al. (2012) compared the method of Kargin and Onatski (2008) and the FPCA and show that the FPCA outperforms the predictive factor in terms of estimation and both methods present the same predictive performance. Hyndman and Shang (2009) and Aue et al. (2015) have respectively proposed to use a univariate and multivariate time series forecasting method since the FPCA scores of a function can display a temporal dependence as the original function. Kokoszka and Young (2016) proposed a unit-root test for the functional time series. More recently, Cerovecki et al. (2019) analyzed the GARCH model for functional time series while Rice et al. (2019) proposed a test and goodness-of-fit for the FGARCH models. So far, there is still a lot to discover in the functional time series models. These approaches are limited by the configuration of the eigenvalues.

This paper is related to but different from the preceding one in the sense that new estimation approaches are suggested. The proposed methods are not based on a prior PCA projection step as is usually done in most of the papers (see Hyndman and Shang (2009), Hays et al. (2012), Aue et al. (2015), and Shang (2017)). Furthermore, a test-based strategy to select the optimal number of lags is proposed in a more generalized estimation approach, including the FPCA. Moreover, this paper exploits the regularization techniques and the estimated parameter for each approach are written as the product of matrices and vectors. The consistency results are analyzed with assumptions closed to the one by Benatia et al. (2017),albeit less restrictive.

To the best of our knowledge, the usage of functional time series is less observed in the financial econometrics application. Only a few authors have started to investigate this strain of literature. Kokoszka and Zhang (2012) proposed to predict an individual stock by using a functional version of the capital asset pricing model (CAPM) and compare a simple functional setting to a fully functional model in the autoregressive framework. Shang (2017) suggested to forecast the U.S. stock market by combining the dynamic update technique with the PCA-based approach proposed by Hyndman and Shang (2009) and Aue et al. (2015). Sancetta (2019) used the FDA framework to predict the end of the day volume in the currency market.

The exploration of the high-frequency data in developing functional time series models is supported by the recent papers related to the intraday returns predictability. Gao et al. (2018) documented the intraday momentum in the U.S. stock market at the 30 minutes

frequency. They show that the first half-hour return contributes to predicting the last half hour return and that the effect is stronger in more volatile days, on higher volume days, recession days, and high impact news release days. Bogousslavsky (2016) identified the infrequent rebalancing and the late-informed investors' effect [3] as the main causes of momentum in the stock market. In the same line, Zhang et al. (2019) documented almost the same results by Gao et al. (2018) in the China stock market. Chu et al. (2019) found that the last half hour is positively predicted by the first half-hour, but they also identified a reversal effect in the second half-hour of the trading day in the Chinese stock market. They also found that this momentum and reversal effect is robust when including previous day return and day-of-week. Heston et al. (2010) discovered a striking pattern of return continuation at half-hour intervals that are exact multiples of a trading day on a 40-day time horizon. Following this intraday return predictability literature, combined with the advantages of the functional time series in exploiting the availability of high-frequency data, there is potential to improve return predictions, discover new insights, and develop new tools.

2.3 The Model Setting

In this paper, for each day i, the shape of the cumulative intraday return of the S&P 500 is observed at the 1-minute frequency. The cumulative intraday returns (CIDRs) by Gabrys et al. (2010) are used to construct the curves. Let $P_i(t_j)$ be the price of a financial asset at time t_j, on a given day i. Since a trading session is opened from 09 :30 AM to 04 :00 PM, the total number of minutes within that period is 390 and therefore, $j \in \{1, ..., 390\}$. The return is defined as

$$R_i(t_j) = 100 * [ln(P_i(t_j)) - ln(P_i(t_1))] \quad with \quad j \in 1, ..., 390. \tag{2.2}$$

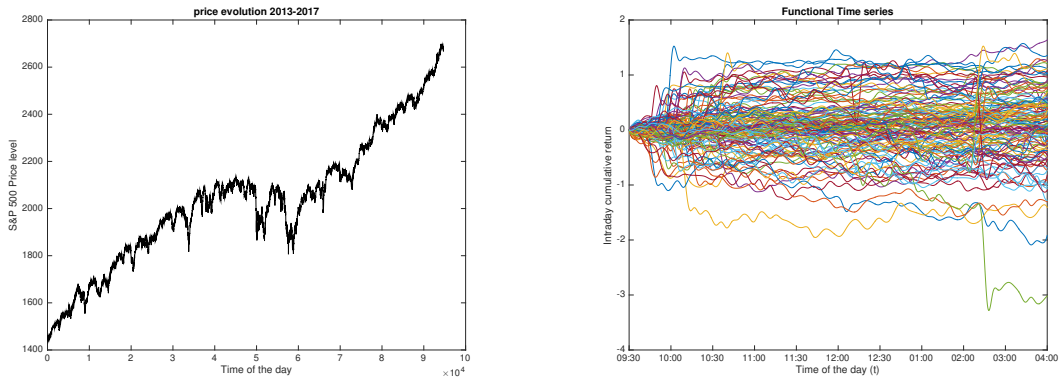
and the CIDRs are defined by

$$X(t) = R_i(t_j) \quad with \quad t_j \in \left(\frac{(j-1)}{390}, \frac{j}{390}\right]. \tag{2.3}$$

Figure 2.1 displays the constructed intraday cumulative returns of the $S\&P500$ based on the raw data for the period 2013 - 2017.

Let $(X_i : i \in \mathbb{Z})$ be an arbitrary stationary functional time series of the S&P 500 CIDRs. It is assumed that each random function X_i is an element of $\mathbb{H} = L^2([0,1])$ (the space of square integrable functions mapping from the compact interval [0,1] to $\mathbb{R}$) endowed with the inner product $< f, g >= \int_0^1 f(t)g(t)dt$ and the norm $||f|| = \left(\int_0^1 f^2(t)dt\right)^{1/2}$ such that $E(||X_i||^2) < \infty$).

3. This is similar to the slow diffusion, analysis, and acceptance of new information identified by Chan (2013)



(a) Price evolution in 2013 - 2017 (b) Collection of functional observations

FIGURE 2.1: Intraday cumulative returns of the $S\&P$ 500 index in 2013-2017

Therefore, a sequence $\{X_1, X_2, ..., X_N\}$ of the realizations of X is observed, where X_i corresponds to the observed curve of S&P 500 of day $i = 1, ..., N$. In this paper, it is assumed that the sequence of $\mathbb{H}$-valued variables $\{X_1, X_2, ..., X_N\}$ follows a functional autoregressive Hilbertian process of order 1 (FAR(1)) presented as follows :

$$X_{n+1}(t) = \int_0^1 \psi(t,s)X_n(s)ds + \varepsilon_{n+1}(t) \;\; n \in \mathbb{Z} \tag{2.4}$$

where for each day n , X_n is a random curve of $\mathbb{H}$. The operator

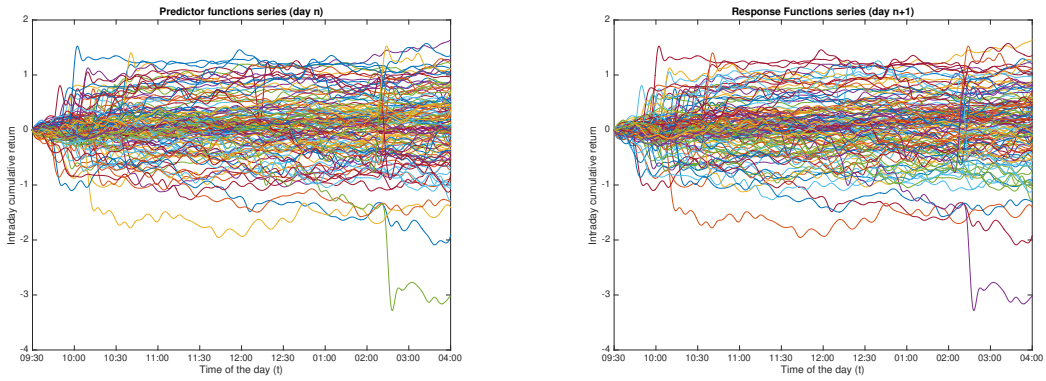
$$\Psi : \mathbb{H} \to \mathbb{H}$$
$$f \to \Psi(f) = \int_0^1 \psi(s,t)f(s)ds$$

is a bounded linear operator and $\varepsilon = (\varepsilon_n,\ n \in \mathbb{Z})$ is a $\mathbb{H}$-valued stationary and ergodic martingale difference with respect to $\{\varepsilon_{n-1}, \varepsilon_{n-2}, ..., X_{n-1}, X_{n-2}, ...\}$. In addition, $E(\varepsilon_n|X_{n-1}) = 0$ and $E(||\varepsilon_n||^2|\mathcal{F}_{n-1}) = \sigma^2 < \infty$. Without loss of generality, it is assumed that $E(X_n) = 0$.

Figure 2.2 represents how the predictor and the predicted functions are displayed. According to what is observed, it can be deduced that if the few outliers are removed from the sample, it is possible to say that each functional observation is generated from the same data generation process. This idea has been argued by Kokoszka and Young (2016) who developed a KPSS unit-root test for functional time series.

Let us denote by $\mathcal{L}$ the space of bounded linear operators on $\mathbb{H}$ equipped with the norm

$$||\Psi||_{\mathcal{L}} = Sup\{||\Psi(f)|| : \;\; ||f|| \leq 1\}. \tag{2.5}$$



(a) Predictor functions (day n)

(b) Response functions (day n+1)

FIGURE 2.2: Functional predictor and functional response

Under the conditions that there exists an integer $j_0 \geq 1$ such that the linear operator $||\Psi^{j_0}||_{\mathcal{L}} < 1$, Equation (2.4) has a unique solution, which is a weakly stationary process in $\mathbb{H}$ given by

$$X_n = \sum_{k=0}^{\infty} \Psi^k(\varepsilon_{n-k}) \tag{2.6}$$

and the series converges almost surely in $\mathbb{H}$. If it is assumed that the Hilbert-Schmidt norm of the operator Ψ is lower than 1, then the existence and the uniqueness of the solution are satisfied (see Lemma 3.1 of Kokoszka and Zhang (2010)). In the next section, various regularization techniques are presented.

2.4 Model estimation

The goal of this paper is to forecast the one day ahead S&P 500 shape X_{n+1}. According to the data generating process, the best linear predictor of X_{n+1} given $X_1, ..., X_n$ is given by $\Psi(X_n)$. Typically, Ψ is unknown and should be estimated consistently by an estimator $\hat{\Psi}$. This section presents four different estimation strategies of the autoregressive operator Ψ. Multiplying Equation (2.4) by X_n and taking the expectation on both sides leads to the following equation :

$$E[< X_{n+1}, f > X_n] = E[< \Psi(X_n), f > X_n], \quad f \in \mathbb{H}.$$

Let us define the covariance operator by

$$K(f) = E[< X_n, f > X_n]$$

Since $E[||X_n||^2] < \infty$, the covariance operator is symmetric, positive, nuclear and

therefore, Hilbert-Schmidt and its spectral system $(v_j, \lambda_j)_{j\geq 1}$ is defined by

$$K(v_j) = \lambda_j v_j, \;\; j \geq 1,$$

with the eigenfunctions v_j forming an orthonormal basis of $\mathbb{H}$ and the eigenvalues are such that $\lambda_1 \geq \lambda_2 \geq ... \geq 0$. Let define the cross-covariance operator by

$$D(f) = E[< X_{n+1}, f > X_n].$$

Then, it is easy to see that

$$D(f) = K\Psi^*(f). \tag{2.7}$$

The operators K and D are unknown and can be estimated by $\hat{K}$ and $\hat{D}$, respectively, where

$$\hat{D}(f) = \frac{1}{N-1}\sum_{n=1}^{N-1} < X_{n+1}, f > X_n$$

and

$$\hat{K}(f) = \frac{1}{N-1}\sum_{n=1}^{N-1} < X_n, f > X_n.$$

The empirical spectral system of $\hat{K}$ is $(\hat{\lambda}_j, \hat{v}_j)_{j\geq 1}$ with $\hat{\lambda}_j \geq \hat{\lambda}_j \geq ... \geq 0$ and $(\hat{v}_j)_{j\geq 1}$ form an orthonormal basis of $\mathbb{H}$.

Given Equation (2.7), one would like to directly estimate the autoregressive operator by writing $\Psi^* = K^{-1}D$, as is usually done in the finite-dimensional context. The problem is that the covariance operator K is compact and is defined in an infinite-dimensional space. Thus, K^{-1} is a noncontinuous operator in the considered space. This result leads to an unstable and noncontinuous estimator of the autoregressive operator. In the inverse problem literature, Equation (2.7) is called an ill-posed problem in the sense that K is only invertible on a subset of $\mathbb{H}$ and its inverse is not continuous.

This paper exploits the functional Yule-Walker Equation (2.7) and estimates the autoregressive operator by 4 different regularization techniques that are the Functional Tikhonov (FT), the Functional Spectral Cut-off (FSC), the Functional Partial Least Squares (FPLS), and the Functional Landweber-Fridman iteration method (FLF). Let now present the considered methods.

2.4.1 The Functional Spectral Cut-off

This approach is almost similar to the FPCA method that is widely proposed in the functional time series literature in order to estimate the autoregressive operator on a

finite subspace of $\mathbb{H}$. Since the operator K is symmetric and nuclear, it admits a spectral decomposition, that is

$$k(s,t) = \sum_{j=1}^{\infty} \lambda_j v_j(s) v_j(t),$$

where $\{v_j\}_{j=1}^{\infty}$ is an orthonormal basis of $\mathbb{H}$. The Functional Spectral Cut-off method consists of selecting the eigenfunctions associated with the eigenvalues greater than some threshold $\alpha > 0$. The inverse of the covariance operator K can be written as

$$K^{-1}(f) = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} < f, v_j > v_j$$

and the regularized inverse of K via FSC approach is given by

$$K_{\alpha}^{-1}(f) = \sum_{\lambda_j > \alpha} \frac{1}{\lambda_j} < f, v_j > v_j.$$

Then, the estimated autoregressive operator is given by

$$\begin{aligned} \Psi_{\alpha}^{*}(f) &= K_{\alpha}^{-1} D(f) \\ &= \sum_{\lambda_j > \alpha} \frac{1}{\lambda_j} < D(f), v_j > v_j \\ &= \sum_{\lambda_j > \alpha} \frac{1}{\lambda_j} \mathbb{E}[< X_{n+1}, f >< X_n, v_j >] v_j \end{aligned}$$

and its empirical counterpart is

$$\hat{\Psi}_{\alpha}^{*}(f) = \frac{1}{N-1} \sum_{j=1}^{N-1} \sum_{n=1}^{N-1} \frac{\hat{Q}_{\alpha,j}}{\hat{\lambda}_j} < X_{n+1}, f >< X_n, \hat{v}_j > \hat{v}_j, \quad for\ each\ f\ \in\ \mathbb{H} \tag{2.8}$$

with $\hat{Q}_{\alpha,j} = I(\hat{\lambda}_j \geq \alpha)$ and α the tuning parameter.

Similar to FSC, the FPCA method consists in projecting the response variable onto the principal components of the covariance operator. Those principal components are nothing else than the eigenfunctions of the operator K associated with the largest eigenvalues. Thus, if m eigenfunctions are selected for the estimation, the FPCA estimator is given by

$$\begin{aligned}\Psi_m^*(f) &= K_m^{-1} D(f) \\ &= \sum_{j=1}^{m} \frac{1}{\lambda_j} < D(f), v_j > v_j \\ &= \sum_{j=1}^{m} \frac{1}{\lambda_j} \mathbb{E}[< X_{n+1}, f >< X_n, v_j >] v_j\end{aligned}$$

and its empirical version is given by

$$\hat{\Psi}_m^*(f) = \frac{1}{N-1} \sum_{j=1}^{N-1} \sum_{n=1}^{N-1} \frac{\hat{Q}_{m,j}}{\hat{\lambda}_j} < X_{n+1}, f >< X_n, \hat{v}_j > \hat{v}_j, \quad for\ each\ f\ \in\ \mathbb{H} \tag{2.9}$$

where $\hat{Q}_{m,j} = I(j \leq m)$. This procedure was also considered by Crambes et al. (2013) for the i.i.d model. Another configuration of the FPCA is proposed by Imaizumi and Kato (2018) and it consists in projecting the predictor and the response function onto the first m principal components respectively, then uses the scores to estimate the Fourier coefficients of the estimated autoregressive operator. The estimated autoregressive operator is then obtained by writing the estimated operator on the basis of the m eigenfunctions of the covariance operator. This configuration by Imaizumi and Kato (2018) is not considered in this paper.

In addition, the eigenfunctions are unique as they are orthonormal. Also, for this model setting, the data are not generated by a factor model. Then, the FPCA is considered as a dimension reduction technique to estimate Ψ^*. If the data are generated by a factor model, the FPCA would estimate the related factors up to a rotation (see Bai and Ng (2002)).

2.4.2 Tikhonov Method

This technique is widely used in the inverse problem literature. It has been studied recently by Benatia et al. (2017) in the context of a fully functional regression. This technique is most widely justified to tackle the high dimensionality problem.

Let α be a positive tuning parameter. Then, the estimated autoregressive operator is given by

$$\Psi_\alpha^*(f) = \Big(\alpha I + K\Big)^{-1} D(f), \quad for\ each\ f\ \in\ \mathbb{H}$$

where I is the identity operator. This estimator can also be characterized in terms of the spectral system of the covariance operator K, as follows

$$\Psi_\alpha^*(f) = \sum_{j=1}^{+\infty} \frac{1}{\alpha + \lambda_j} \mathbb{E}[< X_{n+1}, f >< X_n, v_j >] v_j, \quad for\ each\ f\ \in\ \mathbb{H}.$$

The empirical version is then given by

$$\begin{aligned}\hat{\Psi}_\alpha^*(f) &= \left(\alpha I + \hat{K}\right)^{-1} \hat{D}(f) \\ &= \frac{1}{N-1} \sum_{j=1}^{N-1} \sum_{n=1}^{N-1} \frac{\hat{Q}(\alpha, \hat{\lambda}_j)}{\hat{\lambda}_j} < X_{n+1}, f >< X_n, \hat{v}_j > \hat{v}_j, \quad for\ each\ f\ \in\ \mathbb{H}\end{aligned}$$

where $\hat{Q}(\alpha, \hat{\lambda}_j) = \frac{\hat{\lambda}_j}{\hat{\lambda}_j + \alpha}$ is called the filter factor. The truncation that is operated with the FPCA method is replaced by the shrinkage effect of the parameter α.

2.4.3 Functional Landweber-Fridman(FLF)

The Landweber-Fridman method is basically an iterative method which consist to transform the normal equation $K(\Psi^*) = D$ into a fixed point problem. This method is similar to the gradient descent method used to solve a minimization problem. This method is also very popular in the literature of inverse problem. Let us consider a positive parameter d such that $0 < ||K||_{\mathcal{L}} < 1/d$. Then, the FLF technique can be computed iteratively as follows. Take the initial value

$$\Psi_0^*(f) = dD(f), \quad for\ each\ f\ \in\ \mathbb{H}.$$

For $h = 1, \ldots, \frac{1}{\alpha} - 1$, we have

$$\Psi_h^*(f) = (I - dK)(\Psi_{h-1}^*(f)) + dD(f), \quad for\ each\ f\ \in\ \mathbb{H}$$

where m is the maximum number of iterations. We see that the estimated autoregressive operator can be written as a polynomial function of the covariance operator K and we have

$$\Psi_\alpha^*(f) = d \sum_{l=1}^{1/\alpha} (I - dK)^{l-1} D(f), \quad for\ each\ f\ \in\ \mathbb{H}. \tag{2.10}$$

Since the operators K and D are not observed, they are consistently estimated by $\hat{K}$ and $\hat{D}$, respectively. Then, $\hat{\Psi}_\alpha^*$ is given by

$$\hat{\Psi}_\alpha^*(f) = d \sum_{l=1}^{1/\alpha} (I - d\hat{K})^{l-1} \hat{D}(f), \quad for\ each\ f\ \in\ \mathbb{H}. \tag{2.11}$$

This estimator can be written in terms of the eigensystem of the covariance operator $\hat{K}$, as follows

$$\hat{\Psi}^*_\alpha(f) = \frac{1}{N-1}\sum_{j=1}^{N-1}\sum_{n=1}^{N-1}\frac{\hat{Q}(\alpha,\hat{\lambda}_j)}{\hat{\lambda}_j} < X_{n+1}, f >< X_n, \hat{v}_j > \hat{v}_j, \quad for\ each\ f\ \in\ \mathbb{H} \quad (2.12)$$

where $\hat{Q}(\alpha,\hat{\lambda}_j) = \left(1-(1-d\hat{\lambda}_j)^{\frac{1}{\alpha}}\right)$ is the filter factor.

2.4.4 Functional Partial Least Squares(FPLS)

One of the main drawbacks of FPCA is that X_{n+1} is projected on the eigenfunctions of K associated with its largest eigenvalues regardless of their ability to predict X_{n+1}. Moreover, the selected principal components may not capture the most important information relevant for the prediction of X_{n+1} given X_n. The FPLS may be more adapted in the sense that it extracts the most important factors that explain the relation between the predictand and the predictor function. This method is very popular in the chemometrics field and has been discussed by some prior authors such as Wold et al. (1984) Helland (1988) and Höskuldsson (1988). It was recently introduced in the econometric field by Groen and Kapetanios (2009), Kelly and Pruitt (2015), and Carrasco and Rossi (2016). In the Functional regression context with a scalar response, there are authors like Aguilera et al. (2010), Delaigle and Hall (2012), and, more recently, Zhou (2019).

Practically, for the model setting of this paper, the idea is to identify a new factor $t_h = \int_0^1 X_n(s)\phi_h(s)ds$ at each step $h = 1, ..., m$ such that the covariance with the response function is maximized.

$$\begin{aligned} \max_{v_h, c_h \in L^2([0,1])} \quad & cov^2\left(\int_0^1 X_n(s)\phi_h(s)ds, \int_0^1 X_{n+1}(t)c_h(t)dt\right) \\ \text{subject to} \quad & ||\phi_h|| = 1, ||c_h|| = 1, and \\ & \int_0^1\int_0^1 \phi_\ell(s)k(s,t)\phi_h(t)dsdt = 0,\ \ell = 1, ..., h-1 \end{aligned} \quad (2.13)$$

where $\phi_1, ..., \phi_{h-1}, c_1, ..., c_{h-1}$ are already obtained in the $h-1$ previous step.

There are two main constraints in this optimization problem. The first one represents normalization to one of the eigenfunctions, while the second is nothing else than the orthogonality of the estimated factor with the one retained in the previous iterations.

Proposition 1

For each $t \in [0,1]$, the estimated autoregressive operator is given by :

$$\hat{\psi}_m^*(s,t) = \sum_{l=1}^{m} \hat{\gamma}_{t,l} \hat{K}^{l-1}(\hat{D})(s,t) \tag{2.14}$$

where for each $s,t \in [0,1]$, where for each t, $\hat{\gamma}_t = \hat{R}_t^{-1}\hat{\mu}_t$ is a vector of size m. $\hat{R}_t$ is an $(m \times m)$ matrix with elements $\hat{R}_t(j,l) = \hat{R}_{t,j,l}$:

$$\hat{R}_{t,j,l} = \int_0^1 \int_0^1 \hat{D}(t,u)\hat{K}^{j+l-1}(u,s)\hat{D}(s,t)duds \tag{2.15}$$

and $\hat{\mu}_t = [\hat{\mu}_{t,1}, ..., \hat{\mu}_{t,m}]'$ is a vector of length m :

$$\hat{\mu}_{t,l} = \int_0^1 \int_0^1 \hat{D}(t,u)\hat{K}^{l-1}(u,s)\hat{D}(s,t)duds. \tag{2.16}$$

Moreover, this estimator can be written in terms of the eigensystem of the empirical covariance operator $\hat{K}$ as

$$\hat{\Psi}_m(f) = \frac{1}{N-1} \sum_{j=1}^{N-1} \sum_{n=1}^{N-1} \frac{Q(m,\hat{\lambda}_j)}{\hat{\lambda}_j} < X_{n+1}, f >< X_n, \hat{v}_j > \hat{v}_j, \quad for\ each\ f \in \mathbb{H}. \tag{2.17}$$

with

$$Q(m,\hat{\lambda}_j) = \left(1 - \prod_{l=1}^{m}(1 - \frac{\hat{\lambda}_j}{\hat{\theta}_l})\right)$$

being the filter factor and $\hat{\theta}_2 > \hat{\theta}_2 > ... > \hat{\theta}_m > 0$ are the eigenvalues of the matrix $\hat{R}$.

The results of Proposition 1 rely on an extension of the Alternative Partial Least Squares (APLS) approach proposed Delaigle and Hall (2012) combined with the results of Proposition 1 by Carrasco and Tsafack (2020). It can also be noticed that if $\hat{\theta}_l = \hat{\theta}_r = \hat{\theta}_0$ for each $l, r = 1, ...m$, then FPLS is similar to FLF with $d = \frac{1}{\hat{\theta}_0}$.

Remark 1.

Considering the previous results, the estimated autoregressive operator $\hat{\Psi}_m^*$ can be summarized as

$$\hat{\Psi}_\delta^*(f) = \frac{1}{N-1} \sum_{j=1}^{N-1} \sum_{n=1}^{N-1} \frac{\hat{Q}(\delta,\hat{\lambda}_j)}{\hat{\lambda}_j} < X_{n+1}, f >< X_n, \hat{v}_j > \hat{v}_j, \quad for\ each\ f \in \mathbb{H}. \tag{2.18}$$

where the filter factor $Q(\delta,\hat{\lambda}_j)$ is such that

$$
\begin{aligned}
Q(\delta, \hat{\lambda}_j) = & Q(m, \hat{\lambda}_j) = I(j \leq m) \quad for\ FPCA\ method \\
& Q(\alpha, \hat{\lambda}_j) = I(\hat{\lambda}_j \geq \alpha) \quad for\ SC \\
& Q(\alpha, \hat{\lambda}_j) = \frac{\hat{\lambda}_j}{\hat{\lambda}_j + \alpha} \quad for\ FT \\
& Q(\alpha, \hat{\lambda}_j) = \left(1 - (1 - d\hat{\lambda}_j)^{\frac{1}{\alpha}}\right) \quad for\ FLF \\
& Q(m, \hat{\lambda}_j) = \left(1 - \prod_{l=1}^{m} (1 - \frac{\hat{\lambda}_j}{\hat{\theta}_l})\right) \quad for\ FPLS.
\end{aligned}
\tag{2.19}
$$

with $\alpha > 0$ and $m < N$.

Given the estimated autoregressive operator $\hat{\Psi}_\delta$, the best prediction of the one day ahead $S\&P$ 500 curve is given by

$$
\hat{X}_{n+1}(t) = \int_0^1 \hat{\psi}_\delta(s,t) X_n(s) ds \quad for\ each\ t\ \in [0,1]. \tag{2.20}
$$

2.5 Asymptotic Results

This section is dedicated to studying the convergence rate of the estimator $\hat{\Psi}_\delta$ in the context that the eigenvalues of the covariance operator K are bounded and decline gradually to zero. This situation is analyzed because as far as we are concerned, it encompasses most of the practical case studied in the economic and financial field. For this purpose, the following assumptions are required :

Assumption 1 (A1) : $\{X_1, ..., X_N\}$ is a sequence of zero-mean and square integrable functions following a functional autoregressive process with $E[||X_n||^4] < +\infty$ and there exists an integer $k_0 \geq 1$ such that $||(\Psi^*)^{k_0}||_{\mathcal{L}} < 1$.

Assumption 2 (A2) : ε_n is stationary and ergodic martingale difference sequence that takes values in $\mathbb{H}$ with respect to $\{\varepsilon_{n-1}, \varepsilon_{n-2}, ..., X_{n-1}, X_{n-2}...\}$ with $E[||\varepsilon_n||^2 | \mathcal{F}_{n-1}] = \sigma^2 < +\infty$, $E[||\varepsilon_n||^4 | \mathcal{F}_{n-1}] < \infty$ and $E[||X_n||^4] < \infty$.

Assumption 3 (A3) : The eigenvalues of the covariance operator K and the estimated one $\hat{K}$ are distinct, i.e. $\lambda_1 > \lambda_2 > ... > 0$ and $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_N > 0$.

Assumption 4 (A4) : There is a Hilbert-Schmidt operator R and a positive constant β such that

$$
\Psi^* = K^{\beta/2} R.
$$

This source condition can also be written as

$$
\sum_{j=1}^{\infty} \frac{< \Psi^*(f), v_j >^2}{\lambda_j^\beta} < +\infty \quad for\ all\ f\ \in\ \mathbb{H}.
$$

Assumption 1 ensures that the sequence $\{X_n; n \in \mathbb{H}\}$ is a stationary process and admits a unique solution. Furthermore, since $E[||X_n||^4] < +\infty$, the operator K is trace-class and thereby is Hilbert-Schmidt.

Assumption 2 imposes that the sequence of innovations ε_n is homoskedastic and ensures that the operators K and D are consistently estimated by $\hat{K}$ and $\hat{D}$, respectively. Moreover, it is assumed that the errors ε_n are martingale difference sequences, which is less restrictive than what is usually observed in preceding papers.

Under assumption 3, the eigenvalues λ_j are distinct. Under A2 and A3, the λ_j are consistently estimated by $\hat{\lambda}_j$ (see Lemma 2 by Kokoszka and Reimherr (2013)). This condition guarantees that the null space of the covariance operator K, $\mathcal{N}(K)$ is such that $\mathcal{N}(K) = 0$.

Assumption 4 is a source condition ensuring that the Fourier coefficients $< \Psi^*(f), v_j >$ go to zero not faster than eigenvalues $\lambda_j^{\beta/2}$ as j goes to infinity. β is also interpreted as a parameter used to control the smoothness of Ψ^*. As β gets larger, $\Psi^*(f)$ becomes smoother (see Carrasco et al. (2007) and Benatia et al. (2017)). In the inverse problem literature, this parameter characterizes the severity of the ill-posed problem. As β becomes larger, the ill-posed problem becomes more severe, i.e the eigenvalues λ_j decay more faster (see proposition 3.13 of Engl et al. (1996)). This assumption is necessary to control the rate of convergence of the bias and variance term as a function of β. This assumption is different to the one considered by Imaizumi and Kato (2018) or Crambes et al. (2013), where they considered the fixed design model and their assumption is related to the decreasing rate of the eigenvalues λ_j. This paper does not use such assumptions. The source conditions considered in this paper are more general than the one considered by Imaizumi and Kato (2018) or Crambes et al. (2013).

Let denote the regularized version of Ψ^* by Ψ^*_δ where δ is α for the FT, FLF, FSC and m for FPCA, FPLS methods. Then, for each function $f \in \mathbb{H}$, Ψ^*_δ can be written as

$$\begin{aligned}
\Psi^*_\delta(f) &= \sum_{j=1}^{\infty} \frac{Q(\delta, \lambda_j)}{\lambda_j} < D(f), v_j > v_j \\
&= \sum_{j=1}^{\infty} \frac{Q(\delta, \lambda_j)}{\lambda_j} < K(\Psi^*)(f), v_j > v_j \\
&= \sum_{j=1}^{\infty} Q(\delta, \lambda_j) < \Psi^*(f), v_j > v_j.
\end{aligned}$$

Thus, for each function f,

$$\hat{\Psi}^*_\delta(f) - \Psi^*(f) = \{\hat{\Psi}^*_\delta(f) - \Psi^*_\delta(f)\} + \{\Psi^*_\delta(f) - \Psi^*(f)\}$$

where $\{\Psi^*_\delta(f) - \Psi^*(f)\}$ represents the bias term that goes to zero as δ increases and

$\{\hat{\Psi}^*_\delta(f) - \Psi^*_\delta(f)\}$ and the estimation error term which may increase as δ increases. The conditional MSE is defined by

$$MSE = E\left[||\hat{\Psi}^*_\delta - \Psi^*||^2_{HS}|\mathcal{F}_{N-1}\right].$$

Proposition 2

Under assumptions A1 - A4, if $\alpha^2 N \to \infty$, then

$$\mathbb{E}\left[\left|\left|\hat{\Psi}^*_\delta - \Psi^*\right|\right|^2_{HS}\middle|\mathcal{F}_{N-1}\right] = \begin{cases} O_p\left(\alpha^\beta\right) + O_p\left(\frac{1}{\alpha^2 N}\right) & \text{for FLF and FSC} \\ O_p\left(\alpha^{min\{\beta,2\}}\right) + O_p\left(\frac{1}{\alpha^2 N}\right) & \text{for FT.} \end{cases}$$

Remarks 2.

— Proposition 2 shows that as α goes to zero, the squared bias term decreases while the variance increases. Then α should be optimally chosen such that bias is equal to the variance. Thus, at the optimality,

— If $\alpha \sim N^{-1/(2+\beta)}$, then $MSE \sim N^{\frac{-\beta}{2+\beta}}$ for FLF and FSC.

— For $\beta < 2$, if $\alpha \sim N^{-1/(2+\beta)}$, then $MSE \sim N^{\frac{-\beta}{2+\beta}}$ for FT.

— For $\beta > 2$, if $\alpha^2 \sim N^{-1/2}$, then $MSE \sim N^{-1/2}$ for FT.

— These results lead to the conclusion that FT, FLF, and FSC display the same convergence rate when the signal is difficult to recover ($\beta < 2$), while FT is slower than FLF and FSC when the signal is easy to recover ($\beta > 2$).

— Due to the saturation property (see Carrasco et al. (2007) and chapter 6 of Engl et al. (1996) concerning the saturation property of the Tikhonov regularization) of the FT method, the FLF and FSC approaches should be preferred to FT (see Carrasco et al. (2007)) in terms of estimation. This pattern should be checked in the simulation.

Proposition 3

Under assumptions A1 - A4,

$$\mathbb{E}\left[\left|\left|\hat{\Psi}^{*,PLS}_m - \Psi^*\right|\right|^2_{HS}\middle|\mathcal{F}_{N-1}\right] \leq \mathbb{E}\left[\left|\left|\hat{\Psi}^{*,PCA}_m - \Psi^*\right|\right|^2_{HS}\middle|\mathcal{F}_{N-1}\right]$$

Additionally, if m diverges much slower than N, such that $\frac{m^2}{N\theta^2_m} \to 0$, then

$$\mathbb{E}\left[\left|\left|\hat{\Psi}^*_\delta - \Psi^*\right|\right|^2_{HS} \middle| \mathcal{F}_{N-1}\right] = \begin{cases} O_p\left(\lambda^\beta_{m+1}\right) + O_p\left(\frac{m}{\lambda_m N}\right) & \text{for FPCA} \\ O_p\left(\lambda^\beta_{m+1}\right) + O_p\left(\frac{m^2}{\theta^2_m N}\right) & \text{for FPLS} \end{cases}$$

where θ_m is the smallest root of the residual polynomial $Q_{m,j}$ [4]*. The first O_p term represents the squared bias and the second one is the estimation error term.*

Remarks 3.

— The first result of Proposition 3 shows that the squared bias term obtain with FPLS is smaller than the one obtained with FPCA.

— Proposition 3 shows that as m increases, the squared bias term decreases while the variance increases. Then, m should be optimally chosen i.e. such that bias is equal to the variance. To get more information about the optimal number of functional components m for the FPCA and FPLS, it is necessary to set some additional assumptions on the eigenvalues and the smoothness condition of the autoregressive operator.

— The rate of convergence of FPCA depends on the decreasing rate of the eigenvalues $(\lambda_j)_{j\geq 1}$ of the covariance operator K and therefore depends on the smallest selected eigenvalue λ_m. On the other hand, the FPLS approach depends on the smallest root θ_m of the residual polynomial $Q_{m,j}$. θ_m is also called the smallest eigenvalue of a Hankel Matrix (see Delaigle and Hall (2012)).

— Since θ_m decreases at an exponential rate (see Berg and Szwarc (2011)), it is most of the time expected that the FPLS method presents a larger estimation error of the autoregressive operator estimation than the FPCA method. Furthermore, under some smoothness conditions of the covariance operator and the autoregressive operator, FPCA and FPLS may display the same rate of convergence.

— In contrast to the FPCA and FPLS methods, the rate of convergence with the FSC, FLF, and FT methods do not depend on the configuration of the eigenvalues.

— The convergence rate derived for FPCA and FPLS are more general bound. Both methods display the same upper bound rate for the squared bias while the variance term of FPLS tends to be larger than that of FPCA. The condition $\frac{m^2}{N\theta^2_m} \to 0$ is sufficient for both FPCA and FPLS since $\theta_j < \lambda_j$ for $j \leq m$ (see for instance Lingjaerde and Christophersen (2000) and Carrasco and Tsafack (2020)).

— The rate obtained for the FPCA is different to the one obtained by Imaizumi and Kato (2018). In fact, they considered an i.i.d fixed design model and imposed more restrictive assumptions on the decreasing rate of the eigenvalues λ_j and on

4. For more discussions about θ_m, see Carrasco and Tsafack (2020)

the smoothness of the kernel operator. The results of this paper are also different to the one obtained by Crambes et al. (2013). They also considered an i.i.d model and assumed that the eigenvalues belong to a class of nonnegative decreasing convex functions of the incrementation index, which is still more restrictive than the one proposed in this paper. Moreover, they proposed to estimate directly the operator $\int_0^1 \psi(s,.)X(s)ds$ instead of $\psi(s,t)$. Furthermore, this paper considers an autoregressive model with the error term that is a functional martingale difference process, which is not considered by other papers.

— The assumptions of this paper are similar to the one suggested by Benatia et al. (2017), who developed the convergence rate of the estimated operator in the context of i.i.d data. The model considered in this paper is different to theirs since there is dynamics between observations. In addition, I consider more methods for comparison purposes.

2.6 Asymptotic normality for a fixed value of the tuning parameter

In this section, the asymptotic normality is derived for the simple FAR(1) for a fixed value of δ. The general asymptotic normality result has been presented for the case of i.i.d model setting with a PCA-based estimation approach (see Crambes et al. (2013), Cardot et al. (2007), and Bosq (2000)). More recently, Benatia et al. (2017) derived this result by using the FT method for a fully functional linear regression. They also considered an i.i.d observations. This paper considered that the error term $(\varepsilon_n)_{n=1,...,N}$ is a sequence of functional martingale difference in $\mathbb{H}$ such that $\mathbb{E}[\varepsilon_n|\mathcal{F}_{n-1}] = 0$ and $\mathbb{E}[||\varepsilon_n||^2|\mathcal{F}_{n-1}] = \sigma^2$. The asymptotic normality is only considered for FPCA, FT, FLF, and FSC method. The FPLS is not considered since $\hat{\Psi}_\delta$ is nonlinear in terms of the response function and therefore is more difficult to address.

Proposition 4

Assume that A1 to A4 hold. If $E[||X_i||^4] < \infty$, $E(||X_n||^2||\varepsilon_{n+1}||^2) < \infty$ *and* δ *is fixed, then*

$$\sqrt{n}(\hat{\Psi}^*_\delta - \Psi^*_\delta) \overset{d}{\Longrightarrow} \mathcal{N}(0, \Omega_\delta) \quad as\ N \to \infty \tag{2.21}$$

where $\delta = m$ *for FPCA and* $\delta = \alpha$ *for FT, FLF, FSC and* Ω_δ *is the covariance operator with kernel :*

$$\Omega_\delta = K_\delta^{-1}\mathbb{E}\Big[(X_i \otimes \varepsilon_{i+1})\tilde{\otimes}(\varepsilon_{i+1}\otimes X_i)\Big]K_\delta^{-1} + K_\delta^{-1}(\Psi^*)\mathbb{E}\Big[(X_i\otimes X_i - K)\tilde{\otimes}(X_i\otimes X_i - K)\Big]K_\delta^{-1}(\Psi^*)$$

with $\tilde{\otimes}$ being the tensor product of two operators. Then, for $(A, B) \in \mathbb{H}^{\mathbb{H}} \times \mathbb{H}^{\mathbb{H}}$, $A\tilde{\otimes}B$ is

an element of the Hilbert space of operators from $\mathbb{H}^{\mathbb{H}}$ to $\mathbb{H}^{\mathbb{H}}$.

This result can be useful for testing hypotheses on the shape of the autoregressive operator in one hand and construct the confidence set for the predicted functions in another hand. This asymptotic covariance operator is estimated by replacing the expectation with their empirical version.

Furthermore, this result is different to the one proposed by Crambes et al. (2013) and Cardot et al. (2007), since they derived the asymptotic normality for the predicted response function directly instead of the estimated operator. Also, it is a generalized version to the one proposed by Benatia et al. (2017) for the case of FT estimation method in the i.i.d context. This result is close to the i.i.d model. The main difference relies on the usage of the functional central limit theorem for martingale difference sequences. Moreover, the conditions $E[||X_i||^4] < \infty$ and $E(||X_i||^2||\varepsilon_{i+1}||^2) < \infty$ ensure that the asymptotic covariance operators display finite values.

Corollary 1
Assume that A1 to A4 hold. If $E[||X_i||^4] < \infty$, $E(||X_i||^2||\varepsilon_{i+1}||^2) < \infty$ *and* δ *is fixed, then under the hypothesis that* $\Psi^* = 0$

$$\sqrt{N}(\hat{\Psi}^*_\delta - 0) \overset{d}{\Longrightarrow} \mathcal{N}(0, \Omega_\delta) \quad as\ N \to \infty \tag{2.22}$$

with

$$\Omega_\delta = K_\delta^{-1}\mathbb{E}\Big[(X_i \otimes \varepsilon_{i+1})\tilde{\otimes}(\varepsilon_{i+1} \otimes X_i)\Big]K_\delta^{-1}.$$

This result is obtained by replacing $\Psi^* = 0$ in the result of Proposition 4. Indeed, we have

$$\Psi^*_\delta = K_\delta^{-1}D = K_\delta^{-1}K\Psi^*$$

Based on this result, one can develop a χ^2 test in order to test the significance of the estimated operator. Empirically, one can test the null hypothesis using $\hat{C}_{x\varepsilon} = \hat{D} - \hat{K}\Psi^*$. Under H_0, this operator should be close to zero and we obtain

$$\sqrt{N}\hat{C}_{x\varepsilon} = \sqrt{N}(\hat{D} - \hat{K}\Psi^*) \overset{d}{\Longrightarrow} \mathcal{N}(0, K_{x\varepsilon}) \quad as\ N \to \infty \tag{2.23}$$

where $K_{x\varepsilon} = \mathbb{E}\Big[(X_i \otimes \varepsilon_{i+1})\tilde{\otimes}(\varepsilon_{i+1} \otimes X_i)\Big]$. Based on this result, it is easy to develop a related chi-square test to check the significance of the estimated operator.

2.7 Determining the optimal order of a FAR(p) model

Determining the optimal order of an AR(p) model has been widely discussed in the standard context of time series models. But so far, in the context of functional time series, there is still a lot of work to be done. In fact, only three papers have been identified.

Kokoszka and Reimherr (2013) are the first to propose a PCA-based multistage testing procedure. They argued that there is no necessity of testing this procedure for a large number of p, since each curve $X_i(t)$ already contains a large number of scalar observations. Based on that statement, they considered a maximum lag of $p_{max} = 2$. Similarly, Aue et al. (2015) proposed to project the data on a set of functional principal components and used a standard VAR(1) model on the projection coefficients in order to derive a criterion to optimally choose simultaneously the number of principal components and the order p. They showed how standard multivariate models can be used in the context of functional time series. On the same line, Liu et al. (2016) proposed an F-test by projecting the data on sieve basis for a convolutional functional autoregressive model of order p (CFAR(p)). This paper proposes a generalized approach based on the regularized estimated operator and therefore is adaptable for different estimation methods that is linear in terms of the response function.

Let us consider the FAR(p) model.

$$X_{n+1} = \Psi_1(X_n) + ... + \Psi_p(X_{n-p+1}) + \varepsilon_{n+1}. \tag{2.24}$$

This equation can be transformed into a FAR(1) model

$$Y_{n+1} = \Phi(Y_n) + U_{n+1} \tag{2.25}$$

where,

$Y_{n+1} = [X_{n+1}, X_n, ..., X_{n-p+1}]'$, $Y_n = [X_n, X_{n-1}, ..., X_{n-p}]'$, $U_{n+1} = [\varepsilon_{n+1}, 0, ..., 0]'$.

and

$$\Phi = \begin{pmatrix} \Psi_1 & ... & \Psi_{p-1} & \Psi_p \\ Id & & & 0 \\ & ... & & ... \\ & & Id & 0 \end{pmatrix}$$

where Id and 0 are respectively the identity and the zero operators on $\mathbb{H}$. Y_n is a $p-$vector of functions that belongs to the space $\mathbb{H}_p = (L^2[0,1])^p$. Φ is a matrix of operators that belongs to $(\mathbb{H}_p \times \mathbb{H}_p)$. $\mathbb{H}_p$ is a Hilbert space endowed with the inner product $< x, y >_p = \sum_{\ell=1}^{p} < x_\ell, y_\ell >$ and the norm $||x||_p = \sqrt{< x, x >_p^2}$ (where $x, y \in \mathbb{H}_p$). For the same reasons as in the previous section, the hypothesis testing is only considered for FPCA, FT, FLF, and FSC method.

The hypothesis testing is a sequential procedure in which for $h = 1, ..., p$, one test if the model has h lags.

$$\begin{cases} H_0 : \Psi_h(.) = 0 \\ H_1 : \Psi_h(.) \neq 0. \end{cases}$$

The null hypothesis H_0 means that only $h-1$ lags are necessary while the alternative

H_1 means that there is h lags.

If at this stage H_0 is rejected, we go to step $h+1$. In contrast, if there is not enough evidence to reject H_0, we stop the procedure and the number of lags is $h-1$. At step h, under H_0, the operator $\hat{C}_{x\varepsilon} = \hat{D} - \sum_{\ell=0}^{h-1} \hat{K}_\ell(\Psi_h^*)$ should be close to zero as $X_i \otimes \varepsilon_{i+1}$ are martingale difference functions, we obtain

$$\sqrt{N}(\hat{C}_{x\varepsilon}) \overset{d}{\Longrightarrow} \mathcal{N}(0, K_{x\varepsilon}) \quad as\ N \to \infty \tag{2.26}$$

with

$$K_{x\varepsilon} = \mathbb{E}\Big[(X_i \otimes \varepsilon_{i+1})\tilde{\otimes}(X_i \otimes \varepsilon_{i+1})\Big] = K\tilde{\otimes}V_\varepsilon,$$

$$\hat{K}_h = \frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_{i-h}$$

and

$$V_\varepsilon = \mathbb{E}[\varepsilon_i \otimes \varepsilon_i]$$

Proposition 5
Assume that A1 to A4 hold. If $E[||X_i||^4] < \infty$, $E(||X_i||^2||\varepsilon_i||^2) < \infty$ *and* δ *is fixed (for FPCA, FT, FLF, and FSC), then under* H_0, *the statistics of the test is given by*

$$W_N = N||\hat{C}_{x\varepsilon}||^2_{HS} \overset{d}{\Longrightarrow} \sum_{\ell=1}^{\infty} \frac{Q(\delta,\lambda_\ell)^2}{\lambda_\ell \sigma^2}\chi^2_\ell(1) \quad as\ N \to +\infty \tag{2.27}$$

where λ_ℓ *is the eigenvalue of* K. *Under* H_1

$$W_N \overset{d}{\Longrightarrow} +\infty \quad as\ N \to +\infty \tag{2.28}$$

where $\chi^2_\ell(1)$ *are independent and identically distributed* $\chi^2(1)$ *random variables.*

The asymptotic distribution of W_N is nothing else than a weighted sum of independent and identically distributed $\chi^2(1)$ with the weights represented by the eigenvalues of the covariance operators $\hat{K}_h$. Then, Proposition 5 shows that H_0 is rejected if $W_N > \sum_{\ell=1}^{\infty} \frac{Q(\delta,\lambda_\ell)^2}{\lambda_\ell\sigma^2}\chi^2_\ell(1)$, for a given significance level a. It is also possible to compute directly the p-value associated with W_N. Following Kokoszka and Reimherr (2013), I use only 2 steps in practice for the test procedure. Then, the maximum number of lag considered is $p_{max} = 2$.

2.8 Data driven selection of the tuning parameter

From the previous sections, it is observed that the different estimation methods suggested in this paper depend on a tuning parameter that is m for FPCA and FPLS and α for FSC, FT, and FLF. Those parameters should be selected properly. Usually, this parameter is chosen in such a way that the prediction error is minimized. Because one deals with functional time series, it is proposed to choose the regularization parameter in such a way that the mean squared prediction error (MSPE) is minimized.

$$\underset{\delta \in A(\delta)}{min} \frac{1}{N} \sum_{n=1}^{N-1} \int_0^1 \left[X_{n+1}(t) - \int_0^1 \hat{\psi}_\delta(t,s) X_n(s) ds \right]^2 dt.$$

where $A(\delta)$ is a set of δ values in which the good one should be selected. δ equal to m for FPCA and FPLS, while $\delta = \alpha$ for FLF, FT, and SC.

An alternative criterion could be the usage of the mean absolute prediction error (MAPE) defined as follows

$$\underset{\delta \in A(\delta)}{min} \frac{1}{N} \sum_{n=1}^{N-1} \int_0^1 \left| X_{n+1}(t) - \int_0^1 \hat{\psi}_\delta(t,s) X_n(s) ds \right| dt.$$

It is also possible to consider the average out-of-sample R^2 (AR^2_{oos}) defined as

$$\underset{\delta \in A(\delta)}{max} \frac{1}{N} \sum_{n=1}^{N-1} \int_0^1 R^2_{oos}(t) dt.$$

where

$$R^2_{oos}(t) = 1 - \frac{\sum (X_{n+1}(t) - \hat{\Psi}_\delta(X_n)(t))^2}{\sum (X_{n+1}(t) - \bar{X}_{n+1}(t))^2}$$

This statistic is usually considered to evaluate the return predictability of a considered security based on a predefined prediction model. It takes its values in the interval $(-\infty, 1]$. Then, if the value $R^2_{oos}(t)$ is close to 1, then the return of the considered security is more predictable at the time t of the day, while if the value of $\tilde{R}^2_{oos}(t)$ is going to $-\infty$, then the return of the considered security is less predictable at the time t of the day. The idea is to find the optimal tuning parameter such that the average R^2_{oos} is maximized in this case. The optimal tuning parameter is derived via a "rolling" scheme, in which the training and validation sample shift progressively forward with the new data to be considered. Then, for each rolling window, the related training and the validation sample are used to choose the optimal regularization parameter and the predictive performance of the model is tracked on the hold out sample. The advantage of this approach is to take into account the most recent information.

2.9 Simulation Results

This section is devoted to comparing the performance of the described estimation methods in a finite sample context. The comparisons are made in terms of Monte Carlo Simulations. The main comparisons are done on the mean-square error of the estimated autoregressive operator and the mean-square prediction error of the model. The model setting is the FAR(1)

$$X_{n+1}(t) = \int_0^1 \psi(t,s)X_n(s)ds + \varepsilon_{n+1}(t) \;\; n = 1, ..., N. \tag{2.29}$$

Three error processes are considered where the second and the third have been used by Didericksen et al. (2012). Let $\varepsilon^{(1)}(t)$, $\varepsilon^{(2)}(t)$, and $\varepsilon^{(3)}(t)$ be as follows :

$$d\varepsilon^{(1)}(s) = -\varepsilon^{(1)}(s)ds + \sigma dW(t)$$

where W is the standard Wiener process generated as and $\sigma = 1$.

$$W(\frac{b}{B}) = \frac{1}{\sqrt{B}} \sum_{\ell=1}^{b} Z_\ell \;\; b = 1, ..., B,$$

and Z_ℓ are independent standard normal variables and $Z_0 = 0$.

$$\varepsilon^{(2)}(t) = \xi_1\sqrt{2}sin(2\pi t) + \xi_2\sqrt{2\kappa}cos(2\pi t). \tag{2.30}$$

where ξ_1 and ξ_2 are two independent variables following a normal distribution and κ can be a constant (for the simulations $\kappa = 0.5$).

The third error term configuration is a combination of the previous ones.

$$\varepsilon^{(3)}(t) = a\varepsilon^{(1)}(t) + (1-a)\varepsilon^{(2)}(t). \tag{2.31}$$

where $a \in [0,1]$ is a real constant that represents the strength of the two components ξ_1 and ξ_2. $\varepsilon^{(1)}(t)$ is an infinite series expansion , $\varepsilon^{(2)}(t)$ is a finite series expansion, and $\varepsilon^{(3)}(t)$ is the combination of the previous one. In the simulations I use $a = 0.5$.

The theoretical autoregressive operator Ψ is an integral operator mapping from $\mathbb{H}$ to $\mathbb{H}$. Two configurations of Ψ are considered, which are :
Model 1 : Gaussian operator (see Didericksen et al. (2012))
$\Psi(s,t) = Cexp\Big[-\frac{t^2+s^2}{2}\Big]$,
Model 2 : Factor model operator (see Imaizumi and Kato (2018))
$\Psi(s,t) = \sum_{k=1}^{3} \Psi_{j,k}v_j(s)v_j(t)$,
with $v_1 = 1$, $v_j = \sqrt{2}cos(j\pi t)$, $j \geq 2$; $\Psi_{1,1} = 0.3$ and $\Psi_{j,k} = 4(-1)^{j+k}j^{-\gamma}k^{-\beta}$ for $(j,k) \neq (1,1)$, and $(\beta,\gamma) = (3,3)$,
where $(s,t) \in ([0,1])^2$ and C a constant useful to normalize the autoregressive operator.

The norm of the operator is considered $||\Psi||_{HS} = 0.8$ and $X_0(t) = \varepsilon_0(t)$.

A continuous interval of $[0, 1]$ is considered. This interval consists of 1000 equally-spaced discretization. The sample size considered for the functional time series $N = 1000$. For the numerical integration, the trapezoidal rule is used for all of the operations in the simulation and real data applications. It has been noticed that a good estimation and prediction of the model depend on the choice of the tuning parameter that is, the number of principal components m for the FPCA and FPLS methods and the regularization parameter α for the FT, FLF, and FSC techniques. These parameters are chosen with cross-validation method.

To analyze the estimation error, two criteria are considered. These are the squared error (SE) and the absolute deviation (AD). Those criteria are given by

$$SE = \int_0^1 \int_0^1 \left(\hat{\Psi}(s,t) - \Psi(s,t)\right)^2 dsdt$$

$$AD = \int_0^1 \int_0^1 \left|\hat{\Psi}(s,t) - \Psi(s,t)\right| dt$$

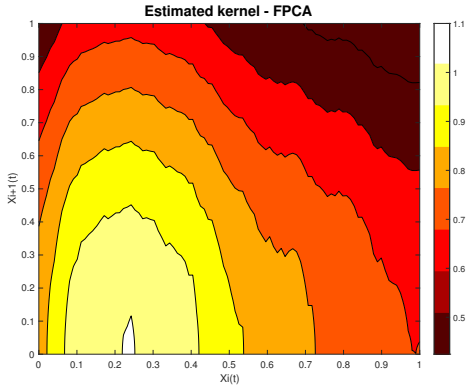
Indeed, we calculate these quantities for each simulation and report the mean, median and standard deviation. To measure the prediction quality, two indicators are considered, which are the integrated squared error (En) and the integrated absolute error (Rn). Those criteria are given by

$$E_n = \int_0^1 \left(\hat{X}_n(t) - X_n(t)\right)^2 dt$$

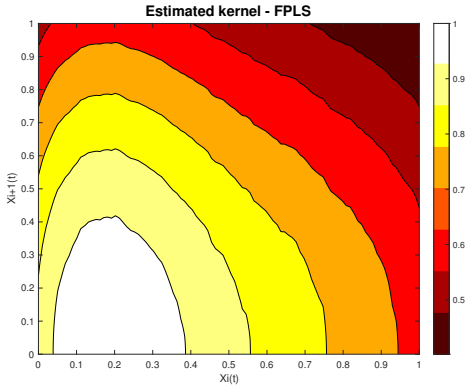
$$R_n = \int_0^1 \left|\hat{X}_n(t) - X_n(t)\right| dt$$

The R^2_{oos} is also considered as a prediction criterion. Similarly to the estimation criteria, we calculate these quantities for each simulation and report the mean, median and standard deviation.

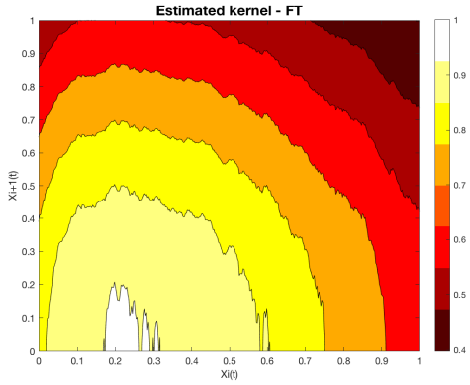
Figure 2.3 shows the estimated kernel via the different estimation methods. It is straightforward to observe that in terms of estimation purpose, the different estimation methods tend to display a shape closed to the true kernel parameter. Moreover, it can be noticed that it is difficult to recover the shape of the kernel representing the relationship between the first tier on the left hand side of the kernels while relationship between the last tier on the right hand side of the kernel is usually well recovered for each method. This pattern is easily observed when looking at the estimation bias on Figures 2.10, 2.4, 2.9 and 2.8. Table 2.1 reports the mean, median, and standard deviation for the different estimation and prediction criteria in the context where the autoregressive operator is the gaussian kernel and the error term is $\varepsilon^{(1)}$. Based on the results of Table 2.1, FPCA tends to outperform the other methods in terms of estimation of the autoregressive operator, while in terms of MSPE and MAPE, it can be observed that all the estimation methods present



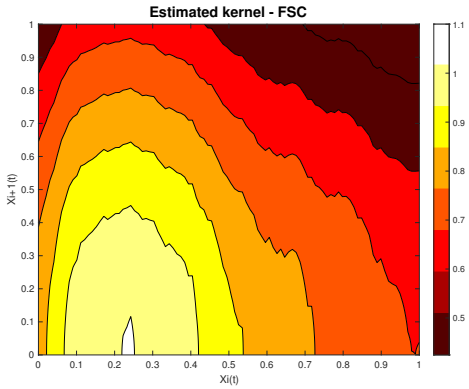
(a) FPCA



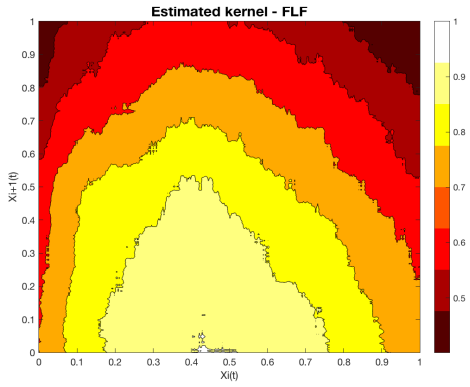
(b) FPLS



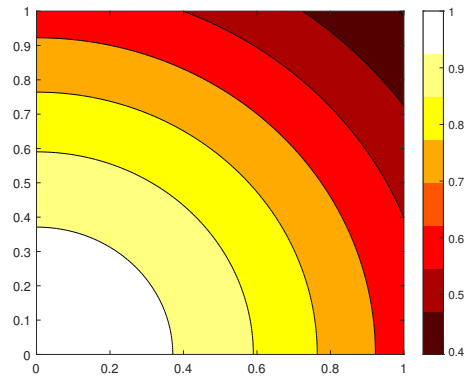
(c) FT



(d) FSC



(e) FLF



(f) True

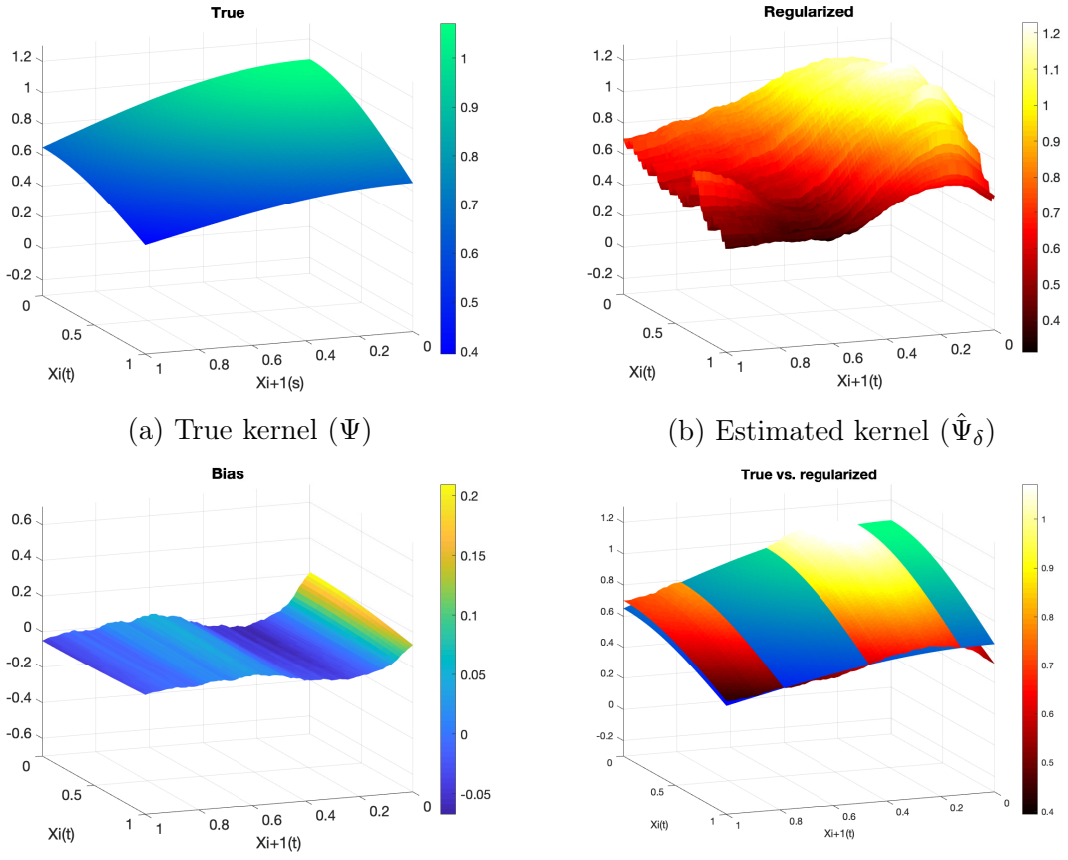
FIGURE 2.3: Estimated gaussian autoregressive operator on the optimal tuning parameter

almost the same predictive performance. This is true for all the performance criteria considered. Moreover, FPCA and FSC methods display similar results. Additionally, it can be observed that the estimated autoregressive operator via these techniques is not close to zero, as documented by Didericksen et al. (2012). Indeed, the approach suggested by the authors is a nonparametric version of the FPCA approach in order to estimate the autoregressive operator. This means that they tend to project the data on some principal component, and do the forecast on the scores as they assume that the scores preserve the stochastic dynamics on the sample of functions. These approaches work well when the sample of functions used in the estimation procedure are smooth enough. They argue that usually the eigenvalues of the covariance operator of the predictor functions drop very quickly to zero. To reduce the quick drop of the eigenvalues, they suggested to add an additional smoothing parameter ; however, the way that parameter is chosen is not discussed. The approach considered in this paper suggests not to project the functions on some orthogonal functions, but to work directly with the functions and operators.

When error terms are less smooth ($\varepsilon^{(2)}$ and $\varepsilon^{(3)}$), the FPLS method tends to outperform the other method when considering the autoregressive operator estimation. The prediction performance is almost the same for all the methods (see Tables 2.6 and 2.7).

Also, when considering the factor-based kernel, the FPLS is still outperforming the other methods when considering the MSE and AD (see Tables 2.8 and 2.9) and the prediction performance is the same for all the methods. Another point to mention is that the FPLS method depends on the decreasing rate of the smallest eigenvalue of the Hankel matrix (θ_m), and its good performance is due to the fact that the eigenfunctions are constructed by taking into account their contribution to predicting the target variable. It would be interesting to see how these methods perform on real data. Then, in overall, the FPLS method tends to outperform the other methods in terms of estimating the autoregressive operator, while the prediction performance is almost the same for the considered methods.

Furthermore, when comparing FT and FLF methods, it can be observed that FT method tends to perform better than FLF method. This is true for estimation and prediction performance almost all the data generating processes considered. This result is in contrast to what was guessed by Benatia et al. (2017).



(a) True kernel (Ψ)

(b) Estimated kernel ($\hat{\Psi}_\delta$)

(c) Bias

(d) True and estimated

FIGURE 2.4: Comparison of the true gaussian and estimated kernels - FPCA

Concerning the selection of the optimal number of functional components, the FPLS tends to select fewer number of components than FPCA, whether the object is estimation or prediction. In fact, the FPCA usually select 10 components in general while FPLS tend to select 1 to 3 components when the purpose is estimation. If the main purpose is to predict, the number of components is 3 in major cases for both FPCA and FPLS (see Table 2.3).

TABLE 2.1: Comparison of the different estimation techniques. Gaussian kernel, $N = 1000$, $M = 1000$ replications, and $\varepsilon^{(1)}$

	Moments	FPCA	FPLS	FT	FLF	FSC
MSE	Mean	0.111	0.355	0.219	0.119	0.111
	Std	0.026	0.157	0.028	0.012	0.026
	Median	0.110	0.322	0.215	0.118	0.110
AD	Mean	0.294	0.504	0.412	0.306	0.294
	Std	0.035	0.105	0.026	0.011	0.020
	Median	0.293	0.476	0.409	0.303	0.293
MSPE	Mean	0.020	0.028	0.020	0.020	0.020
	Std	0.001	0.003	0.001	0.001	0.001
	Median	0.020	0.028	0.020	0.020	0.020
MAPE	Mean	0.023	0.034	0.023	0.023	0.023
	Std	0.001	0.002	0.001	0.001	0.001
	Median	0.023	0.034	0.023	0.023	0.023
R^2_{is}	Mean	0.49	0.49	0.47	0.49	0.49
	Std	0.025	0.027	0.025	0.025	0.025
	Median	0.49	0.49	0.47	0.49	0.49
R^2_{oos}	Mean	0.49	0.49	0.47	0.49	0.49
	Std	0.025	0.027	0.025	0.025	0.025
	Median	0.49	0.49	0.47	0.49	0.49

2.10 Application to the $S\&P$ 500 intraday data

2.10.1 Data

The $S\&P$ 500 Index data is used to analyze the intraday returns predictability. The sample data considered is from 01/01/2010 to 12/31/2017. The data is collected from a website called www.backtestmarket.com.

2.10.2 The Model

To start the empirical analysis, the simple functional autoregressive model is considered where the current cumulative intraday market return is used to predict the next day cumulative return. These results would be tested for the other years of our data base. In the prediction sample, the regression model is given by

$$X_{n+1}(t) = \Psi_0(t) + \int_0^1 \psi(s,t) X_n(s) ds + \varepsilon_{n+1}(t), n = 1, ..., 2013. \tag{2.32}$$

The sample size for this regression period is $N = 2013$. This sample is splitted in 3

TABLE 2.2: Comparison of the different estimation techniques. factor kernel, $N = 1000$, $M = 1000$ replications, and $\varepsilon^{(1)}$

	Moments	FPCA	FPLS	FT	FLF	FSC
MSE	Mean	0.327	0.741	0.240	0.254	0.327
	Std	0.076	0.236	0.026	0.014	0.076
	Median	0.305	0.720	0.235	0.251	0.305
AD	Mean	0.504	0.735	0.431	0.445	0.504
	Std	0.058	0.117	0.023	0.014	0.058
	Median	0.488	0.730	0.426	0.442	0.488
MSPE	Mean	0.020	0.020	0.020	0.020	0.020
	Std	0.001	0.003	0.001	0.001	0.001
	Median	0.020	0.020	0.020	0.020	0.020
MAPE	Mean	0.023	0.023	0.023	0.023	0.023
	Std	0.001	0.002	0.001	0.001	0.001
	Median	0.023	0.023	0.023	0.023	0.023
R^2_{is}	Mean	0.350	0.352	0.323	0.334	0.35
	Std	0.021	0.021	0.021	0.021	0.021
	Median	0.347	0.350	0.321	0.332	0.347
R^2_{oos}	Mean	0.352	0.353	0.325	0.337	0.352
	Std	0.025	0.022	0.024	0.025	0.025
	Median	0.352	0.353	0.325	0.337	0.352

TABLE 2.3: Comparison of the number of selected components, $N = 1000$, $M = 1000$ replications

	Wiener Process				Smoothed				Combo			
	Gaussian		Factor		Gaussian		Factor		Gaussian		Factor	
	PCA	PLS	PCA	PLS	PCA	PLS	PCA	PLS	PCA	PLS	PCA	PLS
MSE	10	3	10	3	10	3	10	3	10	3	8	2
AD	9	2	8	2	9	1	9	2	1	9	9	3
MSPE	6	3	3	3	1	1	1	1	3	3	3	3
MAPE	5	3	3	3	1	1	1	1	3	3	3	3
R2oos	6	3	3	3	1	1	1	1	3	3	3	3

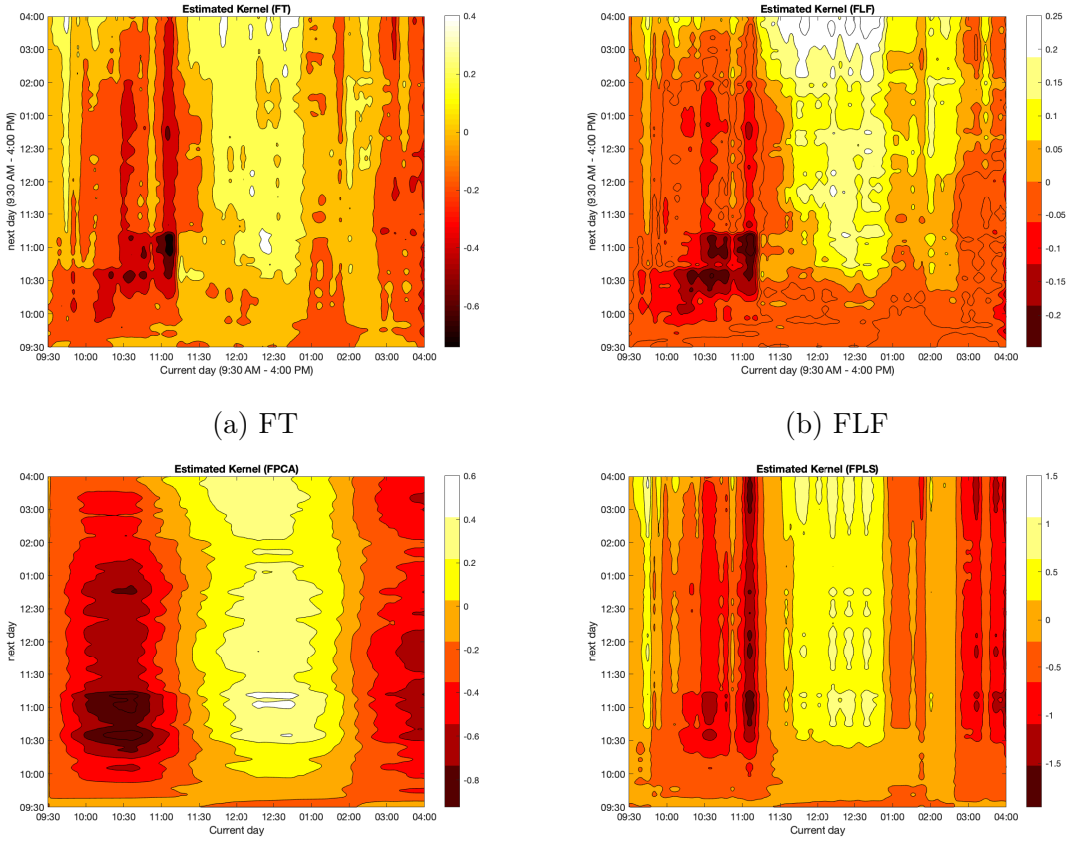
sub-samples using the rolling scheme as described in section 2.8. The training sample is used for the in-sample estimation of the autoregressive operator. The validation sample is used to select the optimal tuning parameter for the estimation and prediction. The testing sample is used to observe the out-of-sample predictive performance of the different estimation methods. Each day is represented by the 390 discretizations points of 1-minute frequency for a trading day. Figure 2.5 displays a contour plot representing the correlation shape between the current day's cumulative return and the next whole day's return on the 1-minute frequency, that is the estimated autoregressive operator $\hat{\psi}_\delta(s,t)$. It can be observed that all the four different methods display almost similar results in terms of estimation. Table 2.4 shows the results of the test to select the optimal number of lag in the model. The test is sequentially driven and Ψ_1, Ψ_2, Ψ_3 correspond to the case where the FAR(1), FAR(2), and FAR(3) are tested respectively. It can be noticed that only one lag is necessary to fit the data and this result holds for the different regularization methods. Moreover, it also indicates that the estimated autoregressive operator $\hat{\Psi}_\delta$ is significantly different from the zero operator for a significant level $a = 0.05$, so I will consider only one lag for estimation and prediction.

TABLE 2.4: The p-values for $\Psi = 0$ under H_0 with $a = 0.05$

p-values	FPCA	FT	FLF	FSC
Ψ_1	0.00001	0.00001	0.00001	0.00001
Ψ_2	0.081	0.0734	0.0687	0.052
Ψ_3	0.773	0.579	0.316	0.373

Based on the estimations using the different methods it can be noticed that the cumulative return for the period 9 :30 AM - 10 :30 AM of the current day is negatively correlated with the whole next day return. Moreover, the cumulative return in the period 11 :30 AM - 1 :00 PM of the current day is negatively correlated with the next day cumulative return in the period 9 :30 AM - 10 :30 AM and positively correlated to the

return for the whole remaining period of the day. This result holds for all the different methods considered. Additionally, the cumulative return in the period 2 :00 PM - 4 :00 PM of the current day tends to be negatively correlated with the whole next day's cumulative return. This result is more pronounced when we take a look at the estimation with FPCA, FPLS and FT methods. In fact, it means that, if investors usually rely only on the first or last tier of the previous trading day (09 :30 AM - 10 :30 AM or 02 :00 PM - 04 :00 PM respectively) to make their analysis and make trading decisions for the next day, they should use a reversal strategy for the next day. Otherwise, if they only rely on the second tier (11 :00 AM - 01 :00 PM) of the current trading day to make next day's decisions, they should use a reversal strategy only in the period 09 :30 AM - 10 :30 AM and a momentum strategy for the remaining period of the the next day. Furthermore, the estimation via FPCA and FSC displays the same results.



(a) FT

(b) FLF

(c) FPCA and FSC

(d) FPLS

FIGURE 2.5: Estimated autoregressive operator.

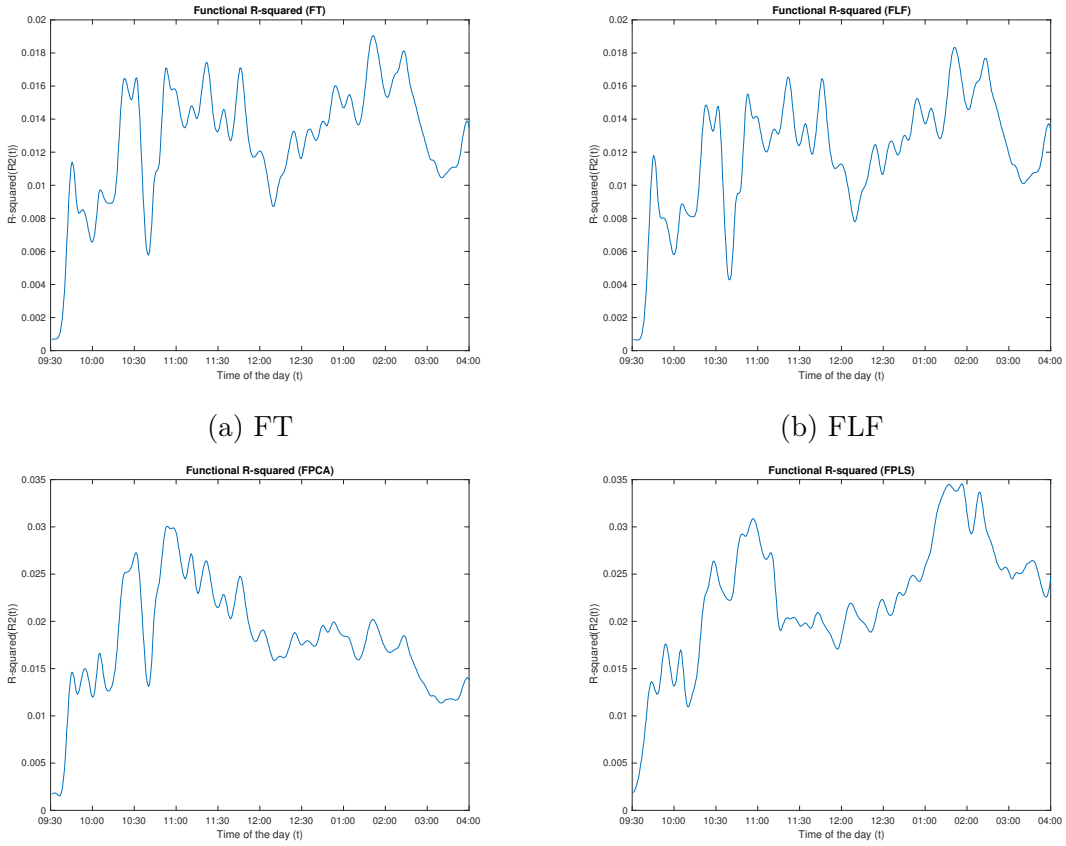
2.10.3 Analyzing the functional out-of-sample R^2

In this section, the predictive functional R^2_{oos} is derived and analyzed. This variable can be used to identify which period of the next day is the most or less predictable periods in terms of returns analysis. From the Figure 2.6, it is easy to see that the most predictable period of the next day based on current day's return is the period 11 :00 AM - 11 :30 AM. This is true for all the considered methods. Indeed, the different estimation methods show a high R^2_{oos} in that period of the day, which ensures that one can expect an edge in that period of the day. The result is almost similar for all of the different estimation methods. FPCA and FPLS tend to catch a remarkable value of 3% in that period of the trading session, while FT and FLF tend to display a predictive R^2_{oos} of 2%. The period 2 :00 PM - 3 :00 PM is also a period to follow-up as a potential predictable one as the different estimation methods tend to display an attractive R^2_{oos} of 2%. This valid only for FPLS, FLF and FT methods. On the other hand, the less predictable period of the next day is the beginning and more specifically the period 09 :30 - 09 :45 as the related predictive R^2_{oos} is closed to 0. This result holds for the different estimation methods considered. This result can be explained by the fact that usually during the pre-market periods, there are news release such as macroeconomic news on the monetary policies or on the business environment that can have an impact on the market right at the opening time. Based on that point, market participant are more uncertain on how the market can react at the opening time. They maybe wait for a few minute at the beginning of the trading session before taking any decision.

2.10.4 Forecast accuracy

In this section, the forecast performance of the considered estimation methods is evaluated. The prediction criteria considered for the comparison are the mean squared prediction error (MSPE), the mean absolute prediction error (MAPE), and the average out-of-sample R-squared (AR^2_{oos}). The performance is evaluated on the test sample with the usage of the optimal tuning parameter obtained on the validation sample. The functional autoregressive model is also compared to the usual AR(1) model on daily frequency. The following table presents the result of the forecasts performance.

Generally, functional autoregressive model tends to produce more forecast accuracy than standard AR(1) model. This suggests that the functional data analysis approach is taking advantage of the additional news and improves the forecast performance. This confirm also the results by Kargin and Onatski (2008). Indeed, they use an estimation method called predictive factor method for their functional autoregressive model. They demonstrate that using the FAR(1) model contribute to improve the forecasting performance compared to the context where the model considered is an ARMA. Furthermore, the different estimation methods tend to display the same predictive performance when we consider the MSPE and MAPE criteria. Indeed, the MSPE is



(a) FT

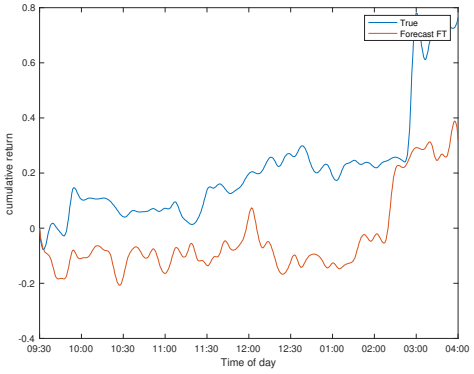
(b) FLF

(c) FPCA and FSC

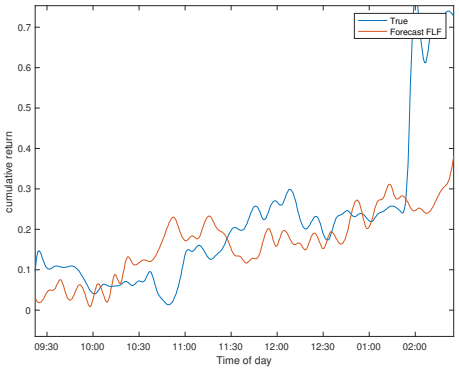
(d) FPLS

FIGURE 2.6: Estimated out-of-sample functional R-squared.

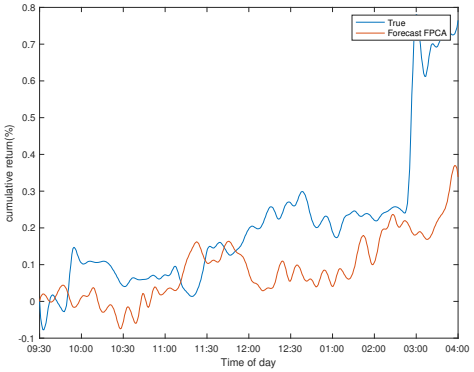
approximately 0.08, while the MAPE is almost 0.19 for all the considered methods. These results are similar to the one obtained in the simulation results. Additionally, when the prediction criterion considered is the AR^2_{oos}, it is easy to observe that FPLS tends to outperform the other methods. In fact, it displays an AR^2_{oos} of 3.9% while the other methods present an AR^2_{oos} of approximately 1%. This supports the fact that FPLS is a supervised methods with the goal to make a good prediction of the response variable. Additionally, Figure 2.7 displays an example of the forecast of the next day return curve in the test sample. It can be observed that when there is no high impact news or jumps, the suggested approach can predict the trend of the curve of the cumulative return, but when there are jumps or high impact announcements, it is difficult to make a good forecast. It could be interesting to introduce variables capturing the news announcements and develop a functional version of the Diebold and Mariano (1995) test in order to check the stability of the forecast performance when the prediction time period changes. Additional materials are required for such an exercise. This goes beyond the scope of this paper.



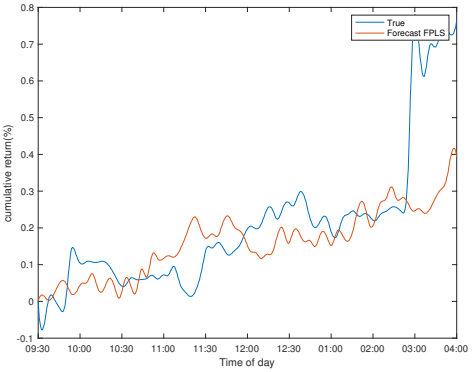
(a) FT



(b) FLF



(c) FPCA and FSC



(d) FPLS

FIGURE 2.7: Forecast of the next day return curve.

TABLE 2.5: Comparison of the forecasting performance of the different methods for $S\&P500$ CIDRs over the testing sample period

	Moments	FPCA	FPLS	FT	FLF	FSC	AR(1)
MSPE	Mean	0.081	0.084	0.081	0.081	0.081	8.551
	Std	0.041	0.0134	0.0162	0.0087	0.042	0.0524
	Median	0.081	0.082	0.081	0.081	0.081	8.532
MAPE	Mean	0.192	0.190	0.191	0.191	0.191	4.328
	Std	0.0543	0.0142	0.0252	0.0109	0.0533	0.0451
	Median	0.192	0.190	0.191	0.191	0.191	4.223
AR^2_{oos}	Mean	0.007	0.039	0.013	0.015	0.007	0.003
	Std	0.0011	0.0020	0.0034	0.0072	0.0011	0.0082
	Median	0.007	0.038	0.013	0.015	0.007	0.003

2.11 Conclusion

This paper investigates the problem of forecasting the stock market intraday return with a functional version of an autoregressive model. The 1-minute frequency data are exploited to construct daily cumulative return curves and the functional data analysis framework is used for estimation purpose. The considered estimation approach is revealed to be interesting for market participants in optimizing their momentum and reversal trading strategy, or in adjusting their portfolio rebalancing. This estimation problem leads to a high dimensionality problem and a comparative analysis of 4 big data techniques including FPLS, FT, FLF, and FSC is developed in order to overcome this issue. The 4 techniques depend on a regularization parameter that is chosen via a data-driven cross-validation approach. I derived the convergence rate of the considered estimation methods. The theoretical results show that the MSE of FT, FLF, and FSC methods display the same convergence rate when the signal to noise is difficult to recover (the parameter is less smooth), while FT is slower than FLF and FSC and can be saturated when it is easy to recover the signal (the parameter is more smooth). On the other hand, FPLS displays a smaller squared bias than the FPCA and the estimation error for the FPLS seems to be much larger than that of FPCA. Also, a testing procedure is developed to select the optimal number of lags in the model and the method is useful regardless of the linear regularization approach.

Monte Carlo simulation results show that in most cases, FPLS methods tends to outperform the others in terms of estimation criteria. On the other hand, the considered methods tend to display the same predictive performance when we consider the MSPE and the MAPE. In addition, when the prediction criterion is the R^2_{oos}, then FPLS and FPCA tend to present the same predictive performance and they usually outperform FT and FLF methods. Moreover, FLF method does not usually outperform FT method, in contrast to what was guessed by Benatia et al. (2017). In overall, the FPLS method tends

slightly to outperform the other methods in most cases in terms of estimation criteria.

The empirical application focuses on the prediction of the next day's cumulative intraday return for the $S\&P$ 500. The results show that all the considered estimation methods tend to display almost similar results in terms of the estimation of the autoregressive operator. In terms of prediction, the considered methods present a similar predictive performance when we consider MSPE and MAPE as criteria, while FPLS method tends to outperform the other methods when we consider the R^2_{oos}. Moreover, concerning the functional R^2_{oos}, the FPLS and FPCA methods tend to document an attractive R^2_{oos} of 3% in the period 11 :00 AM - 11 :30 AM and 2% 2 :00 PM - 3 :00 PM within a trading day, which represent the best predictable period of the next day based on the current day's return and therefore can be considered as potential edge periods in the stock market. Additionally, these periods can be considered as potential time of the day to apply momentum or reversal strategy accordingly. Furthermore, after forecasting the intraday return, it could be interesting to test the prediction accuracy of the different approaches following the idea by Diebold and Mariano (1995). These potential extensions are left for future research.

2.12 Appendix.

Proof of Proposition 1.

For each $t \in [0,1]$, $X_{n+1}(t)$ is a scalar, then one can derive the results by Delaigle and Hall (2012) for each t. The last result is obtained by exploiting the orthogonal polynomial representation as developed by Carrasco and Tsafack (2020).

Lemmas 1 to 3 below will be needed in the proof of Proposition 2.

Lemma 1

Under assumptions A1, A2, and A3, Ψ^*_δ is Hilbert-Schmidt for all δ.

Proof of Lemma 1.

$$||\hat{\Psi}^*_\delta||^2_{HS} = \left|\left|\hat{K}^{-1}_\delta \hat{D}\right|\right|^2_{HS} \leq ||\hat{D}||^2_{HS}||\hat{K}^{-1}_\delta||^2_{op}.$$

If A is Hilbert-Schmidt operator and B is a bounded operator, $||AB|||_{HS} \leq ||A||_{HS}||B||_{op}$ with $||B||_{op} = \underset{||f||\leq 1}{Sup}||B(f)||$ the operator norm, it remains to prove that $||\hat{D}||^2_{HS} < +\infty$ and $||\hat{K}^{-1}_\delta||^2_{op} = O_p(1)$.

If the method is Functional Tikhonov,

$$
\begin{aligned}
||\hat{K}_\delta^{-1}||_{op}^2 &= ||(\alpha I + \hat{K})^{-1}||_{op}^2 \\
&\leq \frac{1}{\alpha^2} \\
&= O_p(1).
\end{aligned}
$$

If the method is Functional Landweber-Fridman,

$$
\begin{aligned}
||\hat{K}_\delta^{-1}||_{op}^2 &= ||d\sum_{l=1}^{1/\alpha}(I - d\hat{K})^{l-1}||_{op}^2 \\
&\leq \underset{j\geq 1}{Sup}\left|\frac{(1-(1-d\hat{\lambda}_j)^{\frac{1}{\alpha}})}{\hat{\lambda}_j}\right|^2 \\
&\leq C\frac{1}{\alpha^2} \\
&= O_p(1).
\end{aligned}
$$

with $||\hat{K}||_{op} < \frac{1}{d}$ and $C > 0$.
If the method is Functional Spectral Cut-off,

$$
\begin{aligned}
||\hat{K}_\delta^{-1}||_{op}^2 &\leq \underset{j\geq 1}{Sup}\left|\frac{I(\hat{\lambda}_j \geq \alpha)}{\hat{\lambda}_j}\right|^2 \\
&\leq \frac{1}{\alpha^2} \\
&= O_p(1).
\end{aligned}
$$

and we get the same result for the FPCA method, that is $||\hat{K}_\delta^{-1}||_{op}^2 = O_p(1)$
If the method is Functional Partial Least Squares

$$
\begin{aligned}
||\hat{K}_\delta^{-1}||_{op}^2 &\leq \underset{j\geq 1}{Sup}\left|\frac{\left(1-\prod_{l=1}^m(1-\frac{\hat{\lambda}_j}{\hat{\theta}_l})\right)}{\hat{\lambda}_j}\right|^2 \\
&\leq \underset{j\geq m+1}{Sup}\left|\frac{\left(1-\prod_{l=1}^m(1-\frac{\hat{\lambda}_j}{\hat{\theta}_l})\right)}{\hat{\lambda}_j}\right|^2 \\
&= O_p(1).
\end{aligned}
$$

using the fact that $\hat{\lambda}_j \leq \hat{\theta}_m$.

Furthermore, $\hat{D}^*$ is a Hilbert-Schmidt operator since it is an integral operator with a degenerated kernel $\hat{D}^*(s,t) = \frac{1}{N-1}\sum_{i=1}^{N-1} X_i(s)X_{i+1}(t)$ and $X_n(t)$ belongs to $\mathbb{H}$.

Lemma 2

Under assumptions $A1 - A4$, $\left|\left|K_\delta^{-1}(K-K_\delta)K^{\beta/2}R\right|\right|_{HS}^2$ *is*

$$\begin{cases} O_p(\alpha^\beta) \; for \; FLF \; and \; FSC \\ O_p(\alpha^{min\{\beta,2\}}) \; for \; FT \\ O_p(\lambda_{m+1}^\beta) \; for \; FPCA \; and \; FPLS. \end{cases}$$

Proof of Lemma 2.

$$\left|\left|K_\delta^{-1}(K-K_\delta)K^{\beta/2}R\right|\right|_{HS}^2 \leq \left|\left|K_\delta^{-1}\right|\right|_{op}^2 \left|\left|(K-K_\delta)K^{\beta/2}R\right|\right|_{HS}^2$$

FT method

For the FT method, see Lemma 10 of Benatia et al. (2017).

FSC method

$$K_\delta = \sum_{\lambda_j \geq \alpha} \lambda_j < v_j, . > v_j$$

$$K - K_\delta = \sum_{\lambda_j < \alpha} \lambda_j < v_j, . > v_j$$

$$K_\delta^{-1} = \sum_{\lambda_j \geq \alpha} \frac{1}{\lambda_j} < v_j, . > v_j$$

Then, $||K_\delta^{-1}||_{op}^2 = O(\frac{1}{\alpha^2})$ and

$$\begin{aligned} \left|\left|(K-K_\delta)K^{\beta/2}R\right|\right|_{HS}^2 &= \sum_{\lambda_j<\alpha} \lambda_j^{\beta+2} < R(v_j), R(v_j) >^2 \\ &\leq \left[Sup_{\lambda_j<\alpha}\lambda_j^{\beta+2}\right] \sum_{\lambda_j<\alpha} < R(v_j), R(v_j) >^2 \\ &\leq C\alpha^\beta. \end{aligned}$$

where $C > 0$ is an arbitrary positive constant. Morevover, $\sum_{\lambda_j<\alpha} < R(v_j), R(v_j) >^2 = ||R||_{HS}^2 < +\infty$ (because R is Hilbert-Schmidt), which leads to the result.

FLF method

$$K_\delta^{-1} = \sum_{j=1}^{\infty} \frac{Q(\alpha, \lambda_j)}{\lambda_j} < v_j, . > v_j$$

where $Q(\alpha, \lambda_j) = (1 - (1 - d\lambda_j)^{\frac{1}{\alpha}})$.

$$\begin{aligned}
\left|\left|(K - K_\delta)K^{\beta/2}R\right|\right|_{HS}^2 &= \sum_{j=1}^{+\infty} \lambda_j^{\beta+2}(1 - Q(\alpha, \lambda_j))^2 < R(v_j), R(v_j) >^2 \\
&\leq \left[Sup_j \lambda_j^{\beta+2}(1 - Q(\alpha, \lambda_j))^2\right] \sum_{j=1}^{+\infty} < R(v_j), R(v_j) >^2 \\
&\leq C\alpha^{\beta+2}.
\end{aligned}$$

Using Proposition 3.11 of Carrasco et al. (2007), $Sup_{j\geq 1}\left[\lambda_j^{\beta}(1-Q(\alpha, \lambda_j))^2\right] \leq C\alpha^{\beta+2}$ (where C is a positive arbitrary constant) and the fact that R is Hilbert-Schmidt leads to the result.

FPCA method

$$K_\delta = \sum_{j=1}^{m} \lambda_j < v_j, . > v_j$$

$$K - K_\delta = \sum_{j\geq m+1} \lambda_j < v_j, . > v_j$$

$$K_\delta^{-1} = \sum_{j=1}^{m} \frac{1}{\lambda_j} < v_j, . > v_j$$

Then, $||K_\delta^{-1}||_{op}^2 = O(\frac{1}{\lambda_{m+1}^2})$ and

$$\begin{aligned}
\left|\left|(K - K_\delta)K^{\beta/2}R\right|\right|_{HS}^2 &= \sum_{j\geq m+1} \lambda_j^{\beta+2} < R(v_j), R(v_j) >^2 \\
&\leq \left[Sup_{\lambda_j<\alpha} \lambda_j^{\beta+2}\right] \sum_{j\geq m+1} < R(v_j), R(v_j) >^2 \\
&\leq C\lambda_{m+1}^{\beta+2}.
\end{aligned}$$

where $C > 0$ is an arbitrary positive constant. Morevover, $\sum_{j\geq m+1} < R(v_j), R(v_j) >^2 = ||R||_{HS}^2 < +\infty$ (because R is Hilbert-Schmidt), which leads to the result.

FPLS method

$$K_\delta^{-1} = \sum_{j=1}^{\infty} \frac{Q(m,\lambda_j)}{\lambda_j} < v_j, . > v_j$$

where $Q(m,\lambda_j) = \left(1 - \prod_{l=1}^{m}(1 - \frac{\lambda_j}{\theta_l})\right)$.

Then,

$$\begin{aligned} ||K_\delta^{-1}||_{op}^2 &= \underset{j\geq 1}{Sup} Q(m,\lambda_j) \\ &\leq \underset{j\geq m+1}{Sup}\ Q(m,\lambda_j) \\ &= O(1). \end{aligned}$$

The second line is true following proofs of Proposition 4 by Carrasco and Tsafack (2020) and

$$\begin{aligned} \left|\left|(K - K_\delta)K^{\beta/2}R\right|\right|_{HS}^2 &= \sum_{j=1}^{+\infty} \lambda_j^\beta (1 - Q(m,\lambda_j))^2 < R(v_j), R(v_j) >^2 \\ &\leq \left[\underset{j}{Sup}\{\lambda_j^\beta (1 - Q(m,\lambda_j))^2\}\right] \sum_{j=1}^{+\infty} < R(v_j), R(v_j) >^2 \\ &\leq \left[\underset{j\geq m+1}{Sup}\ \{\lambda_j^\beta (1 - Q(m,\lambda_j))^2\}\right] \sum_{j=1}^{+\infty} < R(v_j), R(v_j) >^2 \\ &\leq C\lambda_{m+1}^\beta. \end{aligned}$$

Using results of the proofs of Proposition 4 by Carrasco and Tsafack (2020), $\underset{j\geq 1}{Sup}\left[\lambda_j^\beta (1 - Q(\alpha,\lambda_j))^2\right] \leq C\lambda_{m+1}^\beta$ (where C is a positive arbitrary constant) and the fact that R is Hilbert-Schmidt lead to the result. This concludes the proof of Lemma 2.

Lemma 3

Under assumptions A1 − A4, for $N \to \infty$, $\left|\left|\hat{K}_\delta^{-1}\hat{K}(\Psi^*) - K_\delta^{-1}K(\Psi^*)\right|\right|_{HS}^2$ *is*

$$\begin{cases} O_p(\frac{\alpha^\beta}{\alpha^2 N}) \ for\ FLF\ and\ FSC \\ O_p(\frac{\alpha^{min\{\beta,1\}}}{\alpha^2 N}) \ for\ FT \\ O_p(\frac{1}{\lambda_m^2 N}) \ for\ FPCA. \end{cases}$$

Proof of Lemma 3.

$$\begin{aligned}\hat{K}_\delta^{-1}\hat{K}(\Psi^*) - K_\delta^{-1}K(\Psi^*) = &-\hat{K}_\delta^{-1}(\hat{K}_\delta - K_\delta)K_\delta^{-1}(K_\delta - K)(\Psi^*)\\ &- \hat{K}_\delta^{-1}(\hat{K}_\delta - K_\delta)\Psi^*\\ &+ \hat{K}_\delta^{-1}(\hat{K} - K)\Psi^*.\end{aligned}$$

Then,

$$\begin{aligned}\left|\left|\hat{K}_\delta^{-1}\hat{K}(\Psi^*) - K_\delta^{-1}K(\Psi^*)\right|\right|_{HS}^2 &\leq 3\left|\left|\hat{K}_\delta^{-1}(\hat{K}_\delta - K_\delta)K_\delta^{-1}(K_\delta - K)(\Psi^*)\right|\right|_{HS}^2\\ &+ 3\left|\left|\hat{K}_\delta^{-1}(\hat{K}_\delta - K_\delta)\Psi^*\right|\right|_{HS}^2\\ &+ 3\left|\left|\hat{K}_\delta^{-1}(\hat{K} - K)\Psi^*\right|\right|_{HS}^2\\ &= 3(I) + 3(II) + 3(III).\end{aligned}$$

Moreover,

$$\begin{aligned}(I) &= \left|\left|\hat{K}_\delta^{-1}(\hat{K}_\delta - K_\delta)K_\delta^{-1}(K_\delta - K)(\Psi^*)\right|\right|_{HS}^2\\ &\leq \left|\left|\hat{K}_\delta^{-1}\right|\right|_{op}^2\left|\left|\hat{K}_\delta - K_\delta\right|\right|_{op}^2\left|\left|K_\delta^{-1}(K_\delta - K)(\Psi^*)\right|\right|_{HS}^2.\end{aligned}$$

$$\begin{aligned}(II) &= \left|\left|\hat{K}_\delta^{-1}(\hat{K}_\delta - K_\delta)\Psi^*\right|\right|_{HS}^2\\ &\leq \left|\left|\hat{K}_\delta^{-1}\right|\right|_{op}^2\left|\left|\hat{K}_\delta - K_\delta\right|\right|_{op}^2\left|\left|K^{\beta/2}R\right|\right|_{HS}^2.\end{aligned}$$

$$\begin{aligned}(III) &= \left|\left|\hat{K}_\delta^{-1}(\hat{K} - K)\Psi^*\right|\right|_{HS}^2\\ &\leq \left|\left|\hat{K}_\delta^{-1}\right|\right|_{op}^2\left|\left|\hat{K} - K\right|\right|_{op}^2\left|\left|K^{\beta/2}R\right|\right|_{HS}^2.\end{aligned}$$

For FT method :

See proof of Proposition 2 by Benatia et al. (2017).

For FLF and FSC methods :

Furthermore, $||\hat{K}_\delta^{-1}||_{op}^2 = O_p(\frac{1}{\alpha^2})$, $||K_\delta^{-1}||_{op}^2 = O_p(\frac{1}{\alpha^2})$, and from Lemma 2

$$\left|\left|K_\delta^{-1}(K_\delta - K)(\Psi^*)\right|\right|_{HS}^2 = O_p\left(\alpha^\beta\right).$$

Then, $(I) = O_p(\frac{\alpha^\beta}{\alpha^2 N})$. Moreover, $||\hat{K}_\delta - K_\delta||_{op}^2 = O_p(\frac{1}{N})$, $||\hat{K} - K||_{op}^2 = O_p(\frac{1}{N})$
Then, $\left|\left|\hat{K}_\delta^{-1}(\hat{K}_\delta - K_\delta)K_\delta^{-1}(K_\delta - K)(\Psi^*)\right|\right|_{HS}^2 = O_p\left(\frac{\alpha^\beta}{\alpha^2 N}\right)$.

Similarly, $(II) = O_p(\frac{1}{\alpha^2 N})$ and $(III) = O_p(\frac{1}{\alpha^2 N})$. Then, for FLF
$\left|\left|\hat{K}_\delta^{-1}\hat{K}(\Psi^*) - K_\delta^{-1}K(\Psi^*)\right|\right|_{HS}^2 = O_p\left(\frac{\alpha^\beta}{\alpha^2 N}\right)$.

For FPCA method :
$\delta = m$ for FPCA and FPLS

Furthermore, $||\hat{K}_\delta^{-1}||_{op}^2 = O_p(\frac{1}{\lambda_{m+1}^2})$, $||K_\delta^{-1}||_{op}^2 = O_p(\frac{1}{\lambda_{m+1}^2})$, and from Lemma 2 we have

$$\left|\left|K_\delta^{-1}(K_\delta - K)(\Psi^*)\right|\right|_{HS}^2 = O_p\left(\lambda_{m+1}^\beta\right).$$

Then,

$$\left|\left|\hat{K}_\delta^{-1}(\hat{K}_\delta - K_\delta)K_\delta^{-1}(K_\delta - K)(\Psi^*)\right|\right|_{HS}^2 = O_p\left(\frac{\lambda_{m+1}^\beta}{\lambda_{m+1}^2 N}\right).$$

Moreover, $||\hat{K}_\delta - K_\delta||_{op}^2 = O_p(\frac{1}{N})$, $||\hat{K} - K||_{op}^2 = O_p(\frac{1}{N})$, then, $(I) = O_p\left(\frac{1}{\lambda_{m+1}^2 N}\right)$, $(II) = O_p(\frac{1}{\lambda_{m+1}^2 N})$ and $(III) = O_p(\frac{1}{\lambda_{m+1}^2 N})$. This leads to

$$\left|\left|\hat{K}_\delta^{-1}\hat{K}(\Psi^*) - K_\delta^{-1}K(\Psi^*)\right|\right|_{HS}^2 = O_p\left(\frac{\lambda_{m+1}^\beta}{\lambda_{m+1}^2 N}\right) = O_p\left(\frac{1}{\lambda_m^2 N}\right).$$

Proof of Proposition 2.
We have $\delta = \alpha$ for FT, FSC and FLF

$$\begin{aligned}
\mathbb{E}\left[\left|\left|\hat{\Psi}^*_\delta - \Psi^*\right|\right|^2_{HS}\middle|\mathcal{F}_{N-1}\right] &= \mathbb{E}\left[\left|\left|\hat{K}_\delta^{-1}\hat{D} - K_\delta^{-1}D + K_\delta^{-1}D - \Psi^*\right|\right|^2_{HS}\middle|\mathcal{F}_{N-1}\right] \\
&= \mathbb{E}\left[\left|\left|\hat{K}_\delta^{-1}\hat{K}\Psi^* + \hat{K}_\delta^{-1}\hat{C}_{x\varepsilon} - K_\delta^{-1}D + K_\delta^{-1}D - \Psi^*\right|\right|^2_{HS}\middle|\mathcal{F}_{N-1}\right] \\
&= \mathbb{E}\left[\left|\left|\hat{K}_\delta^{-1}\hat{C}_{x\varepsilon} + \hat{K}_\delta^{-1}\hat{K}\Psi^* - K_\delta^{-1}K\Psi^* + K_\delta^{-1}D - \Psi^*\right|\right|^2_{HS}\middle|\mathcal{F}_{N-1}\right] \\
&= \mathbb{E}\Big[||A+B+C||^2_{HS}|\mathcal{F}_{N-1}\Big].
\end{aligned}$$

where $A = \hat{K}_\delta^{-1}\hat{C}_{x\varepsilon}$, with $\hat{C}_{x\varepsilon} = \frac{1}{N-1}\sum_{i=1}^{N-1} X_i \otimes \varepsilon_{i+1}$, $B = \hat{K}_\delta^{-1}\hat{K}\Psi^* - K_\delta^{-1}K\Psi^*$ and $C = K_\delta^{-1}D - \Psi^*$.

Using Lemma 8 by Benatia et al. (2017), the last line of this equation yields

$$\begin{aligned}
\mathbb{E}\left[\left|\left|\hat{\Psi}^*_\delta - \Psi^*\right|\right|^2_{HS}\middle|\mathcal{F}_{N-1}\right] &= \mathbb{E}[||A||^2_{HS}|\mathcal{F}_{N-1}] + ||B+C||^2_{HS} \\
&\leq \mathbb{E}[||A||^2_{HS}|\mathcal{F}_{N-1}] + 2||B||^2_{HS} + 2||C||^2_{HS}.
\end{aligned}$$

Using Lemma 2 lead to $||C||^2_{HS} =$
$\begin{cases} O_p(\alpha^\beta) \; for \; FLF \; and \; FSC \\ O_p(\alpha^{min\{\beta,2\}}) \; for \; FT. \end{cases}$

Following Lemma 3, lead $||B||^2_{HS} =$
$\begin{cases} O_p(\frac{\alpha^\beta}{\alpha^2 N}) \; for \; FLF \; and \; FSC \\ O_p(\frac{\alpha^{min\{\beta,1\}}}{\alpha^2 N}) \; for \; FT. \end{cases}$

Then, it remains to derive the convergence rate of $\mathbb{E}[||A||^2_{HS}|\mathcal{F}_{N-1}]$.

$$\begin{aligned}
\mathbb{E}[||A||^2_{HS}|\mathcal{F}_{N-1}] &= \mathbb{E}[||\hat{K}_\delta^{-1}\hat{C}_{x\varepsilon}||^2_{HS}|\mathcal{F}_{N-1}] \\
&= \mathbb{E}\left[tr[\hat{K}_\delta^{-1}\hat{C}_{x\varepsilon}\hat{C}^*_{x\varepsilon}\hat{K}_\delta^{-1}]\middle|\mathcal{F}_{N-1}\right] \\
&= tr\left[\mathbb{E}[\hat{K}_\delta^{-1}\hat{C}_{x\varepsilon}\hat{C}^*_{x\varepsilon}\hat{K}_\delta^{-1}|\mathcal{F}_{N-1}]\right] \\
&= tr\left[\hat{K}_\delta^{-1}\mathbb{E}[\hat{C}_{x\varepsilon}\hat{C}^*_{x\varepsilon}|\mathcal{F}_{N-1}]\hat{K}_\delta^{-1}\right]
\end{aligned}$$

where $tr(Z)$ is the trace of an arbitrary operator Z, that is $||Z||^2_{HS} = tr(ZZ^*)$. The third line in this equation is true following Lemma 9 by Benatia et al. (2017). Furthermore, since $\hat{C}_{x\varepsilon} = \frac{1}{N-1}\sum_{i=1}^{N-1} X_i \otimes \varepsilon_{i+1}$

$$\hat{C}_{x\varepsilon}\hat{C}^*_{x\varepsilon}(f) = \frac{1}{(N-1)^2}\sum_{i,j=1}^{N-1} X_i < X_j, f >< \varepsilon_{i+1}, \varepsilon_{j+1} >$$

Therefore,

$$\begin{aligned}
\mathbb{E}[\hat{C}_{x\varepsilon}\hat{C}^*_{x\varepsilon}(f)|\mathcal{F}_{N-1}] &= \frac{1}{(N-1)^2}\sum_{i,j=1}^{N-1} X_i < f, X_j > \mathbb{E}[< \varepsilon_{i+1}, \varepsilon_{j+1} > |\mathcal{F}_{N-1}] \\
&= \frac{1}{N-1}\sum_{i=1}^{N-1} < X_i, f > \mathbb{E}[< \varepsilon_{i+1}, \varepsilon_{i+1} > |\mathcal{F}_{N-1}]X_i \\
&= \frac{1}{N-1}\sum_{i=1}^{N-1} < X_i, f > tr(V_\varepsilon)X_i \\
&= \frac{1}{N-1}tr(V_\varepsilon)\sum_{i=1}^{N-1} < X_i, f > X_i \\
&= \frac{1}{N-1}tr(V_\varepsilon)\hat{K}(f)
\end{aligned}$$

where $V_\varepsilon = \mathbb{E}[\varepsilon_i \otimes \varepsilon_i|\mathcal{F}_{N-1}]$. Since ε_i are squared integrable functional martingale difference sequences and v_i are orthonormal, then $\varepsilon_i = \sum_{j=1}^{+\infty} < \varepsilon_i, v_j > v_j$, which leads to $< \varepsilon_i, \varepsilon_i >= \sum_{j=1}^{+\infty} < \varepsilon_i, v_j >^2$. Therefore,

$$\begin{aligned}
\mathbb{E}[< \varepsilon_i, \varepsilon_i > |\mathcal{F}_{N-1}] &= \sum_{j=1}^{+\infty} < V_\varepsilon(v_i), v_i > \\
&= tr(V_\varepsilon).
\end{aligned}$$

Then,

$$\mathbb{E}[||A||^2_{HS}|\mathcal{F}_{N-1}] \leq \frac{1}{N-1}tr\left[\hat{K}_\delta^{-1}\hat{K}\hat{K}_\delta^{-1}\right]tr\left[V_\varepsilon\right].$$

For FT method :

$$
\begin{aligned}
tr\left[\hat{K}_\delta^{-1}\hat{K}\hat{K}_\delta^{-1}\right] &= \sum_{j=1}^{+\infty}\frac{\hat{\lambda}_j}{(\hat{\lambda}_j+\alpha)^2} \\
&\leq \int_0^{+\infty}\frac{x}{(x+\alpha)^2}dx \\
&= \frac{1}{\alpha}\left[\frac{-1}{x+\alpha}\right]_0^{+\infty} \\
&\leq \frac{C}{\alpha^2}.
\end{aligned}
$$

where C is an arbitrary positive number.

For FSC method :

$$
\begin{aligned}
tr\left[\hat{K}_\delta^{-1}\hat{K}\hat{K}_\delta^{-1}\right] &= \sum_{\lambda_j\geq\alpha}\frac{\hat{\lambda}_j}{\hat{\lambda}_j^2} \\
&= \frac{1}{\alpha^2}\sum_{\hat{\lambda}_j\geq\alpha}\hat{\lambda}_j \\
&\leq \frac{C}{\alpha^2}.
\end{aligned}
$$

The last line holds since $\sum_{\lambda_j\geq\alpha}\hat{\lambda}_j < +\infty$ ($\hat{K}$ is nuclear).

For FLF method :

$$
\begin{aligned}
tr\left[\hat{K}_\delta^{-1}\hat{K}\hat{K}_\delta^{-1}\right] &= \sum_{j\geq 1}\frac{(1-(1-d\hat{\lambda}_j)^{1/\alpha})^2}{\hat{\lambda}_j} \\
&\leq \left(\frac{d}{\alpha}\right)^2\sum_{\lambda_j\geq 1}\hat{\lambda}_j \\
&\leq \frac{C}{\alpha^2}
\end{aligned}
$$

where C is an arbitrary positive number. The second line is true following Proposition 3.14 by Carrasco et al. (2007). These results lead to $\mathbb{E}[||A||_{HS}^2|\mathcal{F}_{N-1}] = O_p\left(\frac{1}{\alpha^2 N}\right)$ for FT, FLF and FSC.

The convergence rate of $||C||_{HS}^2$, $||B||_{HS}^2$ and $\mathbb{E}[||A||_{HS}^2|\mathcal{F}_{N-1}]$ leads to

$$\mathbb{E}\left[\left|\left|\hat{\Psi}^*_\delta - \Psi^*\right|\right|^2_{HS}\middle|\mathcal{F}_{N-1}\right] = \begin{cases} O_p\left(\alpha^\beta\right) + O_p\left(\frac{1}{\alpha^2 N}\right) & \text{for the FLF and FSC} \\ O_p\left(\alpha^{min\{\beta,2\}}\right) + O_p\left(\frac{1}{\alpha^2 N}\right) & \text{for the FT.} \end{cases}$$

Proof of Proposition 3.
Following the same argument as in Proposition 1, we have

$$\begin{aligned}\mathbb{E}\left[\left|\left|\hat{\Psi}^*_\delta - \Psi^*\right|\right|^2_{HS}\middle|\mathcal{F}_{N-1}\right] &= \mathbb{E}\left[\left|\left|\hat{K}_\delta^{-1}\hat{D}^* - K_\delta^{-1}D^* + K_\delta^{-1}D^* - \Psi^*\right|\right|^2_{HS}\middle|\mathcal{F}_{N-1}\right] \\ &\leq \mathbb{E}\left[||A||^2_{HS}|\mathcal{F}_{N-1}\right] + 2||B||^2_{HS} + 2||C||^2_{HS}\end{aligned}$$

where $A = \hat{K}_\delta^{-1}\hat{C}_{x\varepsilon}$, with $\hat{C}_{x\varepsilon} = \frac{1}{N-1}\sum_{i=1}^{N-1} X_i \otimes \varepsilon_{i+1}$, $B = \hat{K}_\delta^{-1}\hat{K}\Psi^* - K_\delta^{-1}K\Psi^*$ and $C = K_\delta^{-1}D^* - \Psi^*$.

For FPCA method :

Following Lemma 2, $||C||^2_{HS} = O_p\left(\lambda^\beta_{m+1}\right)$ and using Lemma 3, $||B||^2_{HS} = O_p\left(\frac{1}{\lambda^2_m N}\right)$. Then, it remains to derive the convergence rate of $\mathbb{E}\left[||A||^2_{HS}|\mathcal{F}_{N-1}\right]$

Following the same arguments as in Proposition 2,

$$\mathbb{E}[||A||^2_{HS}|\mathcal{F}_{N-1}] \leq \frac{1}{N-1} tr\left[\hat{K}_\delta^{-1}\hat{K}\hat{K}_\delta^{-1}\right] tr\left[V_\varepsilon\right]$$

and

$$\begin{aligned} tr\left[\hat{K}_\delta^{-1}\hat{K}\hat{K}_\delta^{-1}\right] &= \sum_{j=1}^m \frac{\hat{\lambda}_j}{\hat{\lambda}_j^2} \\ &= \sum_{j=1}^m \frac{1}{\hat{\lambda}_j} \\ &\leq \frac{Cm}{\lambda_m}\end{aligned}$$

where $C > 0$ is an arbitrary constant. Then, $||A||^2_{HS} = O_p\left(\frac{m}{\lambda_m N}\right)$. Combining the rate of convergence for $||A||^2_{HS}$, $||B||^2_{HS}$, $||C||^2_{HS}$ leads to

$$\mathbb{E}\left[\left|\left|\hat{\Psi}^*_\delta - \Psi^*\right|\right|^2_{HS}\middle|\mathcal{F}_{N-1}\right] = O_p\left(\lambda^\beta_{m+1}\right) + O_p\left(\frac{m}{\lambda_m N}\right).$$

For the FPLS method :

Let us define
$\underline{X} = \left[X_1, X_2, ..., X_{N-1}\right]'$
$\underline{Y} = \left[X_2, X_3, ..., X_N\right]'$
and
$\underline{\varepsilon} = \left[\varepsilon_2, \varepsilon_3, ..., \varepsilon_N\right]'$
Let us denote by

$$\Pi_a = Span\{v_j : \lambda_j \leq a\}$$

and

$$\hat{\Pi}_a = Span\{\hat{v}_j : \hat{\lambda}_j \leq a\}$$

the orthogonal projection onto the eigenvectors of the covariance operator K (respectively $\hat{K}$) for which the corresponding eigenvalues λ_j ($\hat{\lambda}_j$) are lower than a, where a is a positive number such that $0 < a \leq |\hat{Q}'_m(0)|^{-1}$. Let us consider the following function $\tilde{\beta}_{PLS}$ defined by

$$\tilde{\Psi}^*_{PLS} = \hat{P}_m(\hat{K})(\hat{K}(\Psi^*)).$$

We have

$$\begin{aligned}
||\hat{\Psi}^*_{PLS} - \Psi^*||_{HS} &\leq ||\hat{\Pi}_a(\hat{\Psi}^*_{PLS} - \Psi^*)||_{HS} + ||(I - \hat{\Pi}_a)(\hat{\Psi}^*_{PLS} - \Psi^*)||_{HS} \\
&\leq ||\hat{\Pi}_a(\hat{\Psi}^*_{PLS} - \tilde{\Psi}^*_{PLS})||_{HS} + ||\hat{\Pi}_a(\tilde{\Psi}^*_{PLS} - \Psi^*)||_{HS} + ||(I - \hat{\Pi}_a)(\hat{\Psi}^*_{PLS} - \Psi^*)||_{HS} \\
&\leq (F1) + (F2) + (F3).
\end{aligned}$$

where $(I - \hat{\Pi}_a)$ is the orthogonal complement of the space $\hat{\Pi}_a$. Let us define $(F1)$, $(F2)$, and $(F3)$ by $(F1) = ||\hat{\Pi}_a(\hat{\Psi}^*_{PLS} - \tilde{\Psi}^*_{PLS})||_{HS}$, $(F2) = ||\hat{\Pi}_a(\tilde{\Psi}^*_{PLS} - \Psi^*)||_{HS}$ and $(F3) = ||(I - \hat{\Pi}_a)(\hat{\Psi}^*_{PLS} - \Psi^*)||$ respectively.

The next step is focused on deriving the upper bound rate of the three terms $(F1)$, $(F2)$ and $(F3)$.

Upper bound rate of $(F1)$ **:**

We have

$$
\begin{aligned}
(F1) &= ||\hat{\Pi}_a(\hat{\Psi}^*_{PLS} - \tilde{\Psi}^*_{PLS})||_{HS} \\
&= ||\hat{\Pi}_a\{\hat{P}_m(\hat{K})(T^*_n(\underline{Y})) - \hat{P}_m(\hat{K})(\hat{K}(\Psi^*))\}||_{HS} \\
&= ||\hat{\Pi}_a\{\hat{P}_m(\hat{K})(T^*_n(\underline{\varepsilon}))||_{HS} \\
&\leq \left\{ \underset{\lambda\in[0,a]}{Sup}\, |\hat{P}_m(\hat{\lambda})| \right\} ||\hat{\Pi}_a(T^*_n(\underline{\varepsilon}))||_{HS} \\
&\leq C|\hat{Q}'_m(0)|||T^*_n(\underline{\varepsilon})||_{HS}.
\end{aligned}
$$

where $C > 0$ is an arbitrary constant, $T^*_n(\underline{Y}) = \frac{1}{N-1}\sum_{i=1}^{N-1} X_i \otimes X_{i+1} = \hat{D}^*$ and $T^*_n(\underline{\varepsilon}) = \hat{C}_{x\varepsilon} = \frac{1}{N-1}\sum_{i=1}^{N-1} X_i \otimes \varepsilon_{i+1}$. The last line holds because for $0 < \lambda < a$,

$$
\hat{P}_m(\lambda) = \frac{1 - \hat{Q}_m(\lambda)}{\lambda} \leq |\hat{Q}'_m(0)| = \sum_{l=1}^{m} \frac{1}{\hat{\theta}_l}.
$$

(See Engl et al. (1996)). On the other hand, we have

$$
|\hat{Q}'_m(0)| = |Q'_m(0)| + o_p\left(\frac{1}{\sqrt{n}}\right)
$$

which follows from the mean-value theorem since $\hat{Q}'_m$ is a continuously differentiable function of $\hat{K}$ and $\hat{D}$, which are consistent estimators of K and D. Also, $|Q'_m(0)| = \sum_{l=1}^{m} \frac{1}{\theta_l} \leq \frac{m}{\theta_m}$. Then,

$$
|\hat{Q}'_m(0)| = O_p\left(\frac{m}{\theta_m}\right).
$$

Moreover,

$$
\begin{aligned}
\mathbb{E}\left[||T^*_n(\underline{\varepsilon})||^2_{HS}|\mathcal{F}_{N-1}\right] &= \mathbb{E}\left[tr(\hat{C}_{x\varepsilon}\hat{C}^*_{x\varepsilon})|\mathcal{F}_{N-1}\right] \\
&= \frac{1}{N-1} tr(V_\varepsilon) tr(\hat{K}) \\
&= O_p\left(\frac{1}{N}\right).
\end{aligned}
$$

Therefore,

$$
(F1)^2 = O_p\left(\frac{m}{\theta_m N}\right).
$$

Upper bound rate of $(F2)$:

$$
\begin{aligned}
(F2) &= ||\hat{\Pi}_a(\tilde{\Psi}^*_{PLS} - \Psi^*)||_{HS} \\
&= ||\hat{\Pi}_a\{\hat{P}_m(\hat{K})(\hat{K}(\Psi^*)) - \Psi^*\}||_{HS} \\
&= ||\hat{\Pi}_a\{\hat{Q}_m(\hat{K})(\Psi^*)||_{HS} \\
&\leq \left\{ \underset{\lambda\in[0,a]}{Sup} |\hat{Q}_m(\lambda)| \right\} ||\hat{\Pi}_a(\Psi^*)||_{HS} \\
&= O_p\left(||\hat{\Pi}_a(\Psi^*)||_{HS} \right).
\end{aligned}
$$

where C is an arbitrary positive constant. This is possible given that $\{ \underset{\lambda\in[0,a]}{Sup} |\hat{Q}_m(\lambda)|\} \leq 1$ by definition of $\hat{Q}_m$ (see Lemma 1 by Carrasco and Tsafack (2020)).

Then

$$
(F2)^2 = O_p\left(\underset{\lambda_j<a}{Sup}\{\lambda_j^{\beta}\} \right).
$$

Upper bound rate of $(F3)$ **:**

We have

$$
\begin{aligned}
(F3) &= ||(I - \hat{\Pi}_a)(\hat{\Psi}^*_{FPLS} - \Psi^*)||_{HS} \\
&= ||(I - \hat{\Pi}_a)(\hat{K}_\delta^{-1}\hat{K})(\{\hat{P}_m(\hat{K})(T_n^*(\underline{Y})) - \Psi^*)||_{HS} \\
&\leq ||\hat{\Pi}_a(\hat{K}_\delta^{-1})||||(I - \hat{\Pi}_a)(\hat{K})(\{\hat{P}_m(\hat{K})(T_n^*(\underline{Y})) - \Psi^*)||_{HS} \\
&\leq \frac{1}{a}||(I - \hat{\Pi}_a)\{\hat{K}\{\hat{P}_m(\hat{K})(T_n^*(\underline{Y})) - \hat{K}(\Psi^*)\}||_{HS} \\
&\leq \frac{1}{a}||(I - \hat{\Pi}_a)\{T_n^*\{\hat{W}_n\{\hat{P}_m(W_n)(\underline{Y}) - T_n(\Psi^*)\}\}||_{HS} \\
&\leq \frac{1}{a}||(I - \hat{\Pi}_a)\{T_n^*\{\hat{W}_n\{\hat{P}_m(W_n)(\underline{Y}) - T_n(\Psi^*)\}\}||_{HS} \\
&\leq \frac{1}{a}||(I - \hat{\Pi}_a)\{T_n^*\{\hat{W}_n\{\hat{P}_m(W_n)(\underline{Y}) - \underline{Y} + \underline{Y} - T_n(\Psi^*)\}\}||_{HS} \\
&\leq \frac{1}{a}||(I - \hat{\Pi}_a)\{T_n^*(\hat{Q}_m(W_n)(\underline{Y}))\}||_{HS} + \frac{1}{a}||(I - \hat{\Pi}_a)\{T_n^*(\underline{\varepsilon})\}||_{HS} \\
&\leq \frac{1}{a}||(I - \hat{\Pi}_a)\{\hat{Q}_m(\hat{K})(T_n^*(\underline{Y}))\}||_{HS} + \frac{1}{a}||(I - \hat{\Pi}_a)\{T_n^*(\underline{\varepsilon})\}||_{HS} \\
&\leq \frac{C\Lambda_m}{a}
\end{aligned}
$$

where $\hat{K}_\delta^{-1}$ is the generalized inverse of $\hat{K}$ using the $FPCA$ regularization and $\Lambda_m = max\{||\{\hat{Q}_m(\hat{K})(T_n^*(\underline{Y}))\}||_{HS}, \frac{1}{\sqrt{N}}\}$. Given the stopping rule, $\Lambda_m = O(\frac{1}{\sqrt{N}})$. The fourth and fifth lines are possible given that $\hat{K} = T_n^*T_n$ and $W_n = T_nT_n^*$. The seventh line is

possible because $\hat{Q}_m(W_n)(\underline{Y}) = \hat{W}_n\{\hat{P}_m(W_n)(\underline{Y}) - \underline{Y}$ and $\underline{\varepsilon} = \underline{Y} - T_n(\Psi^*)$. The last line comes from the fact that $\hat{Q}_m(\hat{K})(T_n^*(\underline{Y})) = T_n^*(\hat{Q}_m(W_n)(\underline{Y}))$ and this is because we have $\hat{K} = T_n^* T_n$ and $W_n = T_n T_n^*$.

When I combine results from $(F1)$, $(F2)$ and $(F3)$, it yields

$$||\hat{\Psi}^*_{FPLS} - \Psi^*||^2_{HS} = O_p\left(||\hat{\Pi}_a(\Psi^*)||_{HS}\right) + O_p\left(\frac{1}{aN}\right) + O_p\left(\frac{m^2}{N\theta_m^2}\right).$$

Since $0 < a \leq |Q'_m(0)|^{-1} \leq \theta_m$, by taking $a = |Q'_m(0)|^{-1}$ we obtain

$$\begin{aligned}
||\hat{\Psi}^*_{FPLS} - \Psi^*||^2_{HS} &= O_p\left(||\hat{\Pi}_{|Q'_m(0)|^{-1}}(\Psi^*)||^2_{HS}\right) + O_p\left(\frac{1}{aN}\right) + O_p\left(\frac{m^2}{N\theta_m^2}\right)) \\
&= O_p\left(||\hat{\Pi}_{\theta_m}(\Psi^*)||^2_{HS}\right) + O_p\left(\frac{1}{|Q'_m(0)|^{-1}N}\right) + O_p\left(\frac{m^2}{N\theta_m^2}\right)) \\
&= O_p\left(||\hat{\Pi}_{\lambda_m}(\Psi^*)||^2_{HS}\right) + O_p\left(\frac{m}{N\theta_m}\right) + O_p\left(\frac{m^2}{N\theta_m^2}\right) \\
&= O_p\left(\lambda^\beta_{m+1}\right) + O_p\left(\frac{m^2}{N\theta_m^2}\right).
\end{aligned}$$

The first term of the last line holds since,

$$\begin{aligned}
||\Pi_{\lambda_m}(\Psi^*)||^2_{HS} &= \sum_{\lambda_j < \lambda_m} < \Psi^*(v_j), \Psi^*(v_j) >^2 \\
&= \sum_{\lambda_j < \lambda_m} < K^{\beta/2} R(v_j), K^{\beta/2} R(v_j) >^2 \\
&= \sum_{\lambda_j \leq \lambda_{m+1}} \lambda_j^\beta < R(v_j), R(v_j) >^2 \\
&= \left[\underset{\lambda_j \leq \lambda_{m+1}}{Sup} \{\lambda_j^\beta\}\right] \sum_{j \geq 1} < R(v_j), R(v_j) >^2 \\
&= O_p\left(\underset{\lambda_j \leq \lambda_{m+1}}{Sup} \{\lambda_j^\beta\}\right) \\
&= O_p\left(\lambda^\beta_{m+1}\right).
\end{aligned}$$

The first term is the bias term and the second one is the variance. The last line is possible since we have $\sum_{j=m+1}^{\infty} \Psi^{*2}_j = ||\Psi^*_{FPCA} - \Psi^*||^2_{HS} = O_p\left(\lambda^\beta_{m+1}\right)$, combined with the fact that λ_{m+1} is the highest eigenvalue of K that is lower than θ_m (see lemma 1 by Carrasco and Tsafack (2020)). Therefore, the result follows.

Let prove that $||\Psi^*_{PLS} - \Psi^*||^2_{HS} \leq ||\Psi^*_{PCA} - \Psi^*||^2_{HS}$

Following the same logic as in proof of Proposition 2 by Carrasco and Tsafack (2020)

$$\begin{aligned}
\left|\left|\Psi^*_{PLS}-\Psi^*\right|\right|^2_{HS} &= \sum_{j=1}^{\infty} Q(m,\lambda_j)^2 < \Psi^*(v_j),\Psi^*(v_j)>^2 \\
&\leq \sum_{j=1}^{\infty}\left[\sum_{(j_1,...,j_m)\in I^+} w_{j_1,...,j_m}\prod_{l=1}^{m}(1-\frac{\lambda_j}{\lambda_{j_l}})\right]^2 < \Psi^*(v_j),\Psi^*(v_j)>^2 \\
&\leq \sum_{j=1}^{\infty} \underset{(j_1,...,j_m)\in I^+}{sup}\prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_{j_l}}\right)^2 < \Psi^*(v_j),\Psi^*(v_j)>^2 \\
&\leq \sum_{j=1}^{\infty}\prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_l}\right)^2 < \Psi^*(v_j),\Psi^*(v_j)>^2
\end{aligned}$$

because for $j=1,...,m$, $\prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_l}\right)^2=0$, while for $j\geq m+1$, we have $0\leq \prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_l}\right)^2\leq 1$, because all of the eigenvalues are nonzero ordered in such a way that $\lambda_1>\lambda_2>....>\lambda_m>\lambda_{m+1}>...>0$. We should also recall that $0\leq w_{j_1,...,j_m}\leq 1$, with $\sum_{(j_1,...,j_m)\in I^+_m} w_{j_1,...,j_m}=1$. The last inequality follow from the fact that the eigenvalues are all distinct and

$$\underset{(j_1,...,j_m)\in I^+}{sup}\prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_{j_l}}\right)^2=\prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_l}\right)^2\leq 1.$$

Therefore, we have

$$\begin{aligned}
\left|\left|\Psi^*_{PLS}-\Psi^*\right|\right|^2_{HS} &= \sum_{j=1}^{\infty} Q(m,\lambda_j)^2 < \Psi^*(v_j),\Psi^*(v_j)>^2 \\
&\leq \sum_{j=1}^{\infty}\prod_{l=1}^{m}\left(1-\frac{\lambda_j}{\lambda_l}\right)^2 < \Psi^*(v_j),\Psi^*(v_j)>^2 \\
&\leq \sum_{j=m+1}^{\infty} < \Psi^*(v_j),\Psi^*(v_j)>^2 \\
&= \left|\left|\Psi^*_{PCA}-\Psi^*\right|\right|^2_{HS}.
\end{aligned}$$

Then, $\left|\left|\Psi^*_{PLS}-\Psi^*\right|\right|^2_{HS}\leq \left|\left|\Psi^*_{PCA}-\Psi^*\right|\right|^2_{HS}$ and $\mathbb{E}\left[\left|\left|\Psi^*_{PLS}-\Psi^*\right|\right|^2_{HS}|\mathcal{F}_{n-1}\right]\leq \mathbb{E}\left[\left|\left|\Psi^*_{PCA}-\Psi^*\right|\right|^2_{HS}|\mathcal{F}_{n-1}\right]$. Therefore,

$$\mathbb{E}\left[\left|\left|\Psi^*_{PLS} - \Psi^*\right|\right|^2_{HS} |\mathcal{F}_{n-1}\right] = O_p\left(\lambda^{\beta}_{m+1}\right).$$

2.12.1 Proof of Proposition 4.

For δ fixed (α for FT, FSC and FLF)

$$\begin{aligned}\hat{\Psi}^*_\delta &= \hat{K}^{-1}_\delta \hat{D}^* \\ &= \hat{K}^{-1}_\delta \hat{K}(\Psi^*) + \hat{K}^{-1}_\delta \hat{C}_{x,\varepsilon}.\end{aligned}$$

and $\Psi^*_\delta = K^{-1}_\delta K(\Psi^*)$. Then,

$$\begin{aligned}\hat{\Psi}^*_\delta - \Psi^*_\delta &= \hat{K}^{-1}_\delta \hat{C}_{x,\varepsilon} + \hat{K}^{-1}_\delta \hat{K}(\Psi^*) - K^{-1}_\delta K(\Psi^*) \\ &= K^{-1}_\delta \hat{C}_{x,\varepsilon} + (\hat{K}^{-1}_\delta - K^{-1}_\delta)\hat{C}_{x,\varepsilon} + \hat{K}^{-1}_\delta \hat{K}(\Psi^*) - K^{-1}_\delta \hat{K}(\Psi^*) + K^{-1}_\delta \hat{K}(\Psi^*) - K^{-1}_\delta K(\Psi^*) \\ &= K^{-1}_\delta \hat{C}_{x,\varepsilon} + (\hat{K}^{-1}_\delta - K^{-1}_\delta)\hat{C}_{x,\varepsilon} + (\hat{K}^{-1}_\delta - K^{-1}_\delta)\hat{K}(\Psi^*) + \hat{K}^{-1}_\delta(\hat{K}_\delta - K_\delta)(\Psi^*) \\ &= K^{-1}_\delta \hat{C}_{x,\varepsilon} + \hat{K}^{-1}_\delta(\hat{K}_\delta - K_\delta)(\Psi^*) + O_p\left(\frac{1}{N}\right) \\ &= K^{-1}_\delta\left[\frac{1}{N}\sum_{i=1}^{N-1} X_i \otimes \varepsilon_{i+1}\right] + K^{-1}_\delta\left[\frac{1}{N}\sum_{i=1}^{N-1}(X_i \otimes X_i - K)\Psi^*\right] + O_p\left(\frac{1}{N}\right).\end{aligned}$$

Let us consider $\tilde{X}_i = K^{-1}_\delta(X_i)$ and $\tilde{X} = K^{-1}_\delta(X)$. Then,

$$\begin{aligned}\hat{\Psi}^*_\delta - \Psi^*_\delta &= \frac{1}{N}\sum_{i=1}^{N-1} \tilde{X}_i \otimes \varepsilon_{i+1} + \frac{1}{N}\sum_{i=1}^{N-1} \tilde{X}_i \otimes \Psi(X_i) - \mathbb{E}[\tilde{X} \otimes \Psi(X)] + O_p\left(\frac{1}{N}\right) \\ &= \frac{1}{N}\sum_{i=1}^{N-1} \tilde{X}_i \otimes \varepsilon_{i+1} + \frac{1}{N}\sum_{i=1}^{N-1}[\tilde{X}_i \otimes X_i - \mathbb{E}[\tilde{X} \otimes X]](\Psi^*) + O_p\left(\frac{1}{N}\right).\end{aligned}$$

Under the assumptions that $X_i \otimes \varepsilon_{i+1}$ are stationary and ergodic martingale difference sequences, $\mathbb{E}[||X_i||^4] < +\infty$ and $\mathbb{E}[||X_i||^2||\varepsilon_{i+1}||^2] < +\infty$,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N-1}\begin{bmatrix} X_i \otimes \varepsilon_{i+1} \\ X_i \otimes X_i - K\end{bmatrix} \overset{d}{\Longrightarrow} N(0, \Omega_1)$$

where $\Omega_1 = \begin{pmatrix} G & 0 \\ 0 & K\end{pmatrix}$ is a (2×2) matrix of covariance operators.

Let consider the following transformation

$$\begin{bmatrix} A \\ B \end{bmatrix} \to K_\delta^{-1}A + \Psi K_\delta^{-1}(B - K)$$

where A and B are arbitrary operators. Then,

$$\begin{aligned}\hat{\Psi}_\delta^* - \Psi_\delta^* &= \frac{1}{N}\sum_{i=1}^{N-1} \tilde{X}_i \otimes \varepsilon_{i+1} + \frac{1}{N}\sum_{i=1}^{N-1}[\tilde{X}_i \otimes X_i - \mathbb{E}[\tilde{X} \otimes X]](\Psi^*) \\ &= K_\delta^{-1}A + \Psi K_\delta^{-1}(B - K)(\Psi^*).\end{aligned}$$

with $A = \frac{1}{N}\sum_{i=1}^{N-1} X_i \otimes \varepsilon_{i+1}$ and $B = \frac{1}{N}\sum_{i=1}^{N-1}[X_i \otimes X_i - \mathbb{E}[X \otimes X]]$.

By the continuous mapping theorem, the asymptotic covariance operator of $\sqrt{N}(\hat{\Psi}_\delta^* - \Psi_\delta^*)$ is given by

$$\Omega_\delta = K_\delta^{-1}\mathbb{E}\Big[(X_i \otimes \varepsilon_{i+1})\tilde{\otimes}(\varepsilon_{i+1} \otimes X_i)\Big]K_\delta^{-1} + K_\delta^{-1}(\Psi^*)\mathbb{E}\Big[(X_i \otimes X_i - K)\tilde{\otimes}(X_i \otimes X_i - K)\Big]K_\delta^{-1}(\Psi^*),$$

where $\tilde{\otimes}$ is the tensor product of two operators. Then, for $(A, B) \in \mathbb{H}^{\mathbb{H}} \times \mathbb{H}^{\mathbb{H}}$, $A\tilde{\otimes}B$ is an element of the Hilbert space of Hilbert-Schmidt operators from $\mathbb{H}^{\mathbb{H}}$ to $\mathbb{H}^{\mathbb{H}}$.

2.12.2 Proof of Proposition 5.

For $h = 1, ..., p$. Assume that we are testing

$$\begin{cases} H_0 : \Psi_h(.) = 0 \\ H_1 : \Psi_h(.) \neq 0. \end{cases}$$

Under H_0, we have $\Psi_h(.) = 0$ and the model FAR(h-1) is estimated

$$\hat{X}_{n+1} = \hat{\Psi}_1(X_n) + ... + \hat{\Psi}_{h-1}(X_{n-h+2}) + \hat{\varepsilon}_{n+1}$$

and

$$\hat{D} = \hat{K}(\hat{\Psi}_1^*) + ... + \hat{K}_{h-1}(\hat{\Psi}_{h-1}^*) + \frac{1}{N}\sum_{n=1}^{N-h+1} X_n \otimes \hat{\varepsilon}_{n+1}$$

On the other hand, $\hat{\varepsilon}_{n+1}$, $X_n \otimes \hat{\varepsilon}_{n+1}$, ... , $X_{n-h+2} \otimes \hat{\varepsilon}_{n+1}$ are martingale difference functions and we have $\frac{1}{N}\sum_{n=1}^{N-h+1} X_n \otimes \hat{\varepsilon}_{n+1}$. Additionally, $E[||X_i||^4] < \infty$, $E(||X_i||^2||\varepsilon_i||^2) < \infty$. Then, the functional central limit theorem for martingale difference sequence leads to

$$\sqrt{N}(\hat{C}_{x\varepsilon} - 0) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N-h+1} X_n \otimes \hat{\varepsilon}_{n+1} \xrightarrow{d} \mathcal{N}(0, K_{x\varepsilon}) \quad as\ N \to \infty.$$

with

$$
\begin{aligned}
K_{x\varepsilon} &= \mathbb{E}\Big[(X_i \otimes \varepsilon_{i+1})\tilde{\otimes}(X_i \otimes \varepsilon_{i+1})\Big] \\
&= \mathbb{E}\Big[(X_i \otimes X_i)\tilde{\otimes}\mathbb{E}[(\varepsilon_{i+1} \otimes \varepsilon_{i+1})|X_1, ..., X_n]\Big] \\
&= \mathbb{E}\Big[(X_i \otimes X_i)\Big]\tilde{\otimes}\mathbb{E}\Big[(\varepsilon_{i+1} \otimes \varepsilon_{i+1})\Big] \\
&= K\tilde{\otimes}V_\varepsilon.
\end{aligned}
$$

Then, we have

$$
\sqrt{N}\hat{K}_{x\varepsilon}^{1/2}(\hat{C}_{x\varepsilon} - 0) \overset{d}{\Longrightarrow} \mathcal{N}(0, \mathcal{I}) \quad as\ N \to \infty.
$$

Furthermore, $\hat{K}_{x\varepsilon} = \hat{K}\tilde{\otimes}\hat{V}_\varepsilon$, then the eigen-decomposition of $\hat{K}_{x\varepsilon}$ is given for each function f_0 by

$$
\begin{aligned}
\hat{K}_{x\varepsilon} &= \frac{1}{N}\sum_{\ell=1}^{N-h-h+2} \hat{\lambda}_\ell < \hat{V}_\varepsilon(f_0), \hat{v}_\ell > \hat{v}_\ell \\
&= \frac{1}{N}\sum_{\ell,n=1}^{N-h+2} \hat{\lambda}_\ell < \hat{\varepsilon}_{n+1}, \hat{\varepsilon}_{\ell+1} >< ., \hat{v}_\ell > \hat{v}_n \\
&= \frac{1}{N}\sum_{\ell=1}^{N-h+2} \hat{\lambda}_\ell < \hat{\varepsilon}_{\ell+1}, \hat{\varepsilon}_{\ell+1} >< ., \hat{v}_\ell > \hat{v}_\ell \overset{P}{\Longrightarrow} \sum_{\ell=1}^{\infty} \lambda_\ell\sigma^2 < ., v_\ell > v_\ell.
\end{aligned}
$$

Therefore

$$
\hat{K}_{x\varepsilon}^{-1} = \sum_{\ell=1}^{N-h+2} \frac{Q(\delta, \hat{\lambda}_\ell)^2}{\hat{\lambda}_\ell\hat{V}_{\varepsilon,n}} < ., \hat{v}_\ell > \hat{v}_n
$$

where $\hat{V}_{\varepsilon,n} = \frac{1}{N}\sum_{\ell,n=1}^{N-h+2} < \hat{\varepsilon}_{n+1}, \hat{\varepsilon}_{\ell+1} >= \hat{\sigma}_\varepsilon^2 \overset{P}{\Longrightarrow} \sigma^2$

$$
\begin{aligned}
N||\hat{K}_{x\varepsilon}^{-1/2}\hat{C}_{x\varepsilon}||_{HS}^2 &= tr\Big[[\hat{K}_{x\varepsilon}^{-1/2}\hat{C}_{x\varepsilon}][\hat{K}_{x\varepsilon}^{-1/2}\hat{C}_{x\varepsilon}]^*\Big] \\
&= \sum_{\ell}^{N} \frac{Q(\delta, \hat{\lambda}_\ell)}{\hat{\lambda}_\ell\hat{V}_{\varepsilon,n}} < \sqrt{N}\hat{C}_{x\varepsilon}(\hat{v}_\ell), \hat{v}_\ell >^2 \\
&\overset{d}{\Longrightarrow} \sum_{\ell=1}^{\infty} \frac{Q(\delta, \lambda_\ell)^2}{\lambda_\ell\sigma^2}\chi_\ell^2(1) \quad as\ \ N \to +\infty
\end{aligned}
$$

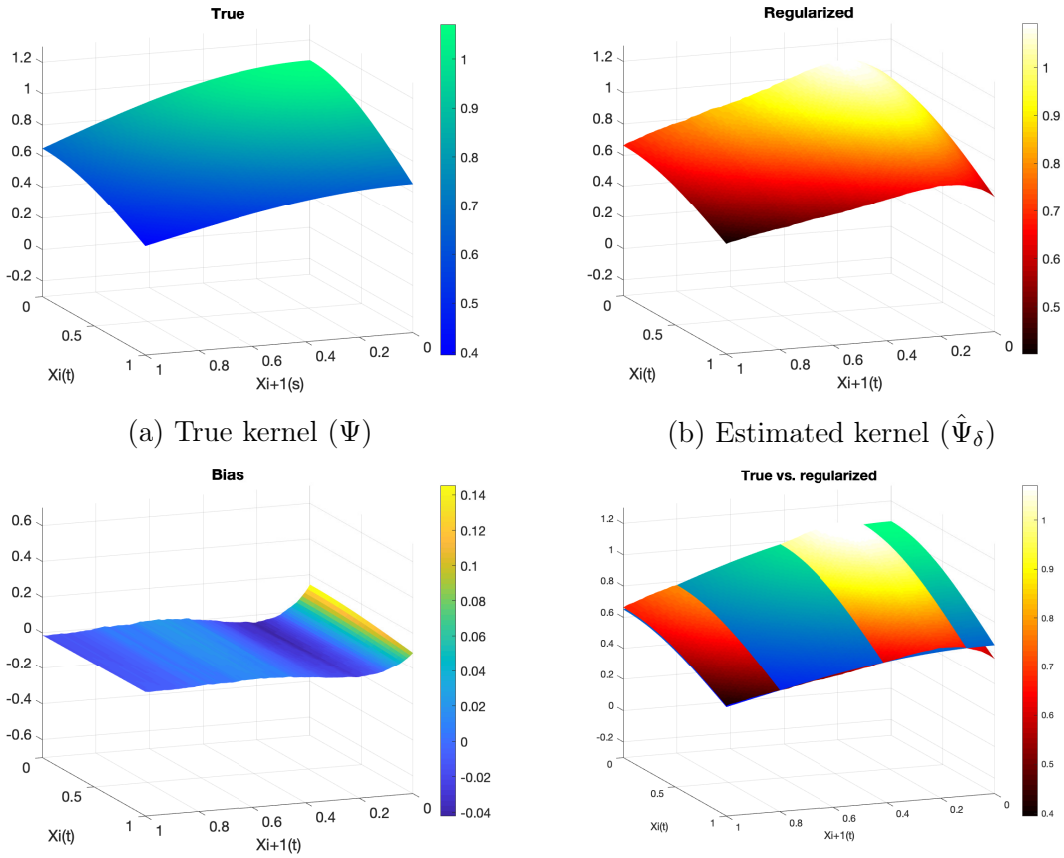
where $\chi_1^2(h, \ell)$ is a random variable that follows a $\chi^2(1)$ distribution for each h and ℓ.

Under H_1, $\Psi_h \neq 0$. Then, $\hat{C}_{x\varepsilon}$ is not close to zero and I obtain

$$N||\hat{K}_{x\varepsilon}^{-1/2}\hat{C}_{x\varepsilon}||_{HS}^2 \overset{d}{\Longrightarrow} +\infty \ as \ N \to +\infty.$$

2.13 Graphics and Tables

2.13.1 Graphics



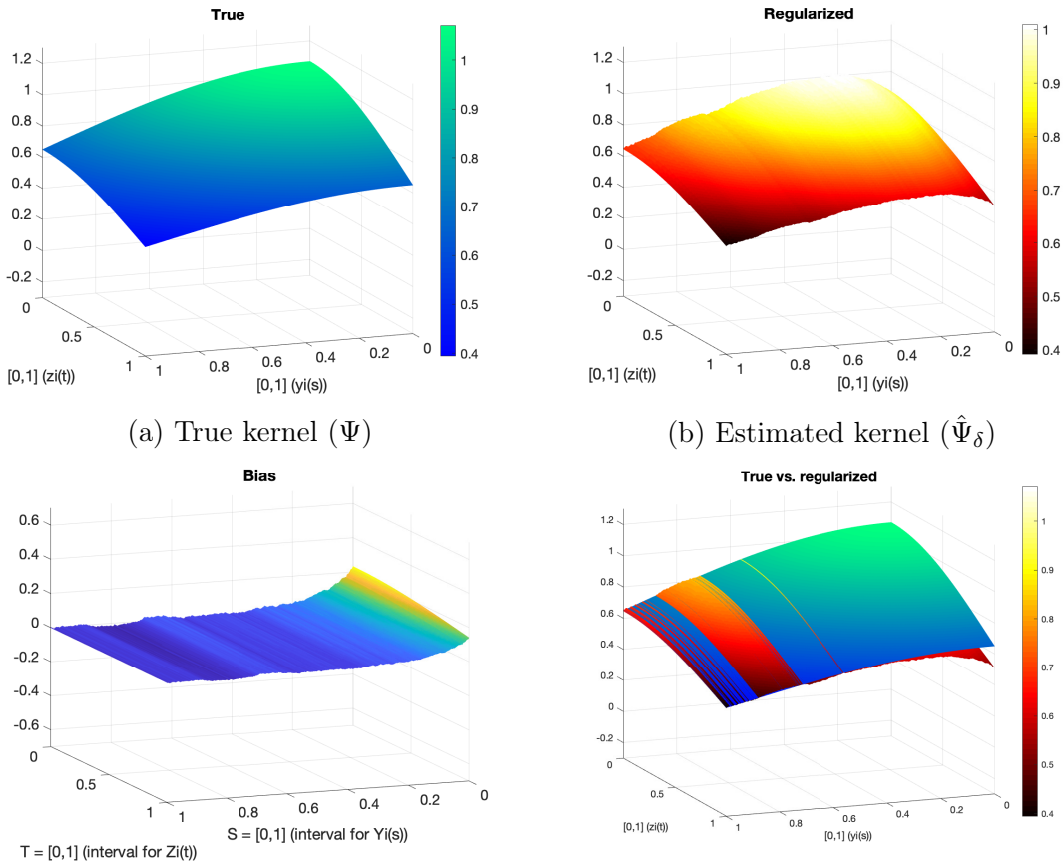
(a) True kernel (Ψ)

(b) Estimated kernel ($\hat{\Psi}_\delta$)

(c) Bias

(d) True and estimated

FIGURE 2.8: Comparison of the true gaussian and estimated kernels - FPLS



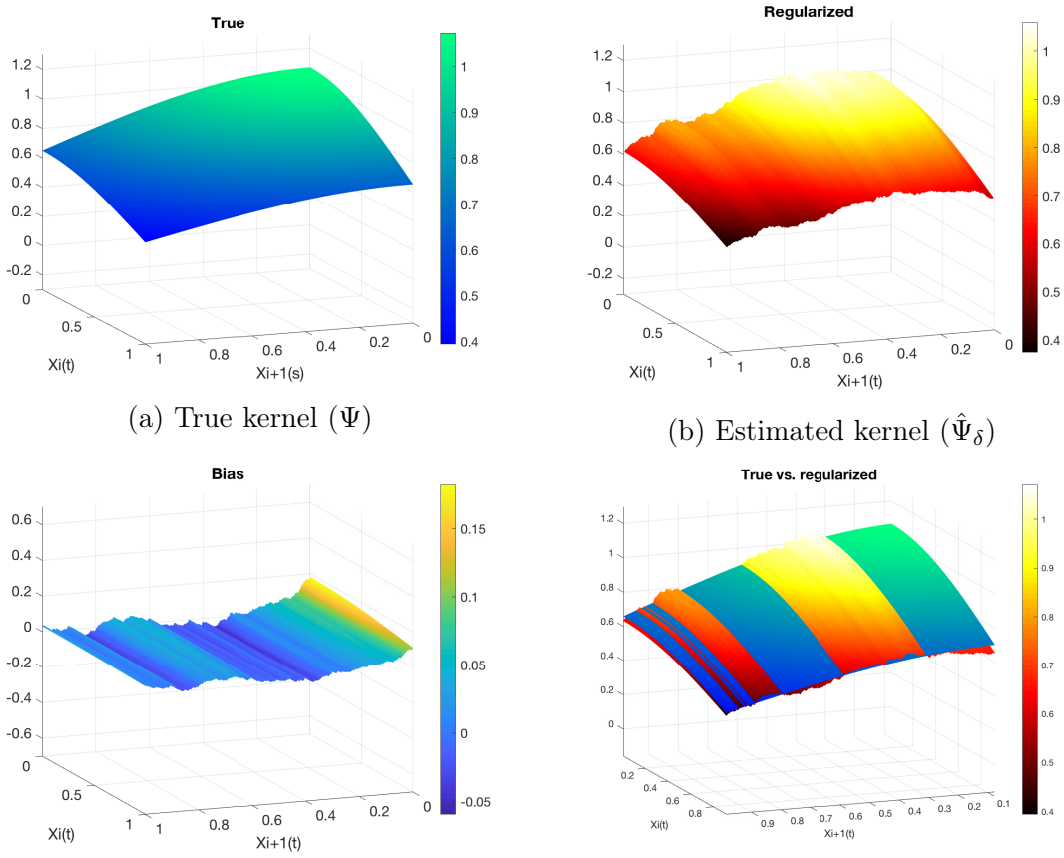
(a) True kernel (Ψ)

(b) Estimated kernel ($\hat{\Psi}_\delta$)

(c) Bias

(d) True and estimated

FIGURE 2.9: Comparison of the true gaussian and estimated kernels - FT



(a) True kernel (Ψ)

(b) Estimated kernel ($\hat{\Psi}_\delta$)

(c) Bias

(d) True and estimated

FIGURE 2.10: Comparison of the true gaussian and estimated kernels - FLF

2.13.2 Tables

TABLE 2.6: Comparison of the different estimation techniques. Gaussian kernel, $N = 1000$, $M = 1000$ replications, and $\varepsilon^{(2)}$

	Moments	FPCA	FPLS	FT	FLF	FSC
MSE	Mean	0.791	0.688	0.862	6.037	0.791
	Std	0.055	0.044	0.003	0.008	0.055
	Median	0.795	0.684	0.862	6.046	0.791
AD	Mean	0.849	0.762	0.875	2.185	0.849
	Std	0.035	0.026	0.002	0.014	0.035
	Median	0.853	0.760	0.874	2.186	0.853
MSPE	Mean	0.075	0.075	0.075	0.082	0.075
	Std	0.0002	0.0003	0.0002	0.0004	0.0002
	Median	0.075	0.075	0.075	0.082	0.075
MAPE	Mean	0.049	0.049	0.049	0.051	0.049
	Std	0.0001	0.0002	0.0001	0.104	0.001
	Median	0.049	0.049	0.049	0.051	0.049
R^2_{is}	Mean	0.001	0.002	0.001	0.002	0.001
	Std	0.0003	0.0002	0.0001	0.0022	0.0003
	Median	0.001	0.002	0.001	0.002	0.001
R^2_{oos}	Mean	0.003	0.002	0.002	0.096	0.003
	Std	0.0003	0.0002	0.0002	0.0022	0.0022
	Median	0.003	0.002	0.002	0.096	0.003

TABLE 2.7: Comparison of the different estimation techniques. Gaussian kernel, $N = 1000$, $M = 1000$ replications, and $\varepsilon^{(3)}$

	Moments	FPCA	FPLS	FT	FLF	FSC
MSE	Mean	0.895	0.788	0.862	7.297	0.895
	Std	0.035	0.028	0.003	0.061	0.035
	Median	0.893	0.787	0.862	7.307	0.895
AD	Mean	0.870	0.761	0.875	2.404	0.870
	Std	0.035	0.026	0.002	0.014	0.035
	Median	0.872	0.760	0.874	2.406	0.872
MSPE	Mean	0.075	0.075	0.075	0.085	0.075
	Std	0.0002	0.0003	0.0002	0.0004	0.0002
	Median	0.075	0.075	0.075	0.085	0.075
MAPE	Mean	0.049	0.049	0.049	0.051	0.049
	Std	0.0001	0.0002	0.0001	0.104	0.001
	Median	0.049	0.049	0.049	0.051	0.049
R^2_{is}	Mean	0.002	0.002	0.001	0.002	0.001
	Std	0.0002	0.0001	0.0001	0.0001	0.138
	Median	0.002	0.006	0.001	0.002	0.139
R^2_{oos}	Mean	0.003	0.004	0.002	0.137	0.003
	Std	0.0003	0.0002	0.0002	0.0022	0.0022
	Median	0.003	0.002	0.002	0.137	0.003

TABLE 2.8: Comparison of the different estimation techniques, factor based kernel, $N = 1000$, $M = 1000$ replications, and $\varepsilon^{(2)}$

	Moments	FPCA	FPLS	FT	FLF	FSC
MSE	Mean	0.858	0.768	0.799	6.110	0.858
	Std	0.0060	0.0056	0.0002	0.050	0.0060
	Median	0.859	0.767	0.799	6.116	0.858
AD	Mean	0.827	0.856	0.797	2.199	0.870
	Std	0.0029	0.0235	0.0001	0.014	0.0091
	Median	0.827	0.856	0.797	2.199	0.870
MSPE	Mean	0.075	0.075	0.075	0.082	0.075
	Std	0.0003	0.0003	0.0003	0.0004	0.0003
	Median	0.075	0.075	0.075	0.082	0.075
MAPE	Mean	0.049	0.049	0.049	0.051	0.049
	Std	0.0001	0.0002	0.0001	0.106	0.0002
	Median	0.049	0.049	0.049	0.051	0.049
R^2_{is}	Mean	0.001	0.001	0.001	0.106	0.001
	Std	0.0002	0.0001	0.0001	0.0001	0.0002
	Median	0.001	0.006	0.001	0.106	0.001
R^2_{oos}	Mean	0.002	0.003	0.002	0.098	0.002
	Std	0.0002	0.0003	0.0002	0.0002	0.0002
	Median	0.002	0.003	0.002	0.098	0.002

TABLE 2.9: Comparison of the different estimation techniques, factor based kernel, $N = 1000$, $M = 1000$ replications, and $\varepsilon^{(3)}$

	Moments	FPCA	FPLS	FT	FLF	FSC
MSE	Mean	0.00013	0.00012	0.00047	0.00023	0.00013
	Std	0.0057	0.053	0.021	0.010	0.0057
	Median	0.00013	0.00012	0.00047	0.00023	0.00013
AD	Mean	0.0002	0.0007	0.0004	0.0003	0.0002
	Std	0.0098	0.0294	0.0194	0.0136	0.0098
	Median	0.0002	0.0007	0.0004	0.0003	0.0002
MSPE	Mean	0.0001	0.0001	0.0001	0.0001	0.0001
	Std	0.0005	0.0006	0.0005	0.0005	0.0005
	Median	0.0001	0.0001	0.0001	0.0001	0.0001
MAPE	Mean	0.0001	0.0001	0.0001	0.0001	0.0001
	Std	0.0006	0.0006	0.0006	0.0006	0.0006
	Median	0.0001	0.0001	0.0001	0.0001	0.0001
R^2_{is}	Mean	0.002	0.002	0.001	0.002	0.001
	Std	0.0002	0.0001	0.0001	0.0001	0.138
	Median	0.002	0.006	0.001	0.002	0.139
R^2_{oos}	Mean	0.0023	0.0027	0.0020	0.097	0.0023
	Std	0.0021	0.0004	0.0016	0.0020	0.0022
	Median	0.0023	0.0027	0.0020	0.097	0.0023

Chapitre 3

Risk Neutral Density Estimation with a Functional Linear Model *

3.1 Introduction

Estimating the risk neutral density (RND) has been an important topic for financial market participants and monetary policymakers. Indeed, this tool is used for derivatives pricing, hedging and market sentiment analysis. Additionally, it is used to analyze the trader's behavior to a potential shock in the financial market and predict the extreme shocks probabilities. For the policymaker, this tool is used to estimate the effectiveness of monetary policies through direct observation of changes in investor's expectations and beliefs to future maturities (see Souissi (2017)).

This concept is also fundamental in the arbitrage-free asset pricing theory. Indeed, the RND is a density measure under which the price of each security in the market is equal to expected value of its future payoff given a risk-free interest rate discounted back to the present. This means that for most of the securities in the market, the number of states of the economy could be very large, which in turns lead to situations where the number of potential future payoffs is very dense. Then the future payoff can be considered as a continuous function of the potential states of the economy. On the same line, since the set of different state of the economy is very large, the RND is a continuous function of the future payoff function and it's form is unknown. Therefore, these properties should be taken into account in the estimation procedure of the RND.

To address the estimation of the functional form of the RND, two main approaches have been suggested in the literature. The first approach relies on parametric modelling (see Black and Scholes (1973) , Bahra (1997), Figlewski (2008)). Most of the authors in this range of the literature focus on considering a specific form of distribution to the RND and then estimate the related parameter. The most used distribution in this context is

∗. This chapter is co-authored with Marine Carrasco.

the log-normal density or a mixture of the log-normal distributions. Unfortunately, this approach is limited to fully explain all types of data generating process (see p.89 by Bondarenko (2003)).

The second approach is the nonparametric technique. In this part of the literature, researchers try to extract the RND with more flexible considerations than in the context of parametric models. Usually, the flexibility comes from not assuming a certain distribution function of the concerned asset. Researchers, have suggested different estimation techniques such as the kernel smoothing of the option prices (see Aıt-Sahalia and Lo (2000), Garcia and Gençay (2000), Aıt-Sahalia and Duarte (2003), Souissi (2017)), regularization of the observed implied volatility curves (see Bliss and Panigirtzoglou (2004), Panigirtzoglou and Skiadopoulos (2004)). Furthermore, there is also another part of the literature considering the nonparametric techniques combined with the arbitrage-free hypothesis. In this context, some authors suggest to estimate the RND via a polynomial approximation method (see Shimko et al. (1993), Rosenberg (1998), Vogt (2014)), while others propose to use the smoothing spline technique (see Bondarenko (2003)).

Based on the precedent literature, it can be noticed that most of the approaches that attempt to estimate the RND end up by making a trade-off between the goodness-of-fit of their model (by considering a form of the distribution or a latent form for approximation) and the connection of their model with the arbitrage-free theory. Whether or not, the methods proposed in the literature tend to exploit a latent form for the density approximation (spline smoothing, polynomial approximation, kernel smoothing, log-normal distribution, sieve) and do not necessarily guarantee the arbitrage-free hypothesis. Recall that the common goal is to develop a model that is consistent with the concept of risk neutral density posited by economic theory and is of use for practical purpose. A naive attempt to connect these purposes is to model RND as a predefined distribution function or a function projected on basis functions for approximation. However, when considering such form for the density function, we automatically make a sort of restriction to the type of assets that may hold with the assumptions.

To fill this gap, this paper proposes to estimate the risk neutral density for European option pricing by exploiting the functional data analysis framework. The advantage of this approach is to connect the fundamental theory on arbitrage-free, the functional feature of the RND and the capacity to realize a good fitting performance. Additionally, the estimation do not rely on any latent form of the distribution for approximation. In fact, we consider that the risk neutral density is a function observed on a very fine grid and there is an infinite possible state of the economy at the maturity. This leads to the fact that the payoff can be observed as a function of all possible states of the economy. Then, we deal with a functional linear regression model where the predictor variables are functions representing the future payoffs functions at the maturity and the response is a scalar representing the call and put prices of the considered security. Then, the call and

put prices are treated as a weighted sum of all the potential pay-offs of the considered option at the time-to-maturity, with the weights represented by the density values.

The estimation of the density function in this context leads to an ill-posed problem. To overcome this issue, we propose to use a regularization technique called the Functional Landweber-Fridman (FLF) method, then apply a density correction procedure in order to get the nonnegativity and the integration to one. The FLF technique is an iterative method that relies on the normal equation related to the regression problem, in such a way that the estimation is made without inverting any operator in the procedure. The advantage of this method is that it relies on neither a basis projection nor a kernel smoothing and it is capable to control the smoothness of the estimated density following that the model is estimated on a fine grid. Further, this estimation approach helps to stabilize the variability of the estimated density coming from the ill-posed problem. Another advantage is the possibility to derive directly the asymptotic normality results and confidence sets for the estimated density and the predictions of options prices.

The contribution of this paper is to use the functional data analysis framework in order to provide the risk neutral density estimation. This alternative approach has not been explored in the precedent literature. We derive consistency results of the estimated RND and the asymptotic normality. Finally, we analyze the performance of considered approach based on some Monte Carlo simulation and real data of S&P 500. Based on empirical analysis, we find that the proposed estimation method yields some reasonable results compared to the approach by Bondarenko (2003).

The rest of the paper is organized as follows. Section 3.2 introduces the theoretical model and the estimation method. Section 3.3 gives a presentation of the consistency results. Section 3.4 derives the asymptotic normality results of the estimations and predictions. Section 3.5 presents the data-driven method to select the optimal tuning parameter. Section 3.6 presents the results of the simulations. Section 3.7 is dedicated to the empirical analysis. Section 3.8 concludes.

3.2 The Theoretical model and estimator

This section is devoted to the presentation of the linear regression model and estimation method of the functional parameter.

3.2.1 The functional linear model

In the intertemporal equilibrium models, the current price of a security can be expressed as the expected net value of its future payoffs discounted back to the present. The present value P_t is calculated in terms of the risk-free interest rate and the expectation is obtained with respect to the marginal rate of substitutions weighted density of the payoffs (see Lucas Jr (1978), Rubinstein (1976), Cox and Ross (1976)) as follows

$$P_t = \delta D(T-t)\mathbb{E}_t^{\mathbb{Q}}[Z(S_T)]$$

where $\mathbb{E}_t^{\mathbb{Q}}[Z(S_T)]$ is the expectation of the marginal rate of substitution and $\delta D(T-t)$ is the discount factor for the maturity $T-t$. More specifically, in the derivative market, an option is defined as a contract giving the right (and not the obligation) to buy or sell a risky asset with price s at a predetermined value called strike price κ at (or within) a given maturity date of the contract. There exist many kind of options in the market. The main known are the American and the European options. In this paper we consider only European options characterized by the fact that the exercise of the contract is possible only at the given maturity date.

Then, in the context of a complete market the price of a European put option P_t with a maturity $T-t$, an underlying price at maturity S_T and a strike price κ, the price is equal to the expected pay-offs $Z(S_T)$ discounted back to the present. In other words, it is given by :

$$P_t = e^{-r_{t,T},(T-t)} \int_0^\infty Z(S_T) f(S_T|(T-t)) dS_T \tag{3.1}$$

where $f(S_T|(T-t))$ is the unobserved density conditional to the maturity T, $r_{t,T}$ is the riskless interest rate between date t and T, $Z(S_T) = max(\kappa - S_T, 0)$ the pay-off. For the reasons of simplification, we will note $\tau = T-t$ and $r_{t,T} = r$. It is important to precise that the data are presented as a cross-sectional configuration of options at a single time t.

The previous equation holds when it is assumed that the market is complete, this means that market participants have all the informations about the risky assets. Because of illiquidity in the market, transaction cost, taxes, measurement errors, the market is usually incomplete (see Gourieroux and Jasiak (2001)). Then it may exist an error term capturing all those uncontrolled informations and this uncertainty may vary according to the time to maturity of the option. The longer the time-to-maturity is, the bigger the variability (see Ait-Sahalia et al. (2018), Driessen et al. (2009)). Then, for each option i at a fixed time t and the same maturity τ, we have the following equation :

$$Y_i = \int_0^\infty Z_i(s) f(s|\tau) ds + \varepsilon_i \tag{3.2}$$

where $Y_i = e^{r,\tau} P_i$, ε_i is a conditionally zero-mean, homoskedastic error term. For the sake of this model we assume that there is an infinite possibility of pay-off at the maturity date, which means that the set of potential payoff is very dense and the conditional density is a function taking its values in a very fine grid. This leads to a functional linear regression with the functional predictor represented by the future payoff $Z_i(s) = max(\kappa_i - s, 0)$ and a scalar response (Y_i). As the pay-off is uncertain, I define a large set (or a very fine grid) of possible values for the underlying s, represented by the potential expiration

prices of the option. Based on that, it is possible to consider the potential pay-offs as independent functions mapping from the set of underlying to the real line $\mathbb{R}$. The call options are also considered in the population model by using the call-put parity, that is $C_{it} = P_{it} + S_{it}e^{-r\tau} - \kappa e^{-r\tau}$.

3.2.2 The literature

This section presents an overview of the techniques suggested to estimate the RND. For the estimation, the properties of the density should be respected. Indeed, the properties of a density probability are the nonnegativity of the function and the integration to one. To solve this issue, authors suggest different techniques that could be either parametric, semi-parametric or nonparametric.

Parametric techniques rely usually on considering a specific distribution of the option prices. The most popular distributions considered in this context are the lognormal density (Jarrow and Rudd (1982)) and the lognormal mixture distribution (Bahra (1997)).

Concerning the nonparametric techniques, Aıt-Sahalia and Lo (2000) suggest to estimate the RND with a kernel smoothing approach based on the observed options prices. Indeed, the authors use a predefined kernel which depends on the choice of the bandwidth. On the same line, Aıt-Sahalia and Duarte (2003) proposed a two-step technique, where in the first step they run a constrained regression in order to guarantee the arbitrage-free condition and in the second step, they use a kernel smoothing technique to smooth the predicted option prices. Additionally, we have the regularization technique suggested by Jackwerth and Rubinstein (1996) to estimate the RND in such a way that the options prices are well predicted. We can also identify the kernel smoothing techniques proposed by Garcia and Gençay (2000), Panigirtzoglou and Skiadopoulos (2004) and Bliss and Panigirtzoglou (2004).

Concerning the semi-parametric techniques, Bondarenko (2003) suggests to estimate the density with a nonparametric method called the positive convolution approximation (PosConv). The idea of this technique is to shrink the infinite-dimensional minimization problem into a finite-dimensional one. In the model setting, the true density is considered as a weighted sum of normal densities, in other words it is a mixture of normal densities. One of the challenge in his paper is to select the number of normal densities to consider in the mixture, and the bandwidth of the distributions in their estimation procedure. The author uses a two-step data driven method to select the tuning parameters. The first step consist in constructing a preliminary undersmoothed estimator for a small bandwidth. The second step uses the preliminary estimator to estimate the optimal bandwidth. Then the estimated density is the one selected in the finite set of candidate densities such that it fits well the options prices. This approach has the advantage of being independent of any data generating process, respecting the non-arbitrage condition and capable to produce small sample results. Furthermore, we can identify the projection on quadratic

polynomial basis. This idea has been suggested by Shimko et al. (1993) to estimate the RND and predict the observed implied volatility. Rosenberg (1998) suggest to use a sigma-shaped polynomial technique for the same problem. Also, Yatchew and Härdle (2006) and Fengler (2009) use a smoothing spline technique. More recently, Vogt (2014) propose to approximate the density with the squared Hermite polynomials to estimate the RND for the prediction of observed implied volatility.

Our method can be considered as a nonparametric method as it does not rely on a data generating process and we do not use either a basis projection or a kernel smoothing technique for the density estimation. In the next section we show how to estimate the density.

3.2.3 Model estimation

Let us define $\mathbb{H} = L^2([0, +\infty))$ the space of square integrable functions mapping from the interval $[0, +\infty)$ to the set of real numbers $\mathbb{R}$. $\mathbb{H}$ is a Hilbert-space endowed with an inner product $< ., . >$ and a norm $||.||$, which are respectively defined as follows : $< f, g >= \int_0^{+\infty} f(t)g(t)dt$ and $||f|| = \left(\int_0^{+\infty} f^2 \right)^{1/2}$.

Let us consider the sample $((\kappa_1, Y_1), ..., (\kappa_n, Y_n))$ of independent pairs following the same distribution as well as in the population version (κ, Y). We consider the functional linear model where $(Z_i)_{i=1...n}$ is the sample of functional predictor variables of the regression representing the set of possible pay-offs at maturity for each option and $(Y_i)_{i=1...n}$ is the scalar response.

For each $i \in \{1, ..., n\}$, $Z_i(s) = max(\kappa_i - s, 0)$, then, Z_i is random only through κ_i. Additionally, the predictor function $(Z_i(s))_{i=1...n}$ is such that $Z_i(s) \geq 0$, this means that $E(Z_i(s)) \geq 0$ for each $s \in [0, +\infty)$ and $E(Y_i) > 0$. We assume that $E[\kappa_i^3] < +\infty$. Indeed, this result guarantees that $\int_0^{+\infty} E(Z^2(s))ds < \infty$, which means that the predictor function is square integrable (see Lemma 1 in appendix). Then, for each time t the model is a cross-sectional regression presented as follows :

$$Y_i = \int_0^{+\infty} Z_i(s)f(s)ds + \varepsilon_i \tag{3.3}$$

where $(f(s))_{s\in[0,+\infty)}$ is a function that belongs to the space $\mathbb{H}$ and $\varepsilon_i; i \in \{1, ..., n\})$ are independent and homoskedastic [1] such that $\mathbb{E}(\varepsilon_i|\kappa_1, ..., \kappa_n) = 0$ and $\mathbb{E}(\varepsilon_i^2|\kappa_1, ..., \kappa_n) = \sigma^2 < \infty$ for each $i \in \{1, ..., n\}$).

By premultiplying both sides of Equation (3.3) by $Z_i(u)$ and take the expectation, we obtain

1. Heteroskedasticity could be introduced if one considers options with different time-to-maturity and the same estimation procedure will be applied after a rescalling procedure of the data in order to obtain homoskedastic error.

$$\mathbb{E}[Z_i(u)Y_i] = \int_0^{+\infty} \mathbb{E}[Z_i(u)Z_i(s)]f(s)ds + \mathbb{E}[Z_i(u)\varepsilon_i].$$

Since $\mathbb{E}[Z_i(u)\varepsilon_i] = 0$, then

$$\mathbb{E}[Z_i(u)Y_i] = \int_0^{+\infty} \mathbb{E}[Z_i(u)Z_i(s)]f(s)ds.$$

In a compact form we can write $C_{zy} = K(f)$, where $C_{zy}(u) = \mathbb{E}[Z_i(u)Y_i]$ is the cross-covariance function between the predictor variable Z and the response variable Y and $k(u,s) = \mathbb{E}[Z_i(u)Z_i(s)]$, for $s, u \in [0, +\infty)$ is the kernel of the covariance operator K of the predictor function. That equation is also known as the normal equation associated with the regression model.

Our main goal is to estimate the function f. If the operator K were invertible, we could estimate f using $f(s) = K^{-1}C_{zy}(s)$ for each $s \in [0, +\infty)$, which is a Fredholm equation of the first kind. But this is not possible to use this equation because we deal with an ill-posed problem as K is a bounded operator mapping from an infinite dimensional space $\mathbb{H}$ to $\mathbb{H}$. This means that the direct inverse of K is not continuous and K is not invertible in $\mathbb{H}$ but only on a subset of $\mathbb{H}$. Estimating f by $\hat{K}^{-1}\hat{C}_{zy}$ would lead to an unstable estimator of the functional parameter. Additionally, this estimator is not continuous. To overcome this issue, we propose to use a regularization technique exploiting the functional data analysis framework called the functional Landweber-Fridman (FLF) method, in order to get a continuous inverse operator and therefore a continuous estimated function. This will also guarantee to obtain a more stable estimator of the density f. In the next section, we present the FLF method.

3.2.4 The functional Landweber-Fridman method

Recall that we have the normal equation related to the regression model

$$C_{zy} = K(f).$$

The main idea of the the functional Landweber-Fridman method is to approach the solution to this equation by an iterative algorithm similar to the fixed point procedure with the goal of minimizing the objective function of the regression problem. The idea is somewhat to the gradient descent technique commonly used for an optimization problem. Instead of iterating all the way to convergence, the algorithm stops after a finite number of iterations. Here, the early termination regularizes the solution of the iterations. The algorithm is presented as follows

— At the first iteration, take $f_0(s) = dC_{zy}(s)$, for each $s \in [0, +\infty)$.

— For $h = 1, ..., \frac{1}{\alpha} - 1$, calculate

$$f_h(s) = f_{h-1}(s) + d(C_{zy}(s) - Kf_{h-1}(s)) \tag{3.4}$$

— where d is a positive parameter such that $0 < ||K||_{op} \leq 1/d$. It is also called the relaxation parameter.
— For convenience, the resulting estimator f_h is denoted f_α with $f_\alpha(s) = K_\alpha^{-1} C_{zy}(s)$, for each $s \in [0, +\infty)$ and K_α^{-1} denotes the regularized inverse of K.

The regularized inverse of K is given by

$$K_\alpha^{-1}(\phi)(s) = d \sum_{l=1}^{\frac{1}{\alpha}-1} (I - dK)^l(\phi)(s)$$

where $s \in [0, +\infty)$. The parameter α is the regularization parameter and it will be chosen via a data driven method. Let us denote $(\lambda_j, v_j)_{j \geq 1}$ as the eigensystem of the covariance operator K, then we can also write K_α^{-1} in terms of the eigensystem of K as follows

$$K_\alpha^{-1}(\phi) = \sum_{j=1}^{\infty} \frac{Q(\alpha, \lambda_j)}{\lambda_j} < v_j, \phi > v_j$$

for each function ϕ and $Q(\alpha, \lambda_j) = \left[1 - (1 - d\lambda_j)^{1/\alpha}\right]$.

K_α^{-1}, K and C_{zy} are unobservable since they are derived on the population version of the model. Then, for the estimation they are replaced by their empirical versions. Then the estimated density function is given by $\hat{f}_\alpha(s) = \hat{K}_\alpha^{-1} \hat{C}_{zy}(s)$. In other words, we have

$$\hat{f}_\alpha(s) = d \sum_{l=1}^{\frac{1}{\alpha}-1} (I - d\hat{K})^l \hat{C}_{zy}(s) \tag{3.5}$$

where

$$\hat{K} = \frac{1}{n} \sum_{i=1}^{n} (Z_i - \overline{Z}) \otimes (Z_i - \overline{Z})$$

$$\hat{k}(u, s) = \frac{1}{n} \sum_{i=1}^{n} (Z_i(u) - \overline{Z}(u))(Z_i(s) - \overline{\tilde{Z}}(s))$$

is the estimated kernel of the operator $\hat{K}$, for $s, u \in [0, +\infty)$, and

$$\hat{C}_{zy} = \frac{1}{n} \sum_{i=1}^{n} (Z_i - \overline{Z}) \otimes (Y_i - \overline{Y}).$$

$\overline{Z}(s) = \frac{1}{n} \sum_{i=1}^{n} Z_i(s)$ for each $s \in [0, +\infty)$ and $\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ are sample mean of the predictor functions and response variable respectively.

3.2.5 Density correction

Our estimator of f, $\hat{f}_\alpha$, is not necessarily positive and does not integrate to one. As the true function f is a density, we propose to transform our estimator $\hat{f}_\alpha$ into a density using the methods proposed by Glad, Hjort, and Ushakov (2003). The correction is different depending on whether $\int_0^{+\infty} \max\left\{0, \hat{f}_\alpha(s)\right\} ds \geq 1$ or $\int_0^{+\infty} \max\left\{0, \hat{f}_\alpha(s)\right\} ds < 1$.

Case 1 : Case where $\int_0^{+\infty} \max\left\{0, \hat{f}_\alpha(s)\right\} ds \geq 1$.

The corrected estimator is given by

$$\tilde{f}_\alpha(s) = \max\left\{0, \hat{f}_\alpha(s) - \xi\right\}$$

where ξ is a positive constant chosen so that $\int_0^{+\infty} \tilde{f}_\alpha(s)\, ds = 1$.

Case 2 : Case where $\int_0^{+\infty} \max\left\{0, \hat{f}_\alpha(s)\right\} ds < 1$.

The corrected estimator $\check{f}_\alpha$ is computed as follows

$$\check{f}_\alpha = \begin{cases} \max\left\{0, \hat{f}_\alpha\right\} + \eta_M & \text{for } |s| \leq M, \\ \max\left\{0, \hat{f}_\alpha\right\} & \text{for } |s| > M, \end{cases}$$

where

$$\eta_M = \frac{1}{2M}\left[1 - \int_0^{+\infty} \max\left\{0, \hat{f}_\alpha(s)\right\} ds\right].$$

Remarks 1.

1. In Case 1, Glad, Hjort, and Ushakov (2003, Theorem 1) show that $\tilde{f}_\alpha$ is always better than $\hat{f}_\alpha(s)$ in the sense that $\left\|\tilde{f}_\alpha - f\right\|^2 \leq \left\|\hat{f}_\alpha - f\right\|^2$ so that the MISE of $\tilde{f}_\alpha$ is always smaller than that of $\hat{f}_\alpha$.

2. In Case 2, Glad, Hjort, and Ushakov (2003, Theorem 2) show

$$E\left\|\check{f}_\alpha - f\right\|^2 \leq E\left\|\hat{f}_\alpha - f\right\|^2 + \frac{3}{2M}.$$

Hence, one can make the MSE of $\check{f}_\alpha$ arbitrary close to that of $\hat{f}_\alpha$ by choosing M dependent of n and large, for instance $M = n$.

3. An algorithm to select ξ in practice is presented in Appendix.

4. An alternative correction of $\hat{f}_\alpha$ could have relied on a normalization

$$\frac{\max\left\{0, \hat{f}_\alpha\right\}}{\int_0^{+\infty} \max\left\{0, \hat{f}_\alpha(s)\right\} ds}.$$

However, there is no garantee that this normalization improves the accuracy of the estimator. The MSE of the normalized estimator may actually be worse than that of

the original estimator as discussed in Glad et al. (2003).

3.2.6 Alternative estimator

The estimated conditional density presented in the previous section may not be twice differentiable. Then, to control the smoothness of the conditional density, one can impose a penality parameter δ on the squared norm of the differentiated density function. Let us denote f'' the second derivative of the conditional density.

$$\hat{f}_{\alpha,\delta} = \underset{f\in L^2([0,+\infty)}{argmin} \left\{ \frac{1}{n}\sum_{i=1}^{n} \left[(Y_i - \overline{Y}) - \int_0^{+\infty} [Z_i(s) - \overline{Z}(s)] f(s) ds \right]^2 + \delta \int_0^{+\infty} (f''(s))^2 ds \right\}.$$

Then, using the Landweber-Fridman regularization approach for estimation, we obtain for each $s \in [0, +\infty)$

$$\hat{f}_{\alpha,\delta}(s) = V\hat{K}^{-1}_{\alpha,\delta}\hat{C}_{zy}(s)$$

with $\hat{f}_{\alpha,\delta} \in L^2([0,+\infty)$ and for each function ϕ

$$\hat{K}^{-1}_{\alpha,\delta}(\phi) = d\sum_{l=1}^{\frac{1}{\alpha}-1} (I - d(\hat{K} + \delta D))^l(\phi)(s)$$

where V is second integral operator defined as

$$V(\phi)(s) = \int_{-\infty}^{s} \left(\int_{-\infty}^{u} \phi(t)dt \right) du.$$

Then, the density correction is applied in order to get a function respecting the density properties. In the rest of the paper, we will not consider this estimator as differentiability is not one of our concerns

3.3 Convergence rate of the risk neutral density

In this section, we analyze the convergence rate of the conditional mean square error (MSE) of $\hat{f}_\alpha$. For this purpose, I use the following assumptions :

A1. (κ_i, Y_i) are i.i.d with the same distribution law as (κ, Y) with $E[\kappa_i^3] < \infty$.

A2. $\int_0^{+\infty} f^2(t)dt < +\infty$, $E[\varepsilon_i|\kappa_1, ..., \kappa_n] = 0$, $E[\varepsilon_i^2|\kappa_1, ..., \kappa_n] = \sigma^2 < +\infty$, $E[\varepsilon_i^4|\kappa_1, ..., \kappa_n] < +\infty$, and $E[\kappa_i^4] < +\infty$.

A3. The eigenvalues of the covariance operator K and the estimated one are distinct, that is $\lambda_1 > \lambda_2 > ... > 0$ and $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_n$.

A4. We assume that for $\mu \geq 0$, f satisfies

$$\sum_{j=1}^{\infty} \frac{< f, v_j >^2}{\lambda_j^{\mu}} < \infty.$$

Assumption **A1** imposes that $(Z_i, Y_i)_{i=1,...n}$ are independent, identically distributed as (Z, Y). It is useful in order to derive the consistency of the covariance operators $\hat{K}$ and $\hat{C}_{zy}$, and to prove the central limit properties of the estimated functions. The fact that $E[\kappa_i^3] < +\infty$ guarantees that the predictor functions $(Z_i)_{i=1,...,n}$ are square integrable. Moreover, it also guarantees that the covariance operator K is nuclear, which in turn is Hilbert-Schmidt (see Lemma 1 in Appendix). This condition is directly obtained if we assume that the predictor variables κ_i follow a Gaussian process.

Assumption **A2** imposes that the error term ε_i is homokedastic and Z_i is exogenous. **A1** and **A2** are sufficient conditions to ensure that $||\hat{K} - K||_{HS}^2 = O_p\left(\frac{1}{n}\right)$, see Proposition 5 of Dauxois et al. (1982). $||.||_{HS}$ is the Hilbert-Schmidt norm of operators.

Assumption **A3** is used to simplify the conditions under which the estimated operators $\hat{K}$ and $\hat{C}_{zy}$ are consistent. Morevover, this assumption ensures that the null space of K, $\mathcal{N}(K) = \{0\}$. Hence, f is the unique solution of $C_{xy} = K(f)$.

Assumption **A4** is a source condition important to derive how the bias and estimation error terms behave. μ is a parameter controlling the degree of smoothness of the true density f. Therefore, as μ becomes more larger, f becomes smoother. It imposes that the eigenvalues of K decline to zero not too fast relatively to the Fourier coefficients of f, $< f, v_j >$. In the inverse problem literature, this parameter characterizes the severity of the ill-posed problem. As μ becomes larger, the ill-posed problem becomes more severe, i.e the eigenvalues λ_j decay more faster (see proposition 3.13 of Engl et al. (1996)).

Proposition 1

Under assumptions A1 - A4, if $\alpha^2 n \to \infty$, then

$$\mathbb{E}\left[\left|\left|\ddot{f}_{\alpha} - f\right|\right|^2 |\kappa_1, ..., \kappa_n\right] = O_p\left(\alpha^{\mu}\right) + O_p\left(\frac{1}{\alpha^2 n}\right) \tag{3.6}$$

and the conditional MSE converges to zero as the sample size increases. Where $\ddot{f}_{\alpha}$ is the estimated density function corrected either with Case 1 or 2.

Remarks 2.

— The first term of the conditional MSE is the squared bias and the second term is the estimation error.

— The bias term vanishes when $\alpha \to 0$.

— As α goes to zero, the squared bias term goes to zero, while the estimation error term increases. Then, there is a trade-off to make in order to guarantee the convergence of the MSE. The optimal parameter α is selected in such a way that the squared bias is equal to the variance term.

— If $\alpha \sim n^{-1/(2+\mu)}$, then $MSE \sim n^{-\frac{\mu}{2+\mu}}$.

3.4 Asymptotic normality

This section focuses on deriving the asymptotic distribution of the estimated density $\tilde{f}_\alpha$, for a fixed value of the tuning parameter α. It follows from theorem 2.7 by Bosq (2000) that under assumption A1,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_i - \overline{Z}) \xrightarrow{d} \mathcal{N}(0, K).$$

Proposition 2

Let α be fixed. Under assumptions A1 - A4, if $\Gamma_\alpha < +\infty$ then

$$\sqrt{n}(\hat{f}_\alpha - f_\alpha) \xrightarrow{d} \mathcal{N}(0, \Gamma_\alpha), \tag{3.7}$$

as $n \to +\infty$, where Γ_α is defined as

$$\Gamma_\alpha = \sigma^2 K_\alpha^{-1} K K_\alpha^{-1} + \mathbb{E}\Big[[K_\alpha^{-1}(Z_i \otimes Z_i - K)] \tilde{\otimes} [(Z_i \otimes Z_i - K) K_\alpha^{-1}]\Big](f \otimes f)$$

with

$$K_\alpha^{-1}(\phi) = d \sum_{l=1}^{\frac{1}{\alpha}-1} (I - dK)^l(\phi)(s)$$

and σ^2 is the variance of the error term.

Corollary 1

Based on the asymptotic normality results, we can derive the confidence interval of the estimated density for a fixed value of the tuning parameter α. Then, a $100(1-\theta)\%$ pointwise confidence interval of $f(s)$, for any $s \in [0, +\infty)$, is given by

$$\mathcal{IC}_{1-\theta} = \hat{f}_\alpha(s) \pm \phi^{-1}(1-\theta/2)[\hat{\Gamma}_\alpha(s,s)]^{1/2}.$$

where $\phi^{-1}(1-\theta/2)$ is the quantile of order θ in the standard normal distribution.

Moreover, a $100(1-\theta)\%$ prediction interval for the average option price $\mathbb{E}[Y_i|\kappa_1, ..., \kappa_n] = \int_0^\infty Z_i(s)\hat{f}_\alpha(s)ds$ is approximated by

$$\mathcal{IC}_{1-\theta}(Y_i) = \hat{Y}_i \pm \phi^{-1}(1-\theta/2)\mathcal{G}_\alpha^{1/2},$$

where

$$\mathcal{G}_\alpha = \mathbb{E}\left[\int_0^\infty \int_0^\infty Z_i(s)\hat{\Gamma}_\alpha(s,t)Z_i(t)dsdt\right].$$

The second result follows from the fact that our density estimator is linear in Y. Indeed, as $\hat{f}_\alpha$ converges in probability to f, then $\hat{Y}_i$ converges in probability to $\mathbb{E}[Y_i|\kappa_i]$ along with the asymptotic distribution under a linear transformation and using the asymptotic distribution of $\hat{f}_\alpha$.

3.5 Data-driven selection of the tuning parameter

According to the consistency results, it can be noticed that the estimation of the RND depends on the tuning parameter α. Then, this parameter should be selected optimally. Since the main goal is to estimate the RND and therefore predict the call and put prices, we define a prediction criterion to select the optimal parameter α. Then, we choose the regularization parameter in such a way that the mean squared prediction error (MSPE) is minimized. We use the K-fold cross-validation for the selection procedure. Let us split the initial sample in M subsamples denoted I_1,...,I_M.

$$\alpha_{op} = \underset{\alpha\in\mathcal{I}_\alpha}{\operatorname{argmin}} \frac{1}{M}\sum_{\ell=1}^{M}\frac{1}{card(I_\ell)}\sum_{j\in I_\ell}\left(Y_j - \hat{Y}_j\right)^2. \tag{3.8}$$

For $\ell \in \{I_1,...,I_M\}$, we estimate the parameter β in the sample $\mathcal{I}_{-\ell}$ representing the observations not in $\mathcal{I}_\ell$. Then, we predict the response variable in $\mathcal{I}_\ell$ considered as the hold-out sample. $\hat{Y}_j$ is the prediction of the j^{th} observation in $\mathcal{I}_\ell$. Hence, we calculate the MSPE for each candidate α. $\mathcal{I}_\alpha$ is the set of candidate α. Indeed, it is a generalization of the leave-one-out cross-validation.

An alternative approach suggested by Engl et al. (1996) is to choose the parameter α such that the following objective function is minimized.

$$\alpha_{op} = \underset{\alpha\in\mathcal{I}_\alpha}{\operatorname{argmin}} \left|\left|\hat{f}_\alpha\right|\right|^2 \left|\left|\hat{C}_{zy} - \hat{K}(\hat{f}_\alpha)\right|\right|^2. \tag{3.9}$$

3.6 Simulations

This section presents simulation results based on the proposed estimation approach. For this purpose, we consider different variety of data generating processes (DGP). From the DGP, the true option prices and RND are observed and we use the proposed estimation method to derive it and study the estimation and prediction performance. We use the K-folds cross-validation to select the optimal α, with $K = 10$. The data generating process considered in this section is the same as the one suggested by Bondarenko (2003). We consider the functional regression model

$$Y_i = \int_0^{+\infty} Z_i(s) f(s) ds + \varepsilon_i$$

where Z and Y are centred random variables. The RND $f(s)$ is specified as a lognormal mixture presented as follows

$$f(s) = \pi_1 \mathcal{LN}(s|\eta_1, \sigma_1) + \pi_2 \mathcal{LN}(s|\eta_2, \sigma_2) + \pi_3 \mathcal{LN}(s|\eta_3, \sigma_3),$$

$$\pi_1 + \pi_2 + \pi_3 = 1,$$

and $\mathcal{LN}(s|\eta_j, \sigma_j)$, $j = 1, 2, 3$ is a lognormal distribution

$$\mathcal{LN}(s|\eta_j, \sigma_j) = \frac{1}{\sqrt{2\pi}\sigma_j s} e^{-\frac{(ln\frac{s}{\eta_j} - \frac{1}{2}\sigma_j^2)^2}{2\sigma_j^2}}.$$

Put options prices with 1-month maturity (21 business days) are generated. The set of strike prices is defined as follows $\mathcal{K} = [430, 431, 432, ..., 540]$, which means that the sample size is $n = 111$ and the underlying follows the aforementioned lognormal mixture distribution. This parametrization is used to match a typical cross-section of the $S\&P$ 500 index options traded at the Chicago Board Options Exchange (CBOE) on March 21, 1995. The closing prices on March 21, 1995 of $S\&P$ 500 options with the maturity date on April 21, 1995 are used to calibrate the RND. The parameters of the RND are presented in Table 3.1.

TABLE 3.1: Parameters of the lognormal mixture density

lognormal	π	η	σ
$\mathcal{LN}(s\|\eta_1, \sigma_1)$	0.1194	475.59	0.0550
$\mathcal{LN}(s\|\eta_2, \sigma_2)$	0.8505	498.17	0.0206
$\mathcal{LN}(s\|\eta_3, \sigma_3)$	0.0301	524.91	0.0146

The predictor functions are defined as $Z_i(s) = max(\kappa_i - s, 0)$, with $i = 1, ...111$, $\kappa_i \in \mathcal{K}$ and s is the underlying. The response variable $Y_i = exp(r\tau)P_i$ and P_i are the generated put option prices. In the simulations we set $exp(r\tau) = 1$. Each predictor function $Z_i(s)$ is represented as a function of a very dense set of 1000 possible underlying values $s \in [493, 523]$ (493 and 523 representing the S&P 500 price at the end of March March 21, 1995 and April 21, 1995 respectively).

To characterize the incompleteness of the market, the error term ε_i is added to the regression model with $\mathbb{E}[\varepsilon_i] = 0$. Additionally, the error term is generated to be proportional to the bid-ask spread $\mathcal{S}$ as follows : $\varepsilon_i \sim \mathcal{U}\Big([-0.5\mathcal{S}, +0.5\mathcal{S}]\Big)$, $i = 1, ..., n$.

Assuming that the prices and pay-offs are all centered, the put prices are calculated

with the following equation

$$Y_i = \int_0^{+\infty} Z_i(s) f(s) ds + \varepsilon_i.$$

All the numerical integrations are performed with the trapezoidal rule. It is also possible to use other integration rule such as the Newton-Cotes or adaptative quadrature.

We simulate 1000 copies of the datasets and evaluate the performance of the estimation method with 2 criteria :

— The Root Mean Squared Prediction Error (RMSPE) between the estimated put prices $\hat{Y}_\alpha$ and the theoretical one Y.

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_{i,\alpha} - Y_i)^2}.$$

— The Root Integrated Squared Error (RISE) between the estimated density $\hat{f}_{\alpha,\delta}$ and the theoretical one f.

$$RISE = \frac{1}{||f||}\sqrt{\int_0^{+\infty}(\hat{f}_\alpha(s) - f(s))^2 ds}.$$

We compare the proposed estimation method with the positive convolution approach (PosConv) suggested by Bondarenko (2003) as it is widely used as a valuable benchmark.

Figure 3.1 shows that the results of the estimated RND and the related cumulative distribution function, when the error term $\varepsilon_i \sim \mathcal{U}\Big([-0.5\mathcal{S}, +0.5\mathcal{S}]\Big)$ with $\mathcal{S} = 1$. It can be noticed that the estimated density tends to mimic the true one, but the magnitudes is slightly different. The estimated RND is unimodal and displays a bell formation with tiny tails and a rise around the strike 500. Table 3.2 shows a comparison of our method to the one suggested by Bondarenko (2003). It can be observed that in terms of estimation of the density (RISE) the PosConv outperforms our method, while in terms of prediction of the option price (MSPE) our method tends to be more accurate mean squared prediction error (MSPE). We observe almost the same result when a different variability of the error term ($\mathcal{S} = 2$) as shown in Figure 3.2 and Table 3.3. In addition, our Monte-Carlo experiments show that for the 1000 simulations, approximately $99,72\%$ of the estimated density are such that $\int_0^\infty max\Big\{0, \hat{f}_\alpha(s)\Big\} ds \leq 1$. Therefore we use the density correction of Case 2 in order to get the integration to one. It is worth nothing that this density correction can change from Case 2 to Case 1 according to what we obtain of the estimated density form the sample data.

The performance level of the proposed method is maybe due to the fact that the method is nonparametric, which means that it is data-demanding. In the considered data generating process the sample size is relatively small and the PosConv method

is semi-parametric and therefore adapted for the data generating process. The goal of Bondarenko (2003) was to produce an alternative estimation method of the RND regardless the interpretability of the results, and the evaluation of the method is only made on the estimation performance and not in terms of prediction (see page 100 of Bondarenko (2003)). In contrast, our main object is to develop a realistic estimator that relies of the economic fundamental and can be used for prediction purpose. Furthermore, Figure 3.3 shows the true smoothed and predicted put prices. It can be noticed that the put prices are reasonably well predicted. This reinforces the argument about the good prediction performance of the proposed method.

3.7 Real data application

This section focuses on a real data example. For this purpose, we consider the S&P 500 index (SPX) as the data of interest to derive the underlying and the strikes. The S&P 500 option is one of the most liquid and tradable asset in the market. It represents the aggregated capitalization of the 500 most important corporations in the U.S. It is also used as a benchmark to see how well the most important companies are behaving. I exploit the available public data by Barratt et al. (2020), extracted from OptionMetric Ivy database via the Wharton Research Data Services.

From the database, the best bid and ask price of all S&P 500 European options are collected for the date of June 3, 2019 with a maturity of 25 days. The price of the index at the end of the same day is also collected. Indeed, it was evaluated at 2744.45 dollars. The expiration price from 1500 to 3999.50 dollars is discretized in 50 cent increments and it leads to a collection of 6000 price values representing the set of underlying. Puts and calls options are both considered in the sample. Also opportunities to buy or sell the puts and calls are allowed.

The price of each option investment is equal to the ask price if buying and the bid price if selling, plus a transaction cost of 65 cent for each option in the sample. The transaction cost when buying or selling the underlying as an investment is 0.3% of the investment. These cost are the one used by TD Ameritrade brokerage at that time. The sample size obtained is $n = 316$ (for more details, see Barratt et al. (2020)).

The results are presented in Figure 3.4. The estimated distribution from the 25-days maturity options displays a bell-shaped formation centrered around the value 2744 dollars. The MSPE from the estimation is equal to 0.91. This density also presents a long tail on the left hand side and a small hill around the strike value 2744.5. This shape tends to be similar to a lognormal distribution shape. Figure 3.5 presents the predicted call and put prices with their respective 95% confidence interval. It can be noticed that the options prices are well predicted via the suggested estimation approach.

3.8 Conclusion

This paper proposes to estimate the risk neutral density for options pricing models with the functional data analysis framework. Indeed, we consider that a European option price of an asset is evaluated as a weighted average of all possible payoffs of the asset, where the weights represents here the risk neutral density of a market participant. To use the functional data analysis framework, we assume that for each asset, one can have an infinity of possible price values at the maturity. This means that at the maturity date, a market participant is exposed to an infinite possibility of payoff. The set of potential payoffs for each option price is then very dense and is considered as a function. On the same line, the risk neutral density is also considered as a collection of values observed on a very fine grid. Therefore, the model setting considered is a functional linear model where the predictors functions are represented by the set of potential payoffs and the response variable is the price of the option, which is a scalar. In contrast to the methods proposed in prior papers, the estimation method proposed in this paper is free of any parametric or semi-parametric configuration and it also takes into account the arbitrage-free theory for options pricing in the model setting.

One of the main issue of this model is the high dimensionality problem as the inverse of the covariance operator of the predictor variable is not continuous. This problem leads to unstable estimated function. To overcome this issue we propose to use a regularization technique called the Functional Landweber-Fridman. We also control for the positivity and the integration to one of the density by applying a density correction. We derive the consistency and asymptotic normality of the estimated density function. Additionally, we provide simulations and an application to evaluate the quality of the estimation method.

Based on the simulations we obtained a reasonable bell-shaped formation of the estimated risk neutral density. Comparing the results of this paper with the one obtained by Bondarenko (2003), it is easy to observe that the approach proposed by Bondarenko (2003) outperforms in terms of estimation of the density but underperforms in terms of prediction of the options prices. The performance of their approach is due to the parametrization considered in their model. The advantage of the approach proposed in this paper is that it lets the data derive the results. The results from real data on S&P 500 options confirm the expected results from the simulations. The proposed approach in this paper can be considered as a promising alternative to the existing ones.

3.9 Appendix

Lemma 1.

If $E[\kappa_i^3] < +\infty$, then K is a nuclear operator and $(Z_i(s))_{i=1,\ldots,n}$ are square integrable functions.

Proof.

Note that

$$\begin{aligned} k(u,s) &= E(Z_i(u) Z_i(s)) \\ &= E(\max(\kappa_i - u, 0) \max(\kappa_i - s, 0)). \end{aligned}$$

Given κ_i has a continuous density on $\mathbb{R}^+$, $k(u,s)$ is continuous on $\mathbb{R}^+ \times \mathbb{R}^+$. Moreover, K is a positive definite operator. Hence to check whether K is nuclear, it is enough to check that

$$\int k(s,s)\ ds < \infty.$$

See for instance, Ferreira, Menegatoo, and Oliveira (2009, Theorem 2.9).

$$\begin{aligned} \int k(s,s)\, ds &= E \int \max(\kappa_i - s, 0)^2\, ds \\ &= E \int_0^{\kappa_i} (\kappa_i - s)^2\, ds \\ &= \frac{1}{3} E(\kappa_i^3) < \infty. \end{aligned}$$

Recall that nuclear implies Hilbert-Schmidt. This proves that K is a nuclear operator and therefore is Hilbert-Schmidt.

* Similarly, one can show that

$$E(\kappa_i^3) < \infty \Rightarrow \int E(Z_i^2(s))\, ds < \infty.$$

We can conclude that Z_i is square integrable.

3.9.1 Algorithm for the correction of the estimator in Case 1

To determine ξ in Case 1, we follow the algorithm outlined by Luedicke and Bernacchia (2014).

First compute $\int \max\left\{0, \hat{f}_\alpha(s)\right\} ds$ using an equally spaced grid. Let N be the number of grid points and Δ the grid interval used in the numerical computation of the integral.

Then, the starting point of the algorithm is

$$\xi_0 = \frac{\int \max\left\{0, \hat{f}_\alpha(s)\right\} dx}{N\Delta}.$$

Let δ be such as $\delta = 10\tau\xi_0$ for some small τ (for instance $\tau = 10^{-4}$).

The search is iterated by adding δ to ξ_l at each iteration

$$\xi_{l+1} = \xi_l + \delta$$

until

$$1 \leq \int \tilde{f}_\alpha(s)\, ds \leq 1 + \tau.$$

This algorithm is fast but not guaranteed to converge. Luedicke and Bernacchia (2014) proposed another algorithm which is sure to converge. In the simulations, we implemented only the first one.

3.9.2 Proof of Proposition 1

We have

$$\hat{f}_\alpha - f = (\hat{f}_\alpha - f_\alpha) + (f_\alpha - f)$$

where the first term of the right hand side of the equation is the estimation error and the second term is the regularization bias. From that equation, we have

$$||\hat{f}_\alpha - f||^2 \leq 2||\hat{f}_\alpha - f_\alpha||^2 + 2||f_\alpha - f||^2.$$

To derive the convergence rate of $||\hat{f}_\alpha - f||^2$, it is sufficient to derive the convergence rate of the squared bias and the estimation error.

Convergence rate of $||f_\alpha - f||^2$:

We have

$$f_\alpha = \sum_{j=1}^{+\infty} < v_j, f > v_j.$$

$$f_\alpha = \sum_{j=1}^{+\infty} Q(\alpha, \lambda_j) < v_j, f > v_j.$$

where $Q(\alpha, \lambda_j) = 1 - (1 - \lambda_j)^{1/\alpha}$. Then,

$$f - f_\alpha = \sum_{j=1}^{+\infty}[1 - Q(\alpha, \lambda_j)] < v_j, f > v_j.$$

and

$$\begin{aligned}
||f - f_\alpha||^2 &= \sum_{j=1}^{+\infty}[1 - Q(\alpha, \lambda_j)]^2 < v_j, f >^2 \\
&= \sum_{j=1}^{+\infty} \lambda_j^\mu [1 - Q(\alpha, \lambda_j)]^2 \frac{< v_j, f >^2}{\lambda_j^\mu} \\
&\leq \underset{j}{Sup}\left[\lambda_j^\mu [1 - Q(\alpha, \lambda_j)]^2\right] \sum_{j=1}^{+\infty} \frac{< v_j, f >^2}{\lambda_j^\mu} \\
&\leq \underset{j}{Sup}\left[\lambda_j^{\mu/2} [1 - Q(\alpha, \lambda_j)\right]^2 \\
&\leq C\alpha^\mu,
\end{aligned}$$

where C is an arbitrary positive constant. The third line of the equation holds since $\sum_{j=1}^{+\infty} \frac{<v_j, f>^2}{\lambda_j^\mu} < +\infty$ according to assumption 3. Following Proposition 3.11 of Carrasco et al. (2007) [2], $\underset{j}{Sup}\left[\lambda_j^{\mu/2}[1 - Q(\alpha, \lambda_j)\right]^2 \leq C\alpha^\mu$, which yields the last line.

Convergence rate of $||\hat{f}_\alpha - f_\alpha||^2$:

$$\begin{aligned}
\hat{f}_\alpha - f_\alpha &= \hat{K}_\alpha^{-1}\hat{C}_{zy} - K_\alpha^{-1}C_{zy} \\
&= \hat{K}_\alpha^{-1}\hat{K}(f) - K_\alpha^{-1}K(f) + \hat{K}_\alpha^{-1}(\hat{V}_{z\varepsilon}),
\end{aligned}$$

where $\hat{V}_{z\varepsilon}(s) = \frac{1}{n}\sum_{i=1}^n (Z_i(s) - \overline{Z}(s))(\varepsilon_i - \overline{\varepsilon})$. The last line holds since $\hat{C}_{zy} = \hat{K}(f) + \hat{V}_{z\varepsilon}$ and $C_{zy} = K(f)$.

Following Lemma 3 of Tsafack (2020), $||\hat{K}_\alpha^{-1}\hat{K}(f) - K_\alpha^{-1}K(f)||^2 = O_p\left(\frac{\alpha^\mu}{\alpha^2 n}\right)$. Furthermore,

$$\hat{K}_\alpha^{-1}(\hat{V}_{z\varepsilon}) = \sum_{j=1}^n \frac{Q(\alpha, \hat{\lambda}_j)}{\hat{\lambda}_j} < \hat{v}_j, \hat{V}_{z\varepsilon} > \hat{v}_j.$$

Then,

2. The eigenvalues of K are denoted λ_j^2 in Carrasco et al. (2007) whereas they are denoted λ_j in the present paper.

$$
\begin{aligned}
||\hat{K}_\alpha^{-1}(\hat{V}_{z\varepsilon})||^2 &= \sum_{j=1}^{n} \frac{[Q(\alpha,\hat{\lambda}_j)]^2}{\hat{\lambda}_j^2} < \hat{v}_j, \hat{V}_{z\varepsilon} >^2 \\
&\leq \underset{j}{Sup}\left[\frac{[Q(\alpha,\hat{\lambda}_j)]^2}{\hat{\lambda}_j^2}\right] \sum_{j=1}^{n} < \hat{v}_j, \hat{V}_{z\varepsilon} >^2 \\
&\leq \left[\frac{d}{\alpha}\right]^2 \sum_{j=1}^{n} < \hat{v}_j, \hat{V}_{z\varepsilon} >^2 .
\end{aligned}
$$

The last line follows from page 5678 of Carrasco et al. (2007). Furthermore,

$$
\begin{aligned}
\mathbb{E}[< \hat{v}_j, \hat{V}_{z\varepsilon} >^2 |Z] &= \mathbb{E}\left[\int_0^\infty (\hat{v}_j(t)\frac{1}{n}\sum_{i=1}^{n}(Z_i(t) - \bar{Z}(t))\varepsilon_i)dt \int_0^\infty (\hat{v}_j(s)\frac{1}{n}\sum_{l=1}^{n}(Z_l(s) - \bar{Z}(s))\varepsilon_l)ds|Z\right] \\
&= \int_0^\infty \int_0^\infty \hat{v}_j(t)\mathbb{E}\left[(\frac{1}{n}\sum_{i=1}^{n}(Z_i(t) - \bar{Z}(t))\varepsilon_i)(\frac{1}{n}\sum_{l=1}^{n}(Z_l(s) - \bar{Z}(s))\varepsilon_l)\right]\hat{v}_j(s)dsdt \\
&= \int_0^\infty \int_0^\infty \hat{v}_j(t)\left[\frac{\sigma^2}{n}\sum_{i=1}^{n}(Z_i(t) - \bar{Z}(t))(Z_l(s) - \bar{Z}(s))\right]\hat{v}_j(s)dsdt \\
&= \frac{\sigma^2}{n} < \hat{v}_j, \hat{K}(\hat{v}_j) > \\
&= \frac{\sigma^2\hat{\lambda}_j}{n}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}[||\hat{K}_\alpha^{-1}(\hat{V}_{z\varepsilon})||^2|Z] &= \sum_{j=1}^{n} \frac{\hat{q}_j^2}{\hat{\lambda}_j^2}\frac{\hat{\sigma^2}\lambda_j}{n} \\
&= \frac{\sigma^2}{n}\sum_{j=1}^{n}\frac{\hat{q}_j^2}{\hat{\lambda}_j^2}\lambda_j \\
&\leq \frac{\sigma^2}{n} sup\left(\frac{\hat{q}_j^2}{\hat{\lambda}_j^2}\right)\sum_{j=1}^{n}\lambda_j \\
&\leq \frac{\sigma^2}{n\alpha^2}
\end{aligned}
$$

Hence, $||\hat{f}_\alpha - f_\alpha||^2 = O_p\left(\frac{\alpha^\mu}{\alpha^2 n}\right) + O_p\left(\frac{1}{\alpha^2 n}\right)$ and $||f_\alpha - f||^2 = O_p\left(\alpha^\mu\right)$. These results lead to

$$||\hat{f}_\alpha - f||^2 = O_p\left(\alpha^\mu\right) + O_p\left(\frac{1}{\alpha^2 n}\right),$$

Furthermore, if the estimated density function is adjusted with Case 1, we have

$$||\ddot{f}_\alpha - f||^2 \leq ||\hat{f}_\alpha - f||^2$$

If we use the second approach for the correction, we obtain

$$\begin{aligned} E[||\ddot{f}_\alpha - f||^2|\kappa_1, ..., \kappa_n] &\leq E[||\hat{f}_\alpha - f||^2|\kappa_1, ..., \kappa_n] + \frac{1}{n} \\ &= O_p\left(\alpha^\mu\right) + O_p\left(\frac{1}{\alpha^2 n}\right) + O_p\left(\frac{1}{n}\right). \end{aligned}$$

Therefore, in both cases, we obtain

$$||\ddot{f}_\alpha - f||^2 = O_p\left(\alpha^\mu\right) + O_p\left(\frac{1}{\alpha^2 n}\right),$$

This concludes the proof.

3.9.3 Proof of Proposition 2

For α fixed,

$$\hat{f}_\alpha - f_\alpha = \hat{K}_\alpha^{-1}\hat{K}(f) - K_\alpha^{-1}K(f) + \hat{K}_\alpha^{-1}(\hat{V}_{z\varepsilon}).$$

Then,

$$\begin{aligned} \hat{f}_\alpha - f_\alpha &= \hat{K}_\alpha^{-1}(\hat{V}_{z,\varepsilon}) + \hat{K}_\alpha^{-1}\hat{K}(f) - K_\alpha^{-1}K(f) \\ &= K_\alpha^{-1}\hat{V}_{z,\varepsilon} + (\hat{K}_\alpha^{-1} - K_\alpha^{-1})\hat{V}_{z,\varepsilon} + \hat{K}_\alpha^{-1}\hat{K}(f) - K_\alpha^{-1}\hat{K}(f) + K_\alpha^{-1}\hat{K}(f) - K_\alpha^{-1}K(f) \\ &= K_\alpha^{-1}\hat{V}_{z,\varepsilon} + (\hat{K}_\alpha^{-1} - K_\alpha^{-1})\hat{V}_{z,\varepsilon} + (\hat{K}_\alpha^{-1} - K_\alpha^{-1})\hat{K}(f) + \hat{K}_\alpha^{-1}(\hat{K}_\alpha - K_\alpha)(f) \\ &= K_\alpha^{-1}\hat{V}_{z,\varepsilon} + \hat{K}_\alpha^{-1}(\hat{K}_\alpha - K_\alpha)(f) + O_p\left(\frac{1}{n}\right) \\ &= K_\alpha^{-1}\left[\frac{1}{n}\sum_{i=1}^n Z_i \otimes \varepsilon_i\right] + K_\alpha^{-1}\left[\frac{1}{n}\sum_{i=1}^n (Z_i \otimes Z_i - K)(f)\right] + O_p\left(\frac{1}{n}\right). \end{aligned}$$

Let us define $\tilde{Z}_i = K_\alpha^{-1}(Z_i)$ and $\tilde{Z} = K_\alpha^{-1}(Z)$. Then,

$$\hat{f}_\alpha - f_\alpha = \frac{1}{n}\sum_{i=1}^{n} \tilde{Z}_i \otimes \varepsilon_i + \frac{1}{n}\sum_{i=1}^{n}[\tilde{Z}_i \otimes Z_i - \mathbb{E}[\tilde{Z} \otimes Z]](f) + O_p\left(\frac{1}{n}\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} \tilde{Z}_i \otimes \varepsilon_i + \frac{1}{n}\sum_{i=1}^{n}[\tilde{Z}_i \otimes Z_i - \mathbb{E}[\tilde{Z} \otimes Z]](f) + O_p\left(\frac{1}{n}\right).$$

Under the assumptions that $Z_i \otimes \varepsilon_i$ are i.i.d, $\mathbb{E}[\kappa_i^3] < +\infty$ we obtain

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{bmatrix} Z_i \otimes \varepsilon_i \\ Z_i \otimes Z_i - E[Z_i \otimes Z_i] \end{bmatrix} \overset{d}{\Longrightarrow} \mathcal{N}(0, \Omega_1)$$

where $\Omega_1 = \begin{pmatrix} G & 0 \\ 0 & K \end{pmatrix}$ is a (2×2) matrix of covariance operators. Let us define the following transformation

$$\begin{bmatrix} A \\ B \end{bmatrix} \to K_\alpha^{-1}A + K_\alpha^{-1}(B - K)$$

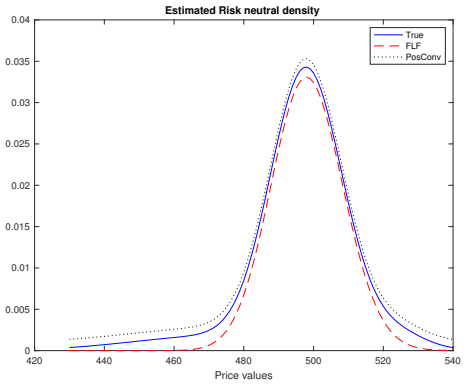
where A and B are two arbitrary operators. By the continuous mapping theorem with the previous transformation with $A = \frac{1}{n}\sum_{i=1}^{n} Z_i \otimes \varepsilon_i$ and $B = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i \otimes Z_i$, yields the asymptotic covariance operator of $\sqrt{n}(\hat{f}_\alpha - f_\alpha)$, Γ_α given by

$$K_\alpha^{-1}(B - K) = \frac{1}{n}\sum_{i=1}^{n}[\tilde{Z}_i \otimes Z_i - \mathbb{E}[\tilde{Z} \otimes Z]].$$

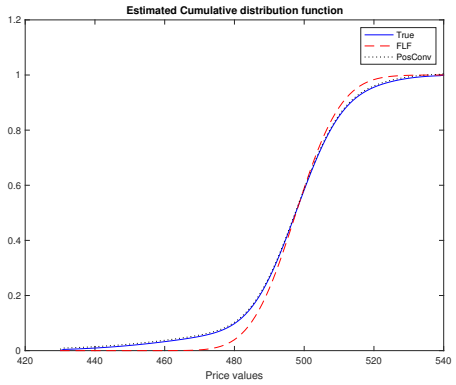
Therefore,

$$\Gamma_\alpha = \mathbb{E}\Big[(K_\alpha^{-1}(Z_i) \otimes \varepsilon_i)\tilde{\otimes}(\varepsilon_i \otimes K_\alpha^{-1}(Z_i))\Big] + \mathbb{E}\Big[[(K_\alpha^{-1}(Z_i) \otimes Z_i - K)]\tilde{\otimes}[(K_\alpha^{-1}(Z_i) \otimes Z_i - K)]\Big](f \otimes f)$$
$$= \mathbb{E}\Big[[K_\alpha^{-1}(Z_i \otimes \varepsilon_i)] \otimes [K_\alpha^{-1}(Z_i \otimes \varepsilon_i)]\Big] + \mathbb{E}\Big[[K_\alpha^{-1}(Z_i \otimes Z_i - K)]\tilde{\otimes}[(Z_i \otimes Z_i - K)K_\alpha^{-1}]\Big](f \otimes f)$$
$$= \mathbb{E}\Big[[K_\alpha^{-1}(Z_i \otimes \varepsilon_i)] \otimes [K_\alpha^{-1}(Z_i \otimes \varepsilon_i)]\Big] + \mathbb{E}\Big[[K_\alpha^{-1}(Z_i \otimes Z_i - K)]\tilde{\otimes}[(Z_i \otimes Z_i - K)K_\alpha^{-1}]\Big](f \otimes f)$$
$$= \sigma^2 K_\alpha^{-1} K K_\alpha^{-1} + \mathbb{E}\Big[[K_\alpha^{-1}(Z_i \otimes Z_i - K)]\tilde{\otimes}[(Z_i \otimes Z_i - K)K_\alpha^{-1}]\Big](f \otimes f).$$

3.9.4 Graphics and tables

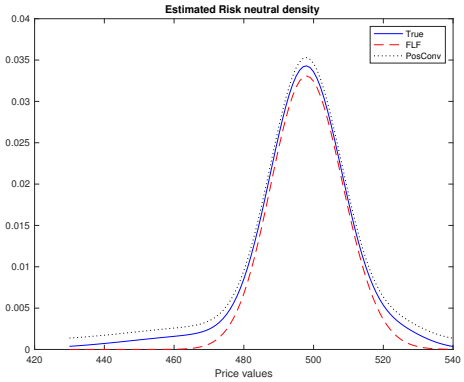


(a) Estimated risk neutral density

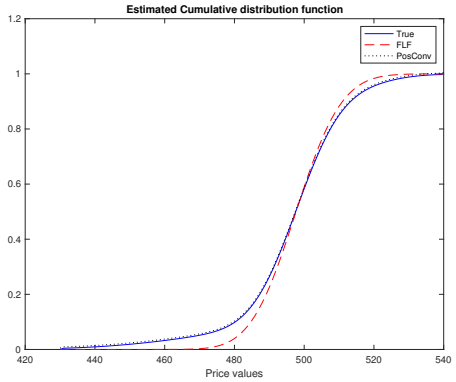


(b) Cumulative distribution function

FIGURE 3.1: Estimated risk neutral density and cumulative distribution function, $\mathcal{S} = 1$



(a) Estimated risk neutral density



(b) Cumulative distribution function

FIGURE 3.2: Estimated risk neutral density and cumulative distribution function, $\mathcal{S} = 2$

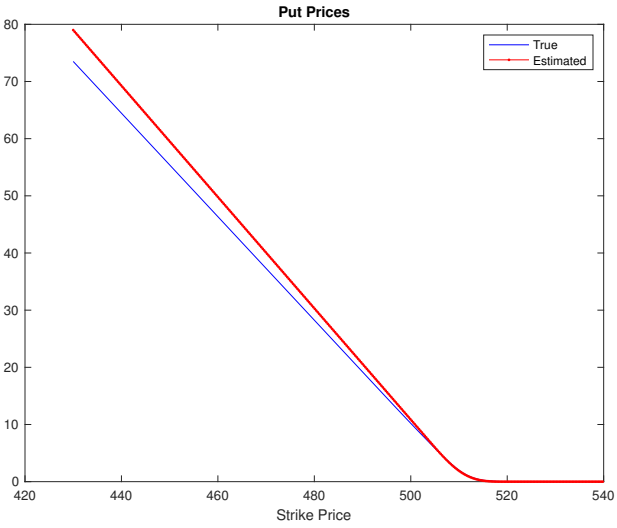
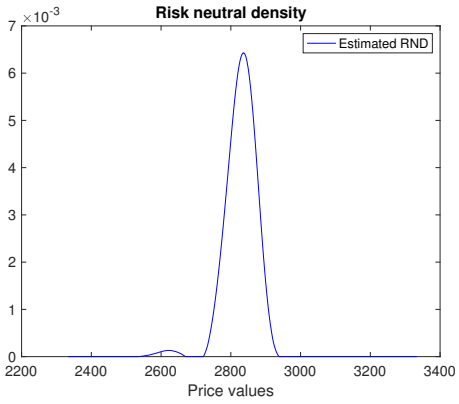
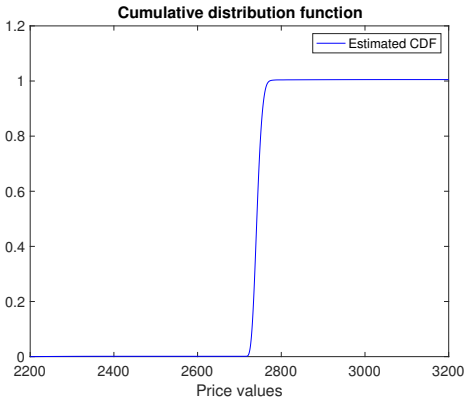


FIGURE 3.3: Comparison of the true and estimated put prices

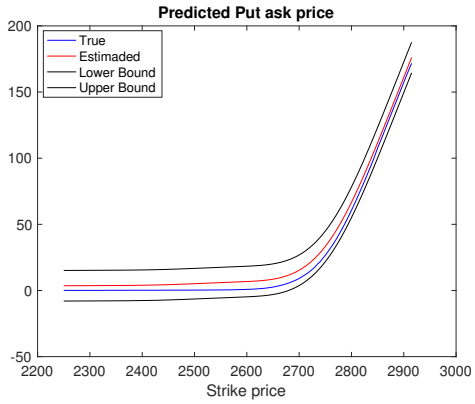


(a) Estimated risk neutral density

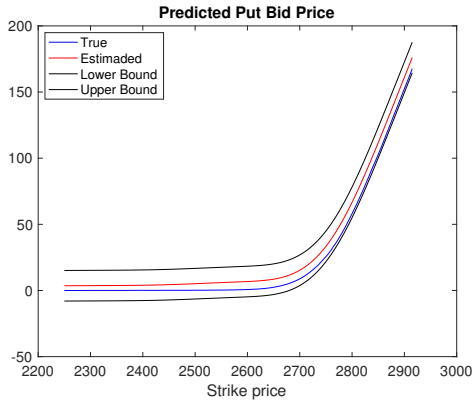


(b) Cumulative distribution function

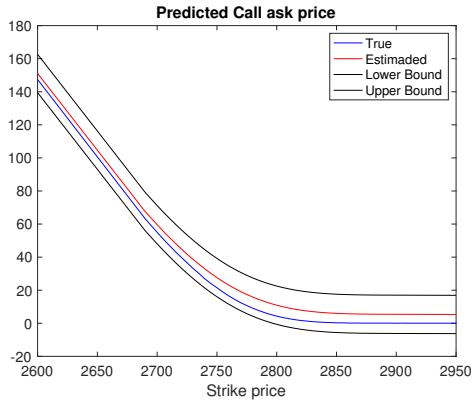
FIGURE 3.4: Estimated RND and CDF functions on S&P 500 options



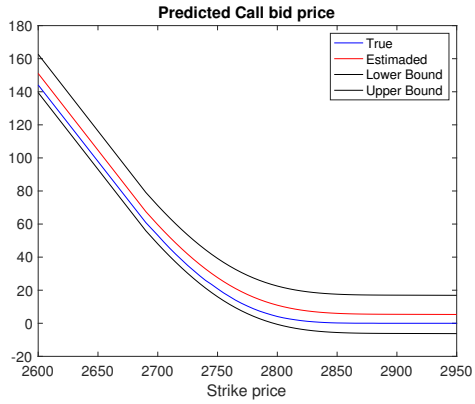
(a) Put ask prices



(b) Put bid prices



(c) Call ask prices



(d) Call bid prices

FIGURE 3.5: SPX predicted call and put prices with 95% confidence bounds

TABLE 3.2: Comparison of the estimation methods, model with $\mathcal{S} = 1$

Criterion	FLF	PosConv
$MSPE$	27.26	42.23
$RISE$	0.264	0.035

TABLE 3.3: Comparison of the estimation methods, model with $\mathcal{S} = 2$

Criterion	FLF	PosConv
$MSPE$	27.14	78.32
$RISE$	0.193	0.037

Conclusion Générale

L'objectif de cette thèse était d'exploiter le cadre d'analyse de données fonctionnelles afin de traiter des questions d'estimation et d'analyses d'inférencielle et de présenter des cas d'application à des sujets sur les marchés financiers.

Au premier chapitre, nous avons effectué une comparaison théorique des techniques d'analyse des composantes principales fonctionnelles (FPCA) et des moindres carrés partiels fonctionnels (FPLS) dans le contexte d'un modele de regression fonctionnelle avec comme variable réponse un scalaire. Nous avons déterminé la vitesse de convergence de l'erreur quadratique moyen de prédiction conditionnelle (MSPE) pour les deux méthodes d'estimation. Nous avons montré que le biais de régularisation de la méthode FPLS est plus petit que celui déterminé via la méthode FPCA, tandis que son erreur d'estimation a tendance à être plus grande que celle de FPCA. De plus, nous avons montré que la méthode FPLS performe mieux que la méthode FPCA en termes de prédiction, ceci avec moins de composantes.

Dans le deuxième chapitre j'ai considéré un modèle autorégressif fonctionnel (FAR) afin d'effectuer la prévision de la courbe de rendement du S&P 500 à la prochaine journée. Contrairement aux modèles AR(1) standard où chaque observation est un scalaire, dans cette étude, chaque courbe de rendement journalière est une collection de 390 points et est considérée comme une observation. L'estimation du paramètre du modèle conduit a un problème d'inversibilité. Ainsi, pour résoudre ce problème, j'ai réalisé une analyse comparative de quatre techniques de Big Data, dont la méthode de Tikhonov fonctionnelle (FT), la technique de Landweber-Fridman fonctionnelle (FLF), la coupure spectrale fonctionnelle (FSC) et les moindres carrés partiels fonctionnels (FPLS). J'ai également déterminé la vitesse de convergence de ces , la distribution asymptotique du paramétre estimé pour les differentes méthodes d'estimation. De plus, j'ai developpé une stratégie de test statistique pour sélectionner le nombre de retard du modèle. Les simulations et les données réelles montrent que les méthodes FPLS et FPCA performent mieux les autres en termes d'estimation tandis que toutes les quatres méthodes affichent des performances similaires en termes de prédiction. Aussi la méthode FPLS tend à présenter un R^2_{oos} hors échantillon assez remarquable, d'une valeur de 3.9% montrant ainsi le démarquage de cette méthode en terme de prediction.

Le troisième chapitre a été consacré à l'estimation de la densité de neutralité au risque (RND) dans le contexte de la tarification des options sur les marchés financiers,

à l'aide d'un modèle fonctionnel. Cette approche repose sur l'idée que la RND peut être considéré comme une fonction qui prend des valeurs sur une grille fine et que l'on peut avoir un nombre infini de possibilités à l'échéance. L'avantage de cette approche est qu'elle exploite la théorie d'absence d'arbitrage et qu'il est possible d'éviter toute sorte de paramétrisation. L'estimation de cette densité se traduit par un problème d'inversibilité et la technique fonctionnelle de Landweber-Fridman (FLF) a été utilisée pour controler la stabilite du parametre estime. Après avoir developpé les résultats asymptotiques du RND estimé, j'ai presenté une analyse empirique pour montrer les performances de la méthode.

En fin de compte, les modèles d'analyse des données fonctionnelles attirent une attention particulière dans le contexte des etudes économiques, mais l'analyses modèles théorique y relatif reste encore un champ à explorer. Les analyses faites dans cette thèse permettent d'ouvrir un champ d'exploration et de fixer un cadre d'analyse des propriétes des estimateurs dans l'utilisation des modèles de prédiction connexes avec des applications en économie et finance.

Bibliographie

Aguilera, Ana M, Manuel Escabias, Cristian Preda, and Gilbert Saporta, "Using basis expansions for estimating functional PLS regression : applications with chemometric data," *Chemometrics and Intelligent Laboratory Systems*, 2010, *104* (2), 289–305.

Aıt-Sahalia, Yacine and Andrew W Lo, "Nonparametric risk management and implied risk aversion," *Journal of econometrics*, 2000, *94* (1-2), 9–51.

_____ **and Jefferson Duarte**, "Nonparametric option pricing under shape restrictions," *Journal of Econometrics*, 2003, *116* (1-2), 9–47.

Ait-Sahalia, Yacine, Mustafa Karaman, and Loriano Mancini, "The term structure of variance swaps and risk premia," *Swiss Finance Institute Research Paper*, 2018, (18-37).

Aue, Alexander, Diogo Dubart Norinho, and Siegfried Hörmann, "On the prediction of stationary functional time series," *Journal of the American Statistical Association*, 2015, *110* (509), 378–392.

Bahra, Bhupinder, "Implied risk-neutral probability density functions from option prices : theory and application," 1997.

Bai, Jushan and Serena Ng, "Determining the number of factors in approximate factor models," *Econometrica*, 2002, *70* (1), 191–221.

Barratt, Shane, Jonathan Tuck, and Stephen Boyd, "Convex Optimization Over Risk-Neutral Probabilities," *arXiv preprint arXiv :2003.02878*, 2020.

Benatia, David, Marine Carrasco, and Jean-Pierre Florens, "Functional linear regression with functional response," *Journal of Econometrics*, 2017, *201* (2), 269–291.

Berg, Christian and Ryszard Szwarc, "The smallest eigenvalue of Hankel matrices," *Constructive Approximation*, 2011, *34* (1), 107–133.

Besse, Philippe C, Hervé Cardot, and David B Stephenson, "Autoregressive forecasting of some functional climatic variations," *Scandinavian Journal of Statistics*, 2000, *27* (4), 673–687.

Black, Fischer and Myron Scholes, "The pricing of options and corporate liabilities," *Journal of political economy*, 1973, *81* (3), 637–654.

Blanchard, Gilles and Nicole Krämer, "Optimal learning rates for kernel conjugate gradient regression," in "Advances in Neural Information Processing Systems" 2010, pp. 226–234.

_____ **and** _____, "Convergence rates of kernel conjugate gradient for random design regression," *Analysis and Applications*, 2016, *14* (06), 763–794.

Blazere, Mélanie, Fabrice Gamboa, and Jean-Michel Loubes, "PLS : a new statistical insight through the prism of orthogonal polynomials," *arXiv preprint arXiv :1405.5900*, 2014.

Blazère, Mélanie, Fabrice Gamboa, and Jean-Michel Loubes, "A Unified Framework to Study the Properties of the PLS Vector of Regression Coefficients," in "International Conference on Partial Least Squares and Related Methods" Springer 2014, pp. 227–237.

Bliss, Robert R and Nikolaos Panigirtzoglou, "Option-implied risk aversion estimates," *The journal of finance*, 2004, *59* (1), 407–446.

Bogousslavsky, Vincent, "Infrequent rebalancing, return autocorrelation, and seasonality," *The Journal of Finance*, 2016, *71* (6), 2967–3006.

Bondarenko, Oleg, "Estimation of risk-neutral densities using positive convolution approximation," *Journal of Econometrics*, 2003, *116* (1-2), 85–112.

Bosq, D, "Linear Processes in Function Spaces : Theory and Applications, volume 149 of Lecture Notes in Statistics," 2000.

Braak, Cajo JF Ter and Sijmen de Jong, "The objective function of partial least squares regression," *Journal of Chemometrics : A Journal of the Chemometrics Society*, 1998, *12* (1), 41–54.

Bro, Rasmus and Lars Elden, "PLS works," *Journal of Chemometrics : A Journal of the Chemometrics Society*, 2009, *23* (2), 69–71.

Cai, T Tony, Ming Yuan et al., "Adaptive covariance matrix estimation through block thresholding," *The Annals of Statistics*, 2012, *40* (4), 2014–2042.

_____ **, Peter Hall et al.**, "Prediction in functional linear regression," *The Annals of Statistics*, 2006, *34* (5), 2159–2179.

Cardot, Hervé, André Mas, and Pascal Sarda, "CLT in functional linear regression models," *Probability Theory and Related Fields*, 2007, *138* (3-4), 325–361.

_____, **Frédéric Ferraty, and Pascal Sarda**, "Functional linear model," *Statistics & Probability Letters*, 1999, *45* (1), 11–22.

_____, _____, **and** _____, "Spline estimators for the functional linear model," *Statistica Sinica*, 2003, pp. 571–591.

Carrasco, Marine and Barbara Rossi, "In-sample inference and forecasting in misspecified factor models," *Journal of Business & Economic Statistics*, 2016, *34* (3), 313–338.

_____ **and Idriss Tsafack**, "Theoretical comparison of functional principal component and functional partial least squares," *Working paper*, 2020.

_____, **Jean-Pierre Florens, and Eric Renault**, "Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization," *Handbook of econometrics*, 2007, *6*, 5633–5751.

Cerovecki, Clément, Christian Francq, Siegfried Hörmann, and Jean-Michel Zakoian, "Functional GARCH models : The quasi-likelihood approach and its applications," *Journal of econometrics*, 2019, *209* (2), 353–375.

Chan, Ernie, *Algorithmic trading : winning strategies and their rationale*, Vol. 625, John Wiley & Sons, 2013.

Chu, Xiaojun, Zherong Gu, and Haigang Zhou, "Intraday momentum and reversal in Chinese stock market," *Finance Research Letters*, 2019, *30*, 83–88.

Chun, Hyonho and Sündüz Keleş, "Sparse partial least squares regression for simultaneous dimension reduction and variable selection," *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 2010, *72* (1), 3–25.

Cox, John C and Stephen A Ross, "The valuation of options for alternative stochastic processes," *Journal of financial economics*, 1976, *3* (1-2), 145–166.

Crambes, Christophe, André Mas et al., "Asymptotics of prediction in functional linear regression with functional outputs," *Bernoulli*, 2013, *19* (5B), 2627–2651.

Dauxois, Jacques, Alain Pousse, and Yves Romain, "Asymptotic theory for the principal component analysis of a vector random function : some applications to statistical inference," *Journal of multivariate analysis*, 1982, *12* (1), 136–154.

Delaigle, Aurore and Peter Hall, "Methodology and theory for partial least squares applied to functional data," *The Annals of Statistics*, 2012, *40* (1), 322–352.

Didericksen, Devin, Piotr Kokoszka, and Xi Zhang, "Empirical properties of forecasts with the functional autoregressive model," *Computational statistics*, 2012, *27* (2), 285–298.

Diebold, Francis X and Robert S Mariano, "Comparing predictive accuracy," *Journal of Business & economic statistics*, 1995, *13* (1), 253–263.

Driessen, Joost, Pascal J Maenhout, and Grigory Vilkov, "The price of correlation risk : Evidence from equity options," *The Journal of Finance*, 2009, *64* (3), 1377–1406.

Engl, Heinz Werner, Martin Hanke, and Andreas Neubauer, *Regularization of inverse problems*, Vol. 375, Springer Science & Business Media, 1996.

Febrero-Bande, Manuel, Pedro Galeano, and Wenceslao González-Manteiga, "Functional Principal Component Regression and Functional Partial Least-squares Regression : An Overview and a Comparative Study," *International Statistical Review*, 2017, *85* (1), 61–83.

Fengler, Matthias R, "Arbitrage-free smoothing of the implied volatility surface," *Quantitative Finance*, 2009, *9* (4), 417–428.

Ferraty, Frédéric and Philippe Vieu, *Nonparametric functional data analysis : theory and practice*, Springer Science & Business Media, 2006.

Figlewski, Stephen, "Estimating the implied risk neutral density," 2008.

Gabrys, Robertas, Lajos Horváth, and Piotr Kokoszka, "Tests for error correlation in the functional linear model," *Journal of the American Statistical Association*, 2010, *105* (491), 1113–1125.

Gao, Lei, Yufeng Han, Sophia Zhengzi Li, and Guofu Zhou, "Market intraday momentum," *Journal of Financial Economics*, 2018, *129* (2), 394–414.

Garcia, René and Ramazan Gençay, "Pricing and hedging derivative securities with neural networks and a homogeneity hint," *Journal of Econometrics*, 2000, *94* (1-2), 93–115.

Glad, Ingrid K, Nils Lid Hjort, and Nikolai G Ushakov, "Correction of density estimators that are not densities," *Scandinavian Journal of Statistics*, 2003, *30* (2), 415–427.

Gourieroux, Christian and Joann Jasiak, *Financial econometrics : Problems, models, and methods*, Princeton University Press, 2001.

Groen, Jan JJ and George Kapetanios, "Revisiting useful approaches to data-rich macroeconomic forecasting," 2009.

Hall, Peter and Joel L Horowitz, "Methodology and convergence rates for functional linear regression," *The Annals of Statistics*, 2007, *35* (1), 70–91.

_____ **and Mohammad Hosseini-Nasab**, "On properties of functional principal components analysis," *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 2006, *68* (1), 109–126.

Hanke, Martin, *Conjugate gradient type methods for linear ill-posed problems*, Pitman Research Notes in Mathematics Series, 1995.

_____, *Conjugate gradient type methods for ill-posed problems*, Routledge, 2017.

Hays, Spencer, Haipeng Shen, and Jianhua Z Huang, "Functional dynamic factor models with application to yield curve forecasting," *The Annals of Applied Statistics*, 2012, *6* (3), 870–894.

Helland, Inge S, "On the structure of partial least squares regression," *Communications in statistics-Simulation and Computation*, 1988, *17* (2), 581–607.

_____, "Partial least squares regression and statistical models," *Scandinavian Journal of Statistics*, 1990, pp. 97–114.

_____ **and Trygve Almøy**, "Comparison of prediction methods when only a few components are relevant," *Journal of the American Statistical Association*, 1994, *89* (426), 583–591.

Heston, Steven L, Robert A Korajczyk, and Ronnie Sadka, "Intraday patterns in the cross-section of stock returns," *The Journal of Finance*, 2010, *65* (4), 1369–1407.

Hörmann, Siegfried and Piotr Kokoszka, "Functional time series," in "Handbook of statistics," Vol. 30, Elsevier, 2012, pp. 157–186.

Horváth, Lajos, Marie Hušková, and Piotr Kokoszka, "Testing the stability of the functional autoregressive process," *Journal of Multivariate Analysis*, 2010, *101* (2), 352–367.

Höskuldsson, Agnar, "PLS regression methods," *Journal of chemometrics*, 1988, *2* (3), 211–228.

Hyndman, Rob J and Han Lin Shang, "Forecasting functional time series," *Journal of the Korean Statistical Society*, 2009, *38* (3), 199–211.

Imaizumi, Masaaki and Kengo Kato, "PCA-based estimation for functional linear regression with functional responses," *Journal of multivariate analysis*, 2018, *163*, 15–36.

Jackwerth, Jens Carsten and Mark Rubinstein, "Recovering probability distributions from option prices," *The Journal of Finance*, 1996, *51* (5), 1611–1631.

Jarrow, Robert and Andrew Rudd, "Approximate option valuation for arbitrary stochastic processes," *Journal of financial Economics*, 1982, *10* (3), 347–369.

Jolliffe, Ian T, "A note on the use of principal components in regression," *Applied Statistics*, 1982, pp. 300–303.

Jong, Sijmen De, "SIMPLS : an alternative approach to partial least squares regression," *Chemometrics and intelligent laboratory systems*, 1993, *18* (3), 251–263.

Jr, Robert E Lucas, "Asset prices in an exchange economy," *Econometrica : Journal of the Econometric Society*, 1978, pp. 1429–1445.

Kargin, Vladislav and Alexei Onatski, "Curve forecasting by functional autoregression," *Journal of Multivariate Analysis*, 2008, *99* (10), 2508–2526.

Kelly, Bryan and Seth Pruitt, "The three-pass regression filter : A new approach to forecasting using many predictors," *Journal of Econometrics*, 2015, *186* (2), 294–316.

Kokoszka, P and X Zhang, "Improved estimation of the kernel of the functional autoregressive process," Technical Report, Technical Report. Utah State University 2010.

Kokoszka, Piotr and Gabriel Young, "KPSS test for functional time series," *Statistics*, 2016, *50* (5), 957–973.

_____ **and Matthew Reimherr**, "Determining the order of the functional autoregressive model," *Journal of Time Series Analysis*, 2013, *34* (1), 116–129.

_____ **and Xi Zhang**, "Functional prediction of intraday cumulative returns," *Statistical Modelling*, 2012, *12* (4), 377–398.

Lingjaerde, Ole C and Nils Christophersen, "Shrinkage structure of partial least squares," *Scandinavian Journal of Statistics*, 2000, *27* (3), 459–473.

Liu, Xialu, Han Xiao, and Rong Chen, "Convolutional autoregressive models for functional time series," *Journal of econometrics*, 2016, *194* (2), 263–282.

Luedicke, Joerg and Alberto Bernacchia, "Self-consistent density estimation," *The Stata Journal*, 2014, *14* (2), 237–258.

Naes, Tormod and Harald Martens, "Comparison of prediction methods for multicollinear data," *Communications in Statistics-Simulation and Computation*, 1985, *14* (3), 545–576.

Nie, Yunlong, Liangliang Wang, Baisen Liu, and Jiguo Cao, "Supervised functional principal component analysis," *Statistics and Computing*, 2018, *28* (3), 713–723.

Panigirtzoglou, Nikolaos and George Skiadopoulos, "A new approach to modeling the dynamics of implied distributions : Theory and evidence from the S&P 500 options," *Journal of Banking & Finance*, 2004, *28* (7), 1499–1520.

Park, Joon Y and Junhui Qian, "Functional regression of continuous state distributions," *Journal of Econometrics*, 2012, *167* (2), 397–412.

Phatak, Aloke and Frank de Hoog, "Exploiting the connection between PLS, Lanczos methods and conjugate gradients : alternative proofs of some properties of PLS," *Journal of Chemometrics : A Journal of the Chemometrics Society*, 2002, *16* (7), 361–367.

Plato, Robert, *A Conjugate Gradient Type Method for Linear Ill Posed Problems with Operator Pertubations*, Citeseer, 1998.

Preda, Cristian and Gilbert Saporta, "Clusterwise PLS regression on a stochastic process," *Computational Statistics & Data Analysis*, 2005, *49* (1), 99–108.

Ramsay, James O and Bernard W Silverman, *Applied functional data analysis : methods and case studies*, Springer, 2007.

Reiss, Philip T and R Todd Ogden, "Functional principal component regression and functional partial least squares," *Journal of the American Statistical Association*, 2007, *102* (479), 984–996.

Rice, Gregory, Tony Wirjanto, and Yuqian Zhao, "Tests for conditional heteroscedasticity with functional data and goodness-of-fit tests for FGARCH models," 2019.

Rosenberg, Joshua V, "Pricing multivariate contingent claims using estimated risk–neutral density functions," *Journal of International Money and Finance*, 1998, *17* (2), 229–247.

Rosipal, Roman and Nicole Krämer, "Overview and recent advances in partial least squares," in "International Statistical and Optimization Perspectives Workshop" Subspace, Latent Structure and Feature Selection"" Springer 2005, pp. 34–51.

Rubinstein, Mark, "The valuation of uncertain income streams and the pricing of options," *The Bell Journal of Economics*, 1976, pp. 407–425.

Sancetta, Alessio, "Intraday End-of-Day Volume Prediction," *Journal of Financial Econometrics*, 2019.

Shang, Han Lin, "Forecasting intraday S&P 500 index returns : A functional time series approach," *Journal of Forecasting*, 2017, *36* (7), 741–755.

_____, **Yang Yang, and Fearghal Kearney**, "Intraday forecasts of a volatility index : functional time series methods with dynamic updating," *Annals of Operations Research*, 2019, *282* (1-2), 331–354.

Shimko, David, Naohiko Tejima, and Donald R Van Deventer, "The pricing of risky debt when interest rates are stochastic," *Journal of Fixed Income*, 1993, *3* (2), 58–65.

Souissi, Nessim, "The Implied Risk Neutral Density Dynamics : Evidence from the S&P TSX 60 Index," *Journal of Applied Mathematics*, 2017, *2017.*

Stock, James H and Mark W Watson, "Business cycle fluctuations in US macroeconomic time series," Technical Report, National Bureau of Economic Research 1998.

Stoica, Petre and Torsten Söderström, "Partial least squares : a first-order analysis," *Scandinavian Journal of Statistics*, 1998, *25* (1), 17–24.

Tsafack, Idriss, "Intraday Stock Market Forecasting via Functional Time Series," *working paper*, 2020, *3* (2), 58–65.

Vogt, Erik, "Option-implied term structures," *Available at SSRN 2541954*, 2014.

Wold, Svante, Arnold Ruhe, Herman Wold, and WJ Dunn III, "The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses," *SIAM Journal on Scientific and Statistical Computing*, 1984, *5* (3), 735–743.

Yatchew, Adonis and Wolfgang Härdle, "Nonparametric state price density estimation using constrained least squares and the bootstrap," *Journal of Econometrics*, 2006, *133* (2), 579–599.

Zhang, Yaojie, Feng Ma, and Bo Zhu, "Intraday momentum and stock return predictability : Evidence from China," *Economic Modelling*, 2019, *76*, 319–329.

Zhou, Zhiyang, "Functional continuum regression," *Journal of Multivariate Analysis*, 2019, *173*, 328–346.