Intraday Stock Market Forecasting via Functional Time Series

Idriss Tsafack ∗

1Department of Economics, University of Montreal

Job Market Paper

Latest version here.

December 24, 2019

Abstract

This paper considers the intraday S&P500 price values at the 1-minute frequency in constructing a collection of return curves sequentially observed each day and an autoregressive model is deployed to forecast the one-day-ahead market return curve using the functional data analysis (FDA). In contrast to the standard AR(1) model where each observation is a scalar, in this paper, each daily return curve is considered as one observation. This approach is practically important because it exploits the potential of available high-frequency data to improve the forecast. Moreover, investors can use forecasts to optimize their momentum, reversal, or portfolio rebalancing strategy. The estimation of this model leads to an ill-posed inverse problem and I conduct a comparative analysis of the four-dimension reduction methods including the Functional Tikhonov method (FT), the Functional Landweber-Fridman technique (FLF), the Functional spectral-cut off (FSC), and the Functional Partial Least Squares (FPLS). The convergence rate, asymptotic distribution, and a test-based strategy to select the optimal lag number are provided. The Monte Carlo simulation results show the performance of the different approaches. Moreover, the empirical application shows that the FPLS method exhibits a remarkable out-of-sample $R^2$ of 8% due to the smartness of the procedure.

Keywords: Forecasting, Functional Data Analysis, Principal Component Analysis, Regularization Method, return predictability, Big data, Tikhonov method, Functional Partial Least Squares.

JEL Codes: C38, C51, C53, C55, C58.

∗Economic Departement, University of Montreal, 3150 rue Jean-Brillant, Montreal, (QC) H3T 1N8, Canada (idriss.tsafack.teufack@umontreal.ca), website: https://www.idrisstsafack.com
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 and daily. Alternatively, lower frequency data is exploited in an autoregressive model or 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 & 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 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 AR(1) on functions is used for estimation and prediction. 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. Each minute data is therefore considered as a realization of a collection of curves observed sequentially on each day segment.

\[
X_{n+1}(t) = \int_0^1 \psi(t,s)X_n(s)ds + \varepsilon_{n+1}(t) \quad n \in \mathbb{Z}
\]

where $X_n(s)$ is the curve of the cumulative intraday return at the minute $s$ of day $n$, $\psi(t,s)$ is the autoregressive operator that is an integral 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.

$^1$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.
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 $R^{2}_{oos}$
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. Therefore, I am interested
in forecasting the next day whole return curve instead of forecasting the next minute return.

One of the most important challenges of this model is to estimate the autoregressive operator $^{2}$ Indeed,
with the high dimensionality, 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 & Zhang (2012), Crambes et al. (2013), Aue et al.
(2015), Shang (2017), Imaizumi & 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. Furthermore, using a PCA-based approach makes it
difficult to identify the estimated kernel operator if the purpose is the estimation. There is also a literature
suggesting nonparametric methods in order to analyze the functional data models (see Besse et al. (2000),
Ramsay & Silverman (2007), Ferraty & Vieu (2006), and Hörmann & Kokoszka (2012)), but using this
approach assumes a certain continuity constraint of the functional data that is not empirically derived and
the basis used for estimations is not data related.

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. These theoretical results are discussed. The
asymptotic normality results would be useful to test the significance of the estimated operator and the
predictability of the returns. Additionally, a test based strategy to identify the optimal lag parameter on a
general context functional autoregressive of order $p$ ($FAR(p)$) is proposed. The advantage of the proposed

$^{2}$The autoregressive operator is similar to the slope parameter in the context of a simple AR(1) model.
approach is that the procedure is not necessarily PCA-based as proposed in prior papers (see Kokoszka & Reimherr (2013), Aue et al. (2015), and Liu et al. (2016)). Moreover, the approach can be applied when using any regularization method that is linear in terms of the response function.

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), the Mean Average Distance (AD), and the Ratio Average Distance (RAD), 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 simulations show that the FLF and the FT tend to outperform the others in terms of estimation of the autoregressive operator when the sample size is large, while, in terms of prediction, the FT and FPLS tend to present the best predictive performance for most cases. Sometimes, the FPLS tend to outperform the others in terms of MSPE and $R^2_{oos}$ and this may come from the supervision of the method to smartly estimate the factors. The performance of FLF and FT in terms of estimation can be attributed to the fact that these methods are not limited by the eigenvalue configuration of the covariance operator of the predictor function.

An overview of the empirical 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. The FT and FLF display almost similar estimation results of the autoregressive operator, while the FPLS tend to capture well the out-of-sample prediction. This means that if investors usually rely on the first half of the previous trading day to make their analysis and make decisions for the next day, they can time their portfolio exposure from the beginning to the end of the next day. Otherwise, if they only rely on the second half of the previous trading day to make next day decisions, they should use a momentum strategy only on the first half and a reversal strategy on the second half of the next day. Moreover, the FPLS approach tends to produce a remarkable $R^2_{oos}$ of 8% in the periods 09:30 AM - 10:30 AM and 02:30 PM - 04:00 PM within a day, that is almost 4 times the one obtained by Gao et al. (2018) and twice the one obtained by Zhang et al. (2019) when using a simple time series momentum model. The FPCA, FSC, and FT methods tend to reach a predictive $R^2_{oos}$ of 5% and 6%, respectively in the period 09:30 AM - 10:30 AM and nearly 2.5% in the period 02:30 PM - 04:00 PM of a trading session for the FT approach. Surprisingly, the FPCA is not able to capture the momentum in the second half period of a trading session. According to the FLF approach, the $R^2_{oos}$ is around 2.5% at the beginning of the trading session, while in the second half of the

\^That is near twice the one obtained by Gao et al. (2018)
day it is approximately 1.2%.

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 & Zhang (2012) and Shang (2017)). Indeed, Kokoszka & 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 & Onatski (2008) proposed a predictive factor method for predicting the next day curve. 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 & Kato (2018), Crambes et al. (2013), and Shang (2017).

The rest of the paper is organized as follows. Section 2 is dedicated to presenting the related literature. Section 3 details the functional econometric model. In section 4, I explain how to estimate the model using the four aforementioned methods. Section 5 analyzes the convergence rate of the estimated autoregressive operator. Section 6 examines the asymptotic normality of the parameter. Section 7 address the selection of the optimal number of lags for a generalized functional autoregressive model. Section 8 discusses on data driven approach to select the tuning parameter. Section 9 presents the comparison of the four methods based on Monte Carlo simulations. Section 10 develops the real data application. Finally, section 11 concludes. The proofs of the main theoretical results are presented in the appendix.

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 & Silverman (2007), Kokoszka & Zhang (2012), and Ferraty & 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 & 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 & Kokoszka (2012) (among others) considered a parametric model and used the Yule-Walker technique for estimation. On the same line, Kargin & 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 & 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 & 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 & 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 & 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 & 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 & 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 & 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 4 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.

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, $j \in \{1, \ldots, 390\}$. The return is defined as

$$R_i(t_j) = 100 \times \left[ \ln(P_i(t_j)) - \ln(P_i(t_1)) \right] \quad \text{with} \quad j \in 1, \ldots, 390. \tag{2}$$

and the CIDRs are defined by

$$X(t) = R_i(t_j) \quad \text{with} \quad t_j \in \left( \frac{(j - 1)}{390}, \frac{j}{390} \right]. \tag{3}$$

Figure 1 displays the constructed intraday cumulative returns of the S&P500 based on the raw data for the period 2013 - 2017.

4This is similar to the slow diffusion, analysis, and acceptance of new information identified by Chan (2013)
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 \(<f,g>_{\mathbb{H}} = \int_0^1 f(t)g(t)dt\) and the norm \(||f||_{\mathbb{H}} = \left(\int_0^1 f^2(t)dt\right)^{1/2}\) such that \(E(||X_i||^4) < \infty\).

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:

\[
X_{n+1}(t) = \int_0^1 \psi(t,s)X_n(s)ds + \varepsilon_{n+1}(t) \quad n \in \mathbb{Z} \tag{4}
\]

where for each day \(n\), \(X_n\) is a random curve of \(\mathbb{H}\),

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

is a bounded linear operator and \(\varepsilon = (\varepsilon_n, n \in \mathbb{Z})\) is a \(\mathbb{H}\)-valued stationary and ergodic martingale difference such that \(E(\varepsilon_n|X_{n-1}) = 0\) 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 represents how the predictor and the predicted functions are displayed. According to what is observed, it can be deduced that if the functional 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 & Young (2016) who developed a KPSS unit-root test for functional time series.

![Predictor functions series (day n)](image)

![Response Functions series (day n+1)](image)

**Figure 2:** Functional Predictor and Functional response

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

$$||\Psi||_{\mathcal{L}} = \sup \{||\Psi(f)|| : ||f|| \leq 1\}$$

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

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

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

In the next section, various regularization techniques are presented.

### 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, $\Psi$ is unknown and should be estimated consistently by an estimator $\hat{\Psi}$. This section presents four different estimation strategies of the autoregressive operator $\Psi$. Multiplying equation (1) by $X_n$ and taking the
expectation on both sides lead to the following equation:

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

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 $(v_j, \lambda_j)_{j \geq 1}$ are 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 $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$.

Let us define the cross-covariance operator by

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

Then, it is easy to see that

$$D(f) = K\Psi^*(f). \quad (7)$$

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 $\hat{\lambda}_j \geq \hat{\lambda}_j \geq \ldots \geq 0$ and $(\hat{v}_j)_{j \geq 1}$ form an orthonormal basis of $\mathbb{H}$.

Given equation (5), one would like to brutally 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 infinitely-dimensional space. Thus, $K^{-1}$ is an unbounded operator, which would lead to an unstable estimator. In the inverse problem literature, equation (5) 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 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).

### 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

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

Then, the estimated autoregressive operator is given by

$$\Psi^\alpha(f) = K^{-1}_\alpha D(f) = \sum_{\lambda_j > \alpha} \frac{1}{\lambda_j} < D(f), v_j > v_j = \sum_{\lambda_j > \alpha} \frac{1}{\lambda_j} \mathbb{E} [< X_{n+1}, f > < X_n, v_j >] v_j.$$

and its empirical counterpart is
\[ \hat{\Psi}^*_n(f) = \frac{1}{N-1} \sum_{j=1}^{N-1} \sum_{n=1}^{N-1} \hat{Q}_{\alpha,j} \frac{\hat{\lambda}_j}{\lambda_j} \langle X_{n+1}, f \rangle < X_n, \hat{v}_j \rangle \hat{v}_j, \text{ for each } f \in \mathbb{H}. \]  
(8)

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

Similar to FSC, the FPCA method consists of 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 greater eigenvalues. Thus, if \( m \) eigenfunctions are selected for the estimation, the FPCA estimator is given by

\[
\Psi_m^*(f) = K_m^{-1} D(f) = \sum_{j=1}^{m} \frac{1}{\lambda_j} < D(f), v_j \rangle v_j = \sum_{j=1}^{m} \frac{1}{\lambda_j} \mathbb{E}[< X_{n+1}, f \rangle < X_n, v_j \rangle] v_j.
\]

and its empirical version is given by

\[
\hat{\Psi}_m^*(f) = \frac{1}{N-1} \sum_{j=1}^{N-1} \sum_{n=1}^{N-1} \hat{Q}_{m,j} \frac{\hat{\lambda}_j}{\lambda_j} \langle X_{n+1}, f \rangle < X_n, \hat{v}_j \rangle \hat{v}_j, \text{ for each } f \in \mathbb{H}
\]  
(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 & Kato (2018) and it consists of 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 & Kato (2018) is not considered in this paper.

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 \( \alpha \) be a positive tuning parameter. Then, the estimated autoregressive operator is given by
\[ \Psi^*_\alpha(f) = \left( \alpha I + K \right)^{-1} D(f), \quad \text{for each } f \in \mathbb{H}, \]

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

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

The empirical version is then given by

\[
\hat{\Psi}^*_\alpha(f) = \left( \alpha I + \hat{K} \right)^{-1} \hat{D}(f)
\]

\[
= \frac{1}{N-1} \sum_{j=1}^{N-1} \sum_{n=1}^{N-1} \hat{Q}(\alpha, \hat{\lambda}_j) < X_{n+1}, f > < X_n, \hat{v}_j > \hat{v}_j, \quad \text{for each } f \in \mathbb{H}.
\]

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

### 4.3 Functional Landweber-Fridman (FLF)

The Landweber-Fridman method is basically an iterative method. 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

\[ \Psi^*_0(f) = dD(f), \quad \text{for each } f \in \mathbb{H}. \]

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

\[ \Psi^*_h(f) = (I - dK)(\Psi^*_{h-1}(f)) + dD(f), \quad \text{for each } f \in \mathbb{H}. \]

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

\[ \Psi^*_\alpha(f) = d \sum_{l=1}^{1/\alpha} (I - dK)^{l-1} D(f), \quad \text{for each } f \in \mathbb{H}. \]  \( (10) \)
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

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

(11)

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

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

(12)

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

### 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 & Kapetanios (2009), Kelly & Pruitt (2015), and Carrasco & Rossi (2016). In the Functional regression context with a scalar response, there are authors like Aguilera et al. (2010), Delaigle & 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.

$$
\max_{\phi_h, c_h \in L^2([0,1])} \text{cov}^2\left( \int_0^1 X_n(s)\phi_h(s)ds, \int_0^1 X_{n+1}(t)c_h(t)dt \right)
$$

subject to $||\phi_h|| = 1, ||c_h|| = 1$, and

$$
\int_0^1 \int_0^1 \phi_\ell(s)K(s,t)\phi_h(t)dtsdt = 0, \ell = 1, ..., h - 1
$$

(13)

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

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

$$
\hat{\psi}_m^* (s, t) = \sum_{l=1}^{m} \hat{\gamma}_{t,l} \hat{K}^{l-1} (\hat{D})(s, t)
= \sum_{l=1}^{m} \hat{\gamma}_{t,l} \int_{0}^{1} \hat{K}^{l-1}(s, u) \hat{D}(u, t) du
$$

(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 $\hat{R}_{t,j,l} = \hat{R}_{t,j,l}$. $\hat{\mu}_t = [\hat{\mu}_{t,1}, ..., \hat{\mu}_{t,m}]'$ is a vector of length $m$.

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

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

(17)

with

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

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

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

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

(18)

where the filter factor $Q(\delta, \hat{\lambda}_j)$ is such that
\( Q(\delta, \hat{\lambda}_j) = Q(m, \hat{\lambda}_j) = I(j \leq m) \) for FPCA method
\( Q(\alpha, \hat{\lambda}_j) = I(\hat{\lambda}_j \geq \alpha) \) for SC
\( Q(\alpha, \hat{\lambda}_j) = \frac{\hat{\lambda}_j}{\hat{\lambda}_j + \alpha} \) for FT
\( Q(\alpha, \hat{\lambda}_j) = \left( 1 - (1 - \hat{\lambda}_j)^{\frac{\alpha}{2}} \right) \) for FLF
\( Q(m, \hat{\lambda}_j) = \left( 1 - \prod_{l=1}^{m} (1 - \frac{\hat{\lambda}_j}{\hat{\theta}_l}) \right) \) for FPLS.

with \( \alpha > 0 \) and \( m < N \).

It can also be noticed that if \( \hat{\theta}_l = \hat{\theta}_r = \hat{\theta}_0 \) for each \( l, r = 1, \ldots, m \), then FPLS is similar to FLF with \( d = \frac{1}{\hat{\theta}_0} \).

Given the estimated autoregressive operator \( \hat{\Psi}_\delta \), the best prediction of the one day ahead S&P 500 curve is given by
\[
\hat{X}_{n+1}(t) = \int_0^1 \hat{\psi}_\delta(s, t) X_n(s) \, ds \quad \text{for each } t \in [0, 1].
\]

The results of proposition 1 rely on an extension of the Alternative Partial Least Squares (APLS) approach proposed Delaigle & Hall (2012) combined with the results of proposition 1 by Carrasco & Tsafack (2019).

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 studies in the economic and financial field. For this purpose, the following assumptions are required:

Assumption 1 (A1) : \( \{X_1, \ldots, X_N\} \) 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) : \( \varepsilon_n \) is stationary and ergodic martingale difference sequence that takes values in \( \mathbb{H} \) with respect
to \{\varepsilon_{n-1}(t), \varepsilon_{n-2}(t), ..., X_{n-1}(t), X_{n-2}(t)\} 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) : There is a Hilbert-Schmidt operator \(R\) and a positive constant \(\beta\) such that

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

This source condition can also be written as

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

Assumption 4 (A4) : 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 1 ensures that the sequence \(\{X_n; n \in \mathbb{H}\}\) is a stationary process and admits a unique solution.

Assumption 2 imposes that the sequence of innovations \(\varepsilon_n\) 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 \(\varepsilon_n\) are martingale difference sequences, which is less restrictive than what is usually observed in preceding papers. Furthermore, since \(E(||X_n||^4) < +\infty\), the operator \(K\) is trace-class and thereby is Hilbert-Schmidt.

Assumption 3 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. This condition guarantees that \(\Psi^*\) belongs to the complement orthogonal to the null space of the operator \(K\). As \(\beta\) gets larger, \(\Psi^*(f)\) becomes smoother (see Carrasco et al. (2007) and Benatia et al. (2017)). This assumption is necessary to establish the rate of convergence of the bias term as a function of \(\alpha\). This assumption is different to the one considered by Imaizumi & 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 \(\lambda_j\). This paper does not use such assumptions. The source conditions considered in this paper are more general than the one by Imaizumi & Kato (2018) or Crambes et al. (2013). Under assumption 4, the eigenvalues \(\lambda_j\) are distinct. Under A2 and A4, the \(\lambda_j\) are consistently estimated by \(\hat{\lambda}_j\) (see lemma 2 by Kokoszka & Reimherr (2013)).

Let us denote the regularized version of \(\Psi^*\) by \(\Psi^*_\delta\) where \(\delta\) is \(\alpha\) for the FT, FLF, FSC and \(m\) for FPCA, FPLS methods. Then, for each function \(f \in \mathbb{H}\), \(\Psi^*_\delta\) can be written as
\[ \Psi_\delta^*(f) = \sum_{j=1}^{\infty} \frac{Q(\delta, \lambda_j)}{\lambda_j} < D(f), v_j > v_j \]

\[ = \sum_{j=1}^{\infty} \frac{Q(\delta, \lambda_j)}{\lambda_j} < K(\Psi^*)(f), v_j > v_j \]

\[ = \sum_{j=1}^{\infty} Q(\delta, \lambda_j) < \Psi^*(f), v_j > v_j. \]

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 \( \delta \) increases. \( \{ \hat{\Psi}_\delta^*(f) - \Psi_\delta^*(f) \} \) and the estimation error term which may increase as \( \delta \) 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

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

**Remarks 1.**

- Proposition 2 shows that as \( \alpha \) goes to zero the squared bias term decreases while the variance increases. Then \( \alpha \) should be optimally chosen i.e. such that bias is equal to the variance. Thus, at the optimality,

- If \( \alpha \sim N^{-1/(2+\beta)} \), then \( MSE \sim N^{2+\beta} \) for FLF and FSC.
- For \( \beta < 2 \), if \( \alpha \sim N^{-1/(2+\beta)} \), then \( MSE \sim N^{2+\beta} \) 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, if $m$ diverges much slower than $N$, such that $\frac{N\lambda m}{m} \to +\infty$, then*

$$
\mathbb{E}\left[\left\| \hat{\Psi}_{\delta}^* - \Psi^* \right\|^2_{HS} \mid \mathcal{F}_{N-1}\right] = \begin{cases} 
O_p\left(\frac{\lambda_{m+1}^\beta}{\frac{m}{\lambda_{m}N}}\right) + O_p\left(\frac{m}{\lambda_{m}N}\right) & \text{for the FPCA} \\
O_p\left(\frac{\lambda_{m+1}^\beta}{\frac{m}{\theta_{m}N}}\right) + O_p\left(\frac{m}{\theta_{m}N}\right) & \text{for the FPLS}
\end{cases}
$$

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

**Remarks 2.**

- Propositions 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 $\lambda_m$. On the other hand, the FPLS approach depends on the smallest root $\theta_m$ of the residual polynomial $Q_{m,j}$. $\theta_m$ is also called the smallest eigenvalue of a Hankel Matrix (see Delaigle & Hall (2012)).

- Since $\theta_m$ decreases at an exponential rate (see Berg & Szwarc (2011)), it is most of the time expected that the FPLS method presents a higher variance 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.

\footnote{For more discussions about $\theta_m$, see Carrasco & Tsafack (2019)}
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 \( \frac{N\lambda m}{m} \to +\infty \) is sufficient for both FPCA and FPLS since \( \theta_j < \lambda_j \) for \( j \leq m \) (see for instance Lingjaerde & Christophersen (2000) and ?).

The rate obtained for the FPCA is different to the one obtained by Imaizumi & 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 \( \lambda_j \) and on 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 assume that the eigenvalue belongs 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). They developed the convergence rate of the estimated operator in the context of the i.i.d model.

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 \( \delta \) and a testing procedure for the true number of lag \( p \) for a general model FAR(p) is developed. 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 model setting. This paper considered that the error term \( (\varepsilon_n)_{n=1,...,N} \) is a sequence of functional martingale difference in \( H \) 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 \( E[||X_i||^4] < \infty, E(||X_i||^2||\varepsilon_{i+1}||^2) < \infty \) and \( \delta \) is fixed, then

\[
\sqrt{n}(\hat{\Psi}_\delta^* - \Psi_\delta^*) \overset{d}{\to} N(0, \Omega_\delta) \quad \text{as } N \to \infty.
\]

(21)
where $\delta = m$ for FPCA and $\delta = \alpha$ for FT, FLF, and FSC.

\[
\Omega_\delta(s, t) = \int_0^1 \int_0^1 K_\delta^{-1}(s, \tau_1)G_\delta(\tau_1, \tau_2)K_\delta^{-1}(\tau_2, t)d\tau_1 d\tau_2.
\]

(22)

and

\[
G_\delta(s, t) = E[\varepsilon_{i+1}(s)\varepsilon_{i+1}(t)X_i(s)X_i(t)'].
\]

(23)

This result can be useful for testing the significance of the shape of the autoregressive operator in one hand and predict the confidence set for the predicted functions in another hand. The population version of this covariance operator is estimated by replacing the expectation with their empirical version. By using the conditional homoskedasticity assumption of the error term, $G_\delta$ can be replaced by

\[
G_\delta(s, t) = \sigma^2(s, t)E[X_i(s)X_i(t)']
\]

\[
= \sigma^2(s, t)k(s, t).
\]

If $s = t$, $G_\delta(s, t) = \sigma^2k(s, t)$ and 0 otherwise. $\sigma^2(s, t) = \sigma^2$ if $s = t$ and 0 otherwise. 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. This result 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.

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 & Reimherr (2013) are the first to propose a PCA-based multistage testing procedure. They stated 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 escalar observations. 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 the different estimation methods that is linear in terms of the response function.

Let us consider the FAR($p$) model.

$$X_{n+1} = \Psi_1(X_n) + \ldots + \Psi_p(X_{n-p+1}) + \varepsilon_{n+1} \quad (24)$$

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

$$Y_{n+1} = \Phi(Y_n) + U_{n+1} \quad (25)$$

where,

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

and

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

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 $\mathbb{H}_p = (L^2[0,1])^p$. $\Phi$ 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 $||x||_p = \sqrt{<x,x>^2_p}$ (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 given by

$$\begin{cases} 
H_0: \Psi_{r+1}(.) = \ldots = \Psi_p(.) \\
H_1: \text{not } H_0
\end{cases}$$

These hypotheses can be written as
\[
\begin{aligned}
H_0: & \ A(\Phi) = Z \\
H_1: & \ A(\Phi) \neq Z
\end{aligned}
\]
where \( A = [O_r, Id_{p-r}]' \), \( \Phi = [\Phi_1, \Phi_2]' \), \( \Phi_1 = [\Psi_1, ..., \Psi_r]' \) and \( \Phi_2 = [\Psi_{r+1}, ..., \Psi_p]' \) and \( Z = O_p \) is \( p \)-vector of zero operators in \((\mathbb{H} \times \mathbb{H})\).

### 7.1 Statistical test

Since proposition 3 holds, the Wald-type statistical test for this restriction hypothesis is given by

\[
W_N = N \sum_{h=r+1}^{p} ||\hat{G}_{\eta_h}^{-1/2} \hat{\Psi}_h \hat{K}_{\eta_h}||^2_{HS}
\]

where

\[
\hat{G}_{\eta_h}(s, t) = \frac{1}{N} \sum_{i=1}^{N-h} \varepsilon_{i+h+1}(s)\varepsilon_{i+h+1}(t)X_i(s)X_i(t)
\]

\[
\hat{K}_{\eta_h}(s, t) = \frac{1}{N} \sum_{i=1}^{N-h} X_i(s)X_i(t)
\]

and \( \hat{K}_{\eta_h}^{-1}\eta_h \) is the regularized version of the inverse of \( \hat{K}_{\eta_h} \). \( \eta_h \) is the \( h \)-th element of the \( p \)-vector of tuning parameters \( \eta = [\eta_1, ..., \eta_p]' \).

**Proposition 5**

Assume that A1 to A4 hold. If \( E(||X||^4) < \infty \), \( E(||X||^2||\varepsilon||^2) < \infty \) and \( \delta \) is fixed (for FPCA, FT, FLF, and FSC), then under \( H_0 \)

\[
W_N \overset{d}{\longrightarrow} \sum_{h=r+1}^{p} \sum_{\ell=1}^{+\infty} \lambda_{\ell,h} \chi^2_{\ell}(1, h) \quad \text{as} \quad N \to +\infty
\]

and under \( H_1 \),

\[
W_N \overset{d}{\longrightarrow} +\infty \quad \text{as} \quad N \to +\infty
\]

where \( \chi^2_{\ell}(1, h) \) 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 4 shows that $H_0$ is rejected if $W_N > \sum_{h=r+1}^p \sum_{\ell=1}^{+\infty} \lambda_{\ell,h} \chi^2_{\ell,\alpha}(1,h)$, for a certain level $\alpha$. This test is implemented in 3 simple steps. The first step is related to implementing the empirical eigenvalues of the covariance operators $\hat{K}_h$. The second step is to sort the eigenvalues in descending order. The third step is to compute the $p$-values of this distribution by the approach proposed by Rice et al. (2019); however, instead of using a Riemann integration approach, this paper uses a trapezoidal approach in order to reduce the discretization bias. The next section is related to the choice of the optimal tuning parameter.

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 $\alpha$ 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.

$$\min_{\delta \in A(\delta)} \frac{1}{N} \sum_{n=1}^{N-1} \int_0^1 \left[ X_{n+1}(t) - \int_0^1 \psi_\delta(t,s)X_n(s) ds \right]^2 dt,$$

where $A(\delta)$ is a set of $\delta$ values in which the good one should be selected. $\delta$ 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

$$\min_{\delta \in A(\delta)} \frac{1}{N} \sum_{n=1}^{N-1} \int_0^1 \left| X_{n+1}(t) - \int_0^1 \psi_\delta(t,s)X_n(s) ds \right| dt,$$

or the average out-of-sample $R^2$ ($AR^2_{oos}$) defined as

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

and 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.
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)

\[ X_{n+1}(t) = \int_0^1 \psi(t, s)X_n(s)ds + \varepsilon_{n+1}(t) \quad n = 1, \ldots, N. \]  

(29)

The three error processes used by Didericksen et al. (2012) \( \varepsilon^{(1)}(t) \), \( \varepsilon^{(2)}(t) \), and \( \varepsilon^{(3)}(t) \) are considered and are defined as follows:

\[ \varepsilon^{(1)}(t) = W(t) - tW(1). \]  

(30)

is the Brownian motion, where \( W \) is the standard Wiener process generated as

\[ W \left( b \right) = \frac{1}{\sqrt{B}} \sum_{\ell=1}^{b} Z_{\ell} \quad b = 1, \ldots, B. \]

and \( Z_{\ell} \) are independent standard normal variables and \( Z_0 = 0 \)

\[ \varepsilon^{(2)}(t) = \xi_1 \sqrt{2} \sin(2\pi t) + \xi_2 \sqrt{2} \kappa \cos(2\pi t) W(t) - tW(1). \]  

(31)

where \( \xi_1 \) and \( \xi_2 \) are two independent variables following a normal distribution and \( \kappa \) can be a constant. and

\[ \varepsilon^{(3)}(t) = a\varepsilon^{(1)}(t) + (1 - a)\varepsilon^{(2)}(t). \]  

(32)

where \( a \in [0, 1] \) is a real constant that represents the strength of the two components \( \xi_1 \) and \( \xi_2 \). \( \varepsilon^{(1)}(t) \) is an infinite series expansion, \( \varepsilon^{(2)}(t) \) is a finite series expansion, and \( \varepsilon^{(3)}(t) \) is the combination of the previous one.

The theoretical autoregressive operator \( \Psi \) is an integral operator mapping from \( H \) to \( H \). Two configurations of \( \Psi \) are considered, which are:

Model 1 : Gaussian operator (see Didericksen et al. (2012))

\[ \Psi(s, t) = C \exp \left[ -\frac{t^2 + s^2}{2} \right], \]

Model 2 : Factor model operator (see Imaizumi & 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 \( \Psi_{j,k} = 4(-1)^j k^{-\gamma} k^{-\beta} \) 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 \( ||\Psi||_\mathcal{L} = 0.8\).

\[ X_0(t) = \sum_{j=1}^{3} j^{-\rho/2} u_j v_j(t), \] where \( u_j \sim \text{Unif}([-3^{1/2}, 3^{1/2}]) \) with \( \rho = 1.2 \). A continuous interval of \([0,1]\) is considered. This interval consists of 1000 equally-spaced discretization. Two sample sizes of functional time series \(N\) is considered, which are 500 and 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 \( \lambda \) that is, the number of principal components \( m \) for the FPCA and FPLS methods and the regularization parameter \( \alpha \) for the FT, FLF, and FSC techniques. These parameters are chosen with respect to some estimation or prediction criteria. a recursive cross-validation is performed to choose the regularization parameter for each estimation method.

9.1 Indicators for estimation error

To analyze the estimation error, three criteria are considered. These are the mean squared error (MSE), the Average distance (AD), and the ratio averaged distance (RAD). Those criteria are given by

\[
MSE = \sqrt{\int_0^1 \int_0^1 \left( \hat{\Psi}(s,t) - \Psi(s,t) \right)^2 dsdt}
\]

\[
AD = \int_0^1 \int_0^1 |\hat{\Psi}(s,t) - \Psi(s,t)| dt
\]

\[
RAD = \int_0^1 \int_0^1 \frac{|\hat{\Psi}(s,t) - \Psi(s,t)|}{|\Psi(s,t)|} dsdt
\]

9.2 Indicators for prediction error

To measure the prediction quality, two indicators are considered, which are the root integrated squared error (En) and the integrated absolute error (Rn). Those criteria are given by

\[
E_n = \sqrt{\int_0^1 \left( \hat{X}_n(t) - X_n(t) \right)^2 dt}
\]
\[ R_n = \int_0^1 |\hat{X}_n(t) - X_n(t)| dt \]

We can also look at the global \( R^2 \) or the functional \( \tilde{R}^2(t) \) defined respectively as

\[ R^2 = \frac{\int_0^1 \text{var}(E[X_{n+1}(s) | X_n]) ds}{\int_0^1 \text{var}(X_{n+1}(s)) ds} \]

\[ \tilde{R}^2(t) = 1 - \frac{\sum \int_0^1 (X_{n+1}(s) - \hat{\psi}(X_n)(s))^2 ds}{\sum \int_0^1 (X_{n+1}(s) - \bar{X}_{n+1}(s))^2 ds} \]

For these last criteria, the tuning parameter is chosen such that the out-of-sample \( R^2 \) is maximized.

Table 5 reports the mean, median, and standard deviation for the different estimation and prediction criteria. The five-dimension reduction methods and the gaussian kernel are considered. One can observe that in terms of estimation purpose, the FLF method displays the best estimation performance. This pattern is observed on the MSE, RAD, and AD criteria. The second best performing method is the FPLS approach. Surprisingly, the FPCA and FSC methods do not display good performances and their patterns are almost similar. This performance of the FPCA and FSC method can be explained by the fact that the convergence rate depends on the high deceasing rate of the eigenvalues \( \lambda_j \) of the covariance operator \( K \), while the FLF and FT do not. In fact, the estimated autoregressive operator via these techniques is close to zero, as documented in figure 3. This pattern from the FPCA has also been noticed by Didericksen et al. (2012). To reduce the quick drop of the eigenvalues, they introduced an additional smoothing parameter; however, the way that parameter is chosen is not yet discussed. Even if the FPLS depend on the decreasing rate of the smallest eigenvalue of the Hankel matrix (\( \theta_m \)), its performance maybe due to fact that the eigenfunctions are constructed by taking into account their contribution to predicting the target variable. This pattern holds even if one looks at the mean or median value of the considered criterion.

When looking at the prediction criteria, one can observe that the FLF is still outperforming the other methods and is still followed by FPLS. This can be observed on the En, Rn criteria and is true for either the mean or the median. When considering the \( R^2_{oos} \) criterion, the FT outperforms the others and is followed by the FPLS, since they display the maximum amount of the \( R^2_{oos} \). The FPCA and FSC continue to underperform the others. When the sample size increases (\( N = 1000 \)), the same results are observed 6.

When considering the factor-based kernel, one can observe that the FT method outperforms based on the MSE and AD criteria. This method is followed by the FPCA and FSC methods. The FLF method

---

6This result can be also observed for the sloping s kernel considered by Didericksen et al. (2012) (see the appendix)
Figure 3: Estimated Gaussian autoregressive operator on the optimal tuning parameter

(a) FPCA
(b) FPLS
(c) FT
(d) FSC
(e) FLF
(f) True
Table 1: Comparison of the different estimation techniques. Model 1 with $M = 1000$ replications, and $\varepsilon^{(1)}$

<table>
<thead>
<tr>
<th></th>
<th>Moments</th>
<th>FPCA</th>
<th>FPLS</th>
<th>FT</th>
<th>FLF</th>
<th>FSC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MSE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.7900</td>
<td>0.3296</td>
<td>0.3546</td>
<td>0.2219</td>
<td>0.7999</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.0124</td>
<td>0.0285</td>
<td>0.0130</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.7800</td>
<td>0.3391</td>
<td>0.3726</td>
<td>0.2317</td>
<td>0.8000</td>
<td></td>
</tr>
<tr>
<td><strong>RAD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.9900</td>
<td>0.3784</td>
<td>0.3960</td>
<td>0.2394</td>
<td>0.9998</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.0151</td>
<td>0.0358</td>
<td>0.0149</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.9800</td>
<td>0.3895</td>
<td>0.4192</td>
<td>0.2507</td>
<td>0.9999</td>
<td></td>
</tr>
<tr>
<td><strong>AD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.7841</td>
<td>0.2709</td>
<td>0.3026</td>
<td>0.1809</td>
<td>0.7842</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.0125</td>
<td>0.0271</td>
<td>0.0113</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.7840</td>
<td>0.2809</td>
<td>0.3209</td>
<td>0.1897</td>
<td>0.7841</td>
<td></td>
</tr>
<tr>
<td><strong>Rn</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.8988</td>
<td>0.5643</td>
<td>0.6276</td>
<td>0.5458</td>
<td>0.8988</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0404</td>
<td>0.0134</td>
<td>0.0162</td>
<td>0.0087</td>
<td>0.0404</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.9236</td>
<td>0.5739</td>
<td>0.6379</td>
<td>0.5516</td>
<td>0.9236</td>
<td></td>
</tr>
<tr>
<td><strong>En</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.0685</td>
<td>0.7218</td>
<td>0.8187</td>
<td>0.7173</td>
<td>1.0685</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0403</td>
<td>0.0142</td>
<td>0.0252</td>
<td>0.0109</td>
<td>0.0403</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>1.0948</td>
<td>0.7323</td>
<td>0.8333</td>
<td>0.7243</td>
<td>1.0948</td>
<td></td>
</tr>
<tr>
<td><strong>R2is</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.0010</td>
<td>0.6480</td>
<td>0.3300</td>
<td>0.6515</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.0318</td>
<td>0.0224</td>
<td>0.2254</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.0011</td>
<td>0.6704</td>
<td>0.3474</td>
<td>0.8163</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td><strong>R2oos</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.0002</td>
<td>0.0152</td>
<td>0.0216</td>
<td>0.0096</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.0020</td>
<td>0.0034</td>
<td>0.0072</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.0002</td>
<td>0.0165</td>
<td>0.0238</td>
<td>0.0156</td>
<td>0.0002</td>
<td></td>
</tr>
</tbody>
</table>

Displays less performance. Based on the RAD criterion, the FPCA, FSC, and FLF document almost the same performance. In terms of prediction, the FT and FLF techniques are outperforming the other methods based on the Rn and En criteria. When looking at the $R^2_{oos}$ criterion, the FPLS technique tends to outperform the other methods and is followed by the FT method (see table 2). The same results are observed when the innovation process is smooth or in between smooth and non-smooth. The same pattern is observed when the sample size increases.

### 9.3 Selection of the optimal number of principal components

Concerning the selection of the optimal number of functional components, the FPLS tend 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 3).
Table 2: Comparison of the different estimation techniques. Model 2 with N = 500, M = 1000 replications, and \( \varepsilon^{(1)} \)

<table>
<thead>
<tr>
<th></th>
<th>Moments</th>
<th>FPCA</th>
<th>FPLS</th>
<th>FT</th>
<th>FLF</th>
<th>FSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>Mean</td>
<td>0.8000</td>
<td>1.1720</td>
<td>0.4748</td>
<td>15.6886</td>
<td>0.8000</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0001</td>
<td>0.1061</td>
<td>0.0249</td>
<td>2.1013</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.8000</td>
<td>1.2381</td>
<td>0.4822</td>
<td>15.9966</td>
<td>0.8000</td>
</tr>
<tr>
<td>RAD</td>
<td>Mean</td>
<td>1.0000</td>
<td>9.4339</td>
<td>1.4314</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0001</td>
<td>0.8699</td>
<td>0.0476</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>1.0000</td>
<td>9.9276</td>
<td>1.4606</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>AD</td>
<td>Mean</td>
<td>0.6365</td>
<td>0.9141</td>
<td>0.3850</td>
<td>11.2205</td>
<td>0.6365</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0001</td>
<td>0.0674</td>
<td>0.0187</td>
<td>1.3794</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.6365</td>
<td>0.9575</td>
<td>0.3883</td>
<td>11.6042</td>
<td>0.6365</td>
</tr>
<tr>
<td>Rn</td>
<td>Mean</td>
<td>0.7752</td>
<td>0.8070</td>
<td>0.6545</td>
<td>0.6702</td>
<td>0.7752</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0425</td>
<td>0.0737</td>
<td>0.0205</td>
<td>0.1214</td>
<td>0.0425</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.8076</td>
<td>0.8541</td>
<td>0.6654</td>
<td>0.7308</td>
<td>0.8076</td>
</tr>
<tr>
<td>En</td>
<td>Mean</td>
<td>0.9932</td>
<td>1.0700</td>
<td>0.8826</td>
<td>0.8161</td>
<td>0.9932</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0420</td>
<td>0.0919</td>
<td>0.0218</td>
<td>0.1808</td>
<td>0.0420</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>1.0247</td>
<td>1.1110</td>
<td>0.8963</td>
<td>0.8518</td>
<td>1.0247</td>
</tr>
<tr>
<td>R2is</td>
<td>Mean</td>
<td>0.0004</td>
<td>0.6834</td>
<td>0.2346</td>
<td>0.0417</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0002</td>
<td>0.0191</td>
<td>0.0326</td>
<td>0.0359</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.0004</td>
<td>0.6911</td>
<td>0.2410</td>
<td>0.0670</td>
<td>0.0003</td>
</tr>
<tr>
<td>R2oos</td>
<td>Mean</td>
<td>0.0004</td>
<td>0.0443</td>
<td>0.0064</td>
<td>0.0005</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0001</td>
<td>0.0745</td>
<td>0.0010</td>
<td>0.0012</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.0004</td>
<td>0.0072</td>
<td>0.0070</td>
<td>0.0001</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

Figure 4: Comparison of the different estimation techniques. Model 2 with N = 500, M = 1000 and \( \varepsilon^{(1)} \)
Table 3: Comparison of the number of selected component with factor based kernel. with
N = 500, M = 1000 replications

<table>
<thead>
<tr>
<th>error1</th>
<th>error2</th>
<th>error3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model1</td>
<td>Model2</td>
</tr>
<tr>
<td>MSE</td>
<td>FPCA</td>
<td>FPLS</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>RAD</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>AD</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rn</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>En</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>R2is</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>R2oos</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>

9.4 Optimal lag selection

To evaluate the performance of the test using simulated data, I use the data generation process proposed by Kokoszka & Reimherr (2013). I consider the following model

$$X_i = \Psi_1(X_{i-1}) + \Psi_2(X_{i-2}) + \epsilon_i,$$

with the kernel of $\Psi_h$ given by

$$\psi_h(s, t) = \frac{C_h}{0.7468} e^{-(s^2 + t^2)/2}.$$

with $C_h \in [0, 1)$ are the Hilbert-Schmidt norm of $\Psi_h$, respectively. The $\epsilon_i$ are standard Brownian motions. The rejection rates are based on one thousand replications, so the standard errors for the empirical size are about 0.009, 0.006, and 0.003, respectively for the nominal sizes of 0.10, 0.05, and 0.01. Two sample sizes are considered, which are $N = 200$ and $N = 500$. 

31
10 Application to the S&P 500 intraday data

10.1 Data

The S&P 500 Index data is used to analyze the intraday returns predictability. The sample data considered is from 01/01/2013 to 12/31/2017. The data is collected from a website called www.backtestmarket.com. This sample would be used in different parts to predict the return.

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

\[ X_{n+1}(t) = \Psi_0(t) + \int_0^1 \psi(s,t)X_n(s)ds + \varepsilon_{n+1}(t), \quad n = 1, \ldots, 1260. \]  

The sample size for this regression period is \( N = 1260 \). This sample is split in 3 sub-samples using the rolling scheme as described in section 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 5 displays a contour plot representing the correlation of the previous day cumulative return shape to the next one on a 1-minute frequency, that is the estimated autoregressive operator \( \hat{\psi}_\delta(s,t) \). It can be observed that the FT and FLF display almost similar results in terms of estimation, while the FPCA and FPLS display different results in terms of estimation. The yellow part represents a positive correlation, while the blue represents a negative correlation. Table 4 shows the results of the test to select the optimal number of lag in the model. The test is sequentially driven and \( \Psi_1, \Psi_2, \Psi_3 \) 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 to the zero operator.

It can be observed that the previous trading day first hour trading session (09:30 AM - 10:30 AM) contributes to predict positively to the next day return, last hour return of the previous day (02:30 PM - 04:00 PM) predicts positively the next day return in the first half trading session (09:30 AM - 10:30 AM), and predicts negatively the market return in the second half of the next trading day (02:30 PM - 04:00 PM).
10.3 Analyzing the functional out-of-sample $R^2$

In this section, the predictive functional $R^2_{\text{oos}}$ is derived and analyzed. From the figure 6, it is easy to see that the most important time of day to buy stocks is around 9:30 AM - 10:30 AM. In fact, the different methods show a very high $R^2_{\text{oos}}$ in that period of the day, which ensures that one can expect a very high volatility in the stock market at the opening hour. The result is similar for all of the different estimation methods. The FPLS tend to catch a remarkable value of 8% at the beginning and end of the trading session, that is almost 4 times the one obtained by Gao et al. (2018) and twice the one obtained by Zhang et al. (2019). The FPCA and FT methods tend to reach a predictive $R^2_{\text{oos}}$ of 5% and 6%, respectively in the period 9:30 AM - 10:30 AM (That is nearly twice the one obtained by Gao et al. (2018)) and nearly 2.5%
at the end of the trading session for the FT approach.

Figure 6: The estimated out-of-sample Functional R-Squared.

Surprisingly, the FPCA is not able to capture the momentum of the ending trading session. According to the FLF approach, the $R^2_{oos}$ is around 2.5% at the beginning of the trading session, while in the second half of the day it is approximately 1.2%. This result indicates that functional time series approaches should perform well in developing momentum or reversal indicators in the stock market.

10.4 Forecast accuracy

In this section, the forecast performance of the four estimation methods is evaluated with the prediction error criteria, which are the mean squared prediction error (MSPE), the mean absolute prediction error (MAPE), and the average out-of-sample R-squared ($AR^2_{oos}$). This result is calculated 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, the functional autoregressive model tends to produce more forecast accuracy than the AR(1) model. This may be due to the large number of parameters observed for the AR(1) model setting. This suggests that the functional data analysis approach is taking advantage of the additional news and improves
Table 5: **Comparison of the forecasting performance of the different methods for S&P500 CIDRs over the testing sample period**

<table>
<thead>
<tr>
<th>Moments</th>
<th>FPCA</th>
<th>FPLS</th>
<th>FT</th>
<th>FLF</th>
<th>FSC</th>
<th>VAR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPE Mean</td>
<td>0.8583</td>
<td>0.5643</td>
<td>0.6276</td>
<td>0.5458</td>
<td>0.8583</td>
<td>8.5510</td>
</tr>
<tr>
<td>Std</td>
<td>0.0410</td>
<td>0.0134</td>
<td>0.0162</td>
<td>0.0087</td>
<td>0.0420</td>
<td>0.0524</td>
</tr>
<tr>
<td>Median</td>
<td>0.9673</td>
<td>0.5739</td>
<td>0.6379</td>
<td>0.5516</td>
<td>0.9673</td>
<td>8.5320</td>
</tr>
<tr>
<td>MAPE Mean</td>
<td>1.0685</td>
<td>0.7318</td>
<td>0.8287</td>
<td>0.7273</td>
<td>1.1550</td>
<td>4.3280</td>
</tr>
<tr>
<td>Std</td>
<td>0.0543</td>
<td>0.0142</td>
<td>0.0252</td>
<td>0.0109</td>
<td>0.0533</td>
<td>0.0451</td>
</tr>
<tr>
<td>Median</td>
<td>1.0685</td>
<td>0.7323</td>
<td>0.8383</td>
<td>0.7234</td>
<td>1.1560</td>
<td>4.2230</td>
</tr>
<tr>
<td>$AR^2_{oos}$ Mean</td>
<td>0.0020</td>
<td>0.0182</td>
<td>0.0356</td>
<td>0.0096</td>
<td>0.0020</td>
<td>0.0030</td>
</tr>
<tr>
<td>Std</td>
<td>0.0011</td>
<td>0.0020</td>
<td>0.0034</td>
<td>0.0072</td>
<td>0.0011</td>
<td>0.0082</td>
</tr>
<tr>
<td>Median</td>
<td>0.0020</td>
<td>0.0165</td>
<td>0.0337</td>
<td>0.0156</td>
<td>0.0020</td>
<td>0.0030</td>
</tr>
</tbody>
</table>

Moreover, the FLF and FPLS models tend to outperform the other prediction methods based on the MSPE and MAPE criteria, while FPCA and FSC tend to underperform and display the same result. Furthermore, FLF outperforms the other methods in terms of the $AR^2_{oos}$. The same result is also observed when predicting the CIDRs constructed for S&P 500 in the period between January to June of 2018. This supports the fact that the obtained results on the functional time series forecasting are robust to the change of the analysis period. It could be interesting to develop a functional version of the Diebold & 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, which goes beyond the scope of this paper.

## 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 to adjust their portfolio rebalancing. This estimation problem leads to a high dimensionality problem and a comparative analysis of 4 big data techniques – FPLS, FT, FLF, and FSC – is developed. The theoretical results show that the MSE of the FT, FLF, and FSC display the same convergence rate when the signal to noise is difficult to recover, while FT is slower than FLF and FSC when it is easy to recover the signal. 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, FLF and FT methods tend to outperform the others in terms of estimation, while FPLS and FT tend to outperform in terms of prediction. The performance of FT and FLF is mainly related to the fact that these techniques do not depend on the decreasing rate of the eigenvalues of the covariance operator $K$, while the others do. More specifically, the FLF most of the time outperforms the FT method, as was already guessed by Benatia et al. (2017). Moreover, the FPLS method tends slightly to outperform the other methods in most cases in terms of prediction criteria. The theoretical results comparing FPCA and FPLS in terms of bias and variance of the MSE are verified in the simulations. FPCA and FSC methods seem to be very limited because these methods depend strongly on the eigen-system configuration of the covariance operator $K$ and the selected eigenfunctions are unsupervised.

The empirical application focuses on the prediction of the next day’s cumulative intraday return for the S&P 500. The results show that the FLF and the FT methods tend to display almost similar results in terms of the estimation of the autoregressive operator, while the FPCA/FSC and FPLS tend to document almost the same results. By relying on the expectations of the simulation results, for the estimation of the autoregressive operator, the focus is placed on the FLF method, while, for the prediction results, the focus is oriented to the FLF, FT, and FPLS. In terms of prediction criteria, the FLF and FPLS tend to outperform the others. Moreover, the FPLS tend to document a remarkable $R^2_{oos}$ of 8% in the period 9:30 AM - 10:30 AM and 2:30 PM - 4:00 PM within a trading day. This result is very interesting and it should be very important to verify the main causes of the momentum in those periods of the day, as was discussed by Gao et al. (2018) and Bogousslavsky (2016). Furthermore, after forecasting the intraday return, it could be interesting to test the prediction accuracy of the different approaches following the idea by Diebold & Mariano (1995). These potential extensions are left for future research.

12 Appendix. Proofs

12.1 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 & Hall (2012) for each $t$. The last result is obtained by exploiting the orthogonal polynomial representation as developed by Carrasco & Tsafack (2019).

Lemmas 1 to 3 below will be used in the proof of proposition 2.
Lemma 1

Under assumptions A1, A2, and A3, $\Psi_\delta^*$ is Hilbert-Schmidt for all $\delta$.

Proof of Lemma 1.

$$||\hat{\Psi}_\delta^*||^2_{HS} = ||\hat{K}_\delta^{-1}\hat{D}||^2_{HS} \leq ||\hat{D}||_{HS}||\hat{K}_\delta^{-1}||^2_{op}. $$

If $A$ is Hilbert-Schmidt operator and $B$ is a bounded operator, $||AB||_{HS} \leq ||A||_{HS}||B||_{op}$ with $||B||_{op} = \sup ||B(f)||$ the operator norm, it remains to prove that $||\hat{D}||^2_{HS} < +\infty$ and $||\hat{K}_\delta^{-1}||^2_{op} = O_p(1)$.

If the method is Functional Tikhonov,

$$||\hat{K}_\delta^{-1}||^2_{op} = ||(\alpha I + \hat{K})^{-1}||^2_{op} \leq \frac{1}{\alpha^2} = O_p(1).$$

If the method is Functional Landweber-Fridman,

$$||\hat{K}_\delta^{-1}||^2_{op} = ||d^{1/\alpha} \sum_{l=1}^{1/\alpha} (I - d\hat{K})^{-1}||^2_{op} \leq \sup_{j \geq 1} \left| \frac{(1 - (1 - d\hat{\lambda}_j)^{\frac{1}{\alpha}})}{\lambda_j} \right|^2 \leq C \frac{1}{\alpha^2} = O_p(1).$$

with $||\hat{K}||_{op} < \frac{1}{\alpha}$ and $C > 0$.

If the method is Functional Spectral Cut-off,

$$||\hat{K}_\delta^{-1}||^2_{op} \leq \sup_{j \geq 1} \left| \frac{I(\hat{\lambda}_j \geq \alpha)}{\lambda_j} \right|^2 \leq \frac{1}{\alpha^2} = O_p(1).$$
and we get the same result for the FPCA method, that is $||\hat{K}_\delta^{-1}||^2_{op} = O_p(1)$

If the method is Functional Partial Least Squares

$$||\hat{K}_\delta^{-1}||^2_{op} \leq \sup_{j \geq m+1} \left| \frac{\lambda_j}{\lambda_j} \right|^2 = O_p(1).$$

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 $\hat{D}^*(s,t) = \frac{1}{N-1} \sum_{n=1}^{N-1} X_i(s)X_{i+1}(t)$ and $X_n(t)$ belongs to $\mathbb{H}$.

**Lemma 2**

Under assumptions $A1 - A4$, $||K_\delta^{-1}(K - K_\delta)K^{\beta/2}R||^2_{HS}$ is

$$\begin{cases} O_p(\alpha^3) & \text{for FLF and FSC} \\
O_p(\alpha^{\min(\beta,2)}) & \text{for FT} \\
O_p(\lambda^{\beta}_{m+1}) & \text{for FPCA and FPLS}.
\end{cases}$$

**Proof of Lemma 2.**

$$||K_\delta^{-1}(K - K_\delta)K^{\beta/2}R||^2_{HS} \leq ||\hat{K}_\delta^{-1}||^2_{op} ||(K - K_\delta)K^{\beta/2}R||^2_{HS}$$

**FT method**

For the FT method, see lemma 10 of Benatia et al. (2017).

**FSC method**

$$K_\delta = \sum_{\lambda_j \geq \alpha} \lambda_j < v_{j+1} > v_j$$
\[
K - K_\delta = \sum_{\lambda_j < \alpha} \lambda_j < v_j, > v_j
\]

\[
K_\delta^{-1} = \sum_{\lambda_j \geq \alpha} \frac{1}{\lambda_j} < v_j, > v_j
\]

Then, \( \|K_\delta^{-1}\|_{op}^2 = O(\frac{1}{\alpha^2}) \) and

\[
\left\| (K - K_\delta) K^{\beta/2} R \right\|_{HS}^2 = \sum_{\lambda_j < \alpha} \lambda_j^{\beta+2} < R(v_j), R(v_j) >^2
\]

\[
\leq \left[ Sup_{\lambda_j < \alpha} \lambda_j^{\beta+2} \right] \sum_{\lambda_j < \alpha} < R(v_j), R(v_j) >^2
\]

\[
\leq C \alpha^\beta.
\]

where \( C > 0 \) is an arbitrary positive constant. Moreover, \( \sum_{\lambda_j < \alpha} < R(v_j), R(v_j) >^2 = \|R\|_{HS}^2 < +\infty \) (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 \( Q(\alpha, \lambda_j) = (1 - (1 - d\lambda_j)^{\frac{1}{2}}) \).

\[
\left\| (K - K_\delta) K^{\beta/2} R \right\|_{HS}^2 = \sum_{j=1}^{+\infty} \lambda_j^{\beta+2} (1 - Q(\alpha, \lambda_j))^2 < R(v_j), R(v_j) >^2
\]

\[
\leq \left[ Sup_{j} \lambda_j^{\beta+2} (1 - Q(\alpha, \lambda_j))^2 \right] \sum_{j=1}^{+\infty} < R(v_j), R(v_j) >^2
\]

\[
\leq C \alpha^{\beta+2}.
\]

Using Proposition 3.11 by 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**
\[ K_\delta = \sum_{j=1}^{m} \lambda_j < v_j, \ldots > v_j \]

\[ K - K_\delta = \sum_{j \geq m+1} \lambda_j < v_j, \ldots > v_j \]

\[ K_\delta^{-1} = \sum_{j=1}^{m} \frac{1}{\lambda_j} < v_j, \ldots > v_j \]

Then, \(||K_\delta^{-1}||^2_{\text{op}} = O\left(\frac{1}{\lambda_{m+1}}\right)\) and

\[ \left\| (K - K_\delta)K^{\beta/2}R \right\|_{\text{HS}}^2 = \sum_{j \geq m+1} \lambda_j^{\beta+2} < R(v_j), R(v_j) >^2 \]

\[ \leq \left[ \sup_{\lambda_j < \alpha} \lambda_j^{\beta+2} \right] \sum_{j \geq m+1} < R(v_j), R(v_j) >^2 \]

\[ \leq C\lambda_{m+1}^{\beta+2}. \]

where \( C > 0 \) is an arbitrary positive constant. Moreover, \( \sum_{j \geq m+1} < R(v_j), R(v_j) >^2 = ||R||^2_{\text{HS}} < +\infty \) (because \( R \) is Hilbert-Schmidt), which leads to the result.

**FPLS method**

\[ K_\delta^{-1} = \sum_{j=1}^{\infty} Q(m, \lambda_j) < v_j, \ldots > v_j \]

where \( Q(m, \lambda_j) = \left(1 - \prod_{l=1}^{m} \left(1 - \frac{\lambda_j}{\theta_l}\right)\right) \).

Then,

\[ ||K_\delta^{-1}||^2_{\text{op}} = \sup_{j \geq 1} Q(m, \lambda_j) \]

\[ \leq \sup_{j \geq m+1} Q(m, \lambda_j) \]

\[ = O(1). \]
The second line is true following proofs of Proposition 4 by Carrasco & Tsafack (2019) and

\[
\left\| (K - K_\delta)K^{\beta/2}R \right\|_{HS}^2 = \sum_{j=1}^{+\infty} \lambda_j^\beta (1 - Q(m, \lambda_j))^2 < R(v_j), R(v_j) >^2
\]

\[
\leq \left[ \sup_j \lambda_j^\beta (1 - Q(m, \lambda_j))^2 \right] \sum_{j=1}^{+\infty} < R(v_j), R(v_j) >^2
\]

\[
\leq \left[ \sup_{j \geq m+1} \lambda_j^\beta (1 - Q(m, \lambda_j))^2 \right] \sum_{j=1}^{+\infty} < R(v_j), R(v_j) >^2
\]

\[
\leq C\lambda_{m+1}^\beta.
\]

Using results of the proofs of Proposition 4 by Carrasco & Tsafack (2019), \( \sup_{j \geq 1} \left[ \lambda_j^\beta (1 - Q(\alpha, \lambda_j))^2 \right] \leq C\lambda_{m+1}^\beta \) (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.

### 12.2 Lemma 3

Under assumptions \( A1 - A4 \), for \( N \to \infty \), \( \left\| \hat{K}_\delta^{-1}(\Psi^*) - K^{-1}_\delta K(\Psi^*) \right\|_{HS}^2 \) is

\[
\begin{cases}
O_p\left(\frac{\alpha^\beta}{a^2N}\right) \text{ for FLF and FSC} \\
O_p\left(\frac{\text{min}(8,1)}{a^2N}\right) \text{ for FT} \\
O_p\left(\frac{1}{\lambda_{mN}}\right) \text{ for FPCA.}
\end{cases}
\]

**Proof of Lemma 3.**

\[
\hat{K}_\delta^{-1}(\Psi^*) - K^{-1}_\delta K(\Psi^*) = -\hat{K}_\delta^{-1}(\hat{K}_\delta - K_\delta)K_\delta^{-1}(K_\delta - K)(\Psi^*)
\]

\[
-\hat{K}_\delta^{-1}(\hat{K}_\delta - K_\delta)\Psi^*
\]

\[
+ \hat{K}_\delta^{-1}(K - K)(\Psi^*).
\]

Then,
\[
\left\| \hat{K}_\delta^{-1} \hat{K}(\Psi^*) - K_{\delta}^{-1}K(\Psi^*) \right\|_{HS}^2 \leq 3 \left\| \hat{K}_\delta^{-1}(\hat{K}_\delta - K_{\delta})K_{\delta}^{-1}(K_{\delta} - K)(\Psi^*) \right\|_{HS}^2 \\
+ 3 \left\| \hat{K}_\delta^{-1}(\hat{K}_\delta - K_{\delta})\Psi^* \right\|_{HS}^2 \\
+ 3 \left\| \hat{K}_\delta^{-1}(\hat{K} - K)\Psi^* \right\|_{HS}^2 \\
= 3(I) + 3(II) + 3(III).
\]

Moreover,

\[
(I) = \left\| \hat{K}_\delta^{-1}(\hat{K}_\delta - K_{\delta})K_{\delta}^{-1}(K_{\delta} - K)(\Psi^*) \right\|_{HS}^2 \\
\leq \left\| \hat{K}_\delta^{-1} \right\|_{op}^2 \left\| \hat{K}_\delta - K_{\delta} \right\|_{op}^2 \left\| K_{\delta}^{-1}(K_{\delta} - K)(\Psi^*) \right\|_{HS}^2.
\]

\[
(II) = \left\| \hat{K}_\delta^{-1}(\hat{K}_\delta - K_{\delta})\Psi^* \right\|_{HS}^2 \\
\leq \left\| \hat{K}_\delta^{-1} \right\|_{op}^2 \left\| \hat{K}_\delta - K_{\delta} \right\|_{op}^2 \left\| K_{\delta}^{3/2}R \right\|_{HS}^2.
\]

\[
(III) = \left\| \hat{K}_\delta^{-1}(\hat{K} - K)\Psi^* \right\|_{HS}^2 \\
\leq \left\| \hat{K}_\delta^{-1} \right\|_{op}^2 \left\| \hat{K} - K \right\|_{op}^2 \left\| K^{3/2}R \right\|_{HS}^2.
\]

For FT method:

See proof of Proposition 2 by Benatia et al. (2017).

For FLF and FSC methods:

Furthermore, \( \left\| \hat{K}_\delta^{-1} \right\|_{op}^2 = O_p(\frac{1}{N}) \), \( \left\| K_{\delta}^{-1} \right\|_{op}^2 = O_p(\frac{1}{N^2}) \), and from Lemma 2

\[
\left\| K_{\delta}^{-1}(K_{\delta} - K)(\Psi^*) \right\|_{HS}^2 = O_p\left( \alpha^\beta \right).
\]

Then, \( (I) = O_p\left( \frac{\alpha^\beta}{\alpha^{2N}} \right) \). Moreover, \( \left\| \hat{K}_\delta - K_{\delta} \right\|_{op}^2 = O_p\left( \frac{1}{N} \right) \), \( \left\| \hat{K} - K \right\|_{op}^2 = O_p\left( \frac{1}{N} \right) \)
Then, \( \left\| \hat{K}^{-1}_\delta (\hat{K} - K)K^{-1}_\delta (K - \bar{K})(\Psi^*) \right\|_{HS}^2 = O_p\left( \frac{\alpha_\delta}{\alpha N} \right) \).

Similarly, \((II) = O_p\left( \frac{1}{\alpha N} \right)\) and \((III) = O_p\left( \frac{1}{\alpha N} \right)\). Then, for FLF
\[
\left\| \hat{K}^{-1}_\delta \hat{K}(\Psi^*) - K^{-1}_\delta K(\Psi^*) \right\|_{HS}^2 = O_p\left( \frac{\alpha_\delta}{\alpha N} \right)
\]

**For FPCA method:**

Furthermore, \(||\hat{K}_\delta^{-1}||_{op}^2 = O_p\left( \frac{1}{\lambda_{m+1}^2} \right)\), \(||K_\delta^{-1}||_{op}^2 = O_p\left( \frac{1}{\lambda_{m+1}^2} \right)\), and from Lemma 2 we have
\[
\left\| K_\delta^{-1}(K_\delta - \bar{K})(\Psi^*) \right\|_{HS}^2 = O_p\left( \frac{\lambda^2}{\lambda_{m+1}^2} \right).
\]

Then,
\[
\left\| \hat{K}^{-1}_\delta (\hat{K} - K)K^{-1}_\delta (K - \bar{K})(\Psi^*) \right\|_{HS}^2 = O_p\left( \frac{\lambda^2}{\lambda_{m+1}^2 N} \right).
\]

Moreover, \(||\hat{K}_\delta - K||_{op}^2 = O_p\left( \frac{1}{N} \right)\), \(||\hat{K} - K||_{op}^2 = O_p\left( \frac{1}{N} \right)\), then, \((I) = O_p\left( \frac{1}{\lambda_{m+1}^2 N} \right)\), \((II) = O_p\left( \frac{1}{\lambda_{m+1}^2 N} \right)\) and \((III) = O_p\left( \frac{1}{\lambda_{m+1}^2 N} \right)\). Which lead to
\[
\left\| \hat{K}^{-1}_\delta \hat{K}(\Psi^*) - K^{-1}_\delta K(\Psi^*) \right\|_{HS}^2 = O_p\left( \frac{\lambda^2}{\lambda_{m+1}^2 N} \right) = O_p\left( \frac{1}{\lambda_{m}^2 N} \right).
\]

**Proof of the proposition 2.**

\[
\mathbb{E}\left[ \left\| \hat{\Psi}_\delta^* - \Psi^* \right\|_{HS}^2 \middle| \mathcal{F}_{N-1} \right] = \mathbb{E}\left[ \left\| \hat{K}^{-1}_\delta \hat{D} - K^{-1}_\delta D + K^{-1}_\delta D - \Psi^* \right\|_{HS}^2 \middle| \mathcal{F}_{N-1} \right]
\]
\[
= \mathbb{E}\left[ \left\| \hat{K}^{-1}_\delta \hat{K}\Psi^* + \hat{K}^{-1}_\delta \hat{C}_{xx} - K^{-1}_\delta D + K^{-1}_\delta D - \Psi^* \right\|_{HS}^2 \middle| \mathcal{F}_{N-1} \right]
\]
\[
= \mathbb{E}\left[ \left\| \hat{K}^{-1}_\delta \hat{C}_{xx} + \hat{K}^{-1}_\delta \hat{K}\Psi^* - K^{-1}_\delta K\Psi^* + K^{-1}_\delta D - \Psi^* \right\|_{HS}^2 \middle| \mathcal{F}_{N-1} \right]
\]
\[
= \mathbb{E}\left[ \left\| A + B + C \right\|_{HS}^2 \middle| \mathcal{F}_{N-1} \right].
\]

where \(A = \hat{K}^{-1}_\delta \hat{C}_{xx}, \) with \(\hat{C}_{xx} = \frac{1}{N-1} \sum_{i=1}^{N-1} X_i \otimes \varepsilon_{i+1}, \) \(B = \hat{K}^{-1}_\delta \hat{K}\Psi^* - K^{-1}_\delta K\Psi^* \) and \(C = K^{-1}_\delta D - \Psi^* .\)

Using Lemma 8 by Benatia et al. (2017), the last line of this equation yields
\[
E \left[ \left\| \hat{\Psi}_\delta - \Psi^* \right\|_{HS}^2 \bigg| \mathcal{F}_{N-1} \right] = E[\| A \|_{HS}^2 | \mathcal{F}_{N-1}] + \| B + C \|_{HS}^2 \\
\leq E[\| A \|_{HS}^2 | \mathcal{F}_{N-1}] + 2\| B \|_{HS}^2 + 2\| C \|_{HS}^2.
\]

Using Lemma 2 lead to \( \| C \|_{HS}^2 = \)
\[
\begin{cases}
O_p(\alpha^3) & \text{for FLF and FSC} \\
O_p(\alpha^{\min(\beta,2)}) & \text{for FT}.
\end{cases}
\]

Following Lemma 3, lead \( \| B \|_{HS}^2 = \)
\[
\begin{cases}
O_p(\frac{\alpha^2}{N}) & \text{for FLF and FSC} \\
O_p(\frac{\alpha^{\min(\beta,1)}}{\alpha N^2}) & \text{for FT}.
\end{cases}
\]

Then, it remains to derive the convergence rate of \( E[\| A \|_{HS}^2 | \mathcal{F}_{N-1}] \).

\[
E[\| A \|_{HS}^2 | \mathcal{F}_{N-1}] = E[\| \hat{K}_\delta^{-1} \hat{C}_{xx} \|_{HS}^2 | \mathcal{F}_{N-1}]
\]
\[
= E \left[ \operatorname{tr}[\hat{K}_\delta^{-1} \hat{C}_{xx} \hat{C}_{xx}^* \hat{K}_\delta^{-1}] \bigg| \mathcal{F}_{N-1} \right]
\]
\[
= \operatorname{tr} \left[ E[\hat{K}_\delta^{-1} \hat{C}_{xx} \hat{C}_{xx}^* \hat{K}_\delta^{-1}] | \mathcal{F}_{N-1} \right]
\]
\[
= \operatorname{tr} \left[ \hat{K}_\delta^{-1} E[\hat{C}_{xx} \hat{C}_{xx}^* | \mathcal{F}_{N-1}] \hat{K}_\delta^{-1} \right]
\]

where \( \operatorname{tr}(Z) \) is the trace of an arbitrary operator \( Z \), that is \( \| Z \|_{HS}^2 = \operatorname{tr}(ZZ^*) \). The third line in this equation is true following Lemma 9 by Benatia et al. (2017). Furthermore, since \( \hat{C}_{xx} = \frac{1}{N-1} \sum_{i=1}^{N-1} X_i \otimes \varepsilon_{i+1} \)

\[
\hat{C}_{xx} \hat{C}_{xx}^*(f) = \frac{1}{(N-1)^2} \sum_{i,j=1}^{N-1} X_i < X_j, f > < \varepsilon_{i+1}, \varepsilon_{j+1} >
\]

Therefore,
\[\mathbb{E}[\hat{C}_x \hat{C}_x^*(f)|\mathcal{F}_{N-1}] = \frac{1}{(N-1)^2} \sum_{i,j=1}^{N-1} X_i < f, X_j > \mathbb{E}[< \varepsilon_{i+1}, \varepsilon_{j+1} > |\mathcal{F}_{N-1}] \]

\[= \frac{1}{N-1} \sum_{i=1}^{N-1} < X_i, f > \mathbb{E}[< \varepsilon_{i+1}, \varepsilon_{i+1} > |\mathcal{F}_{N-1}]X_i \]

\[= \frac{1}{N-1} \sum_{i=1}^{N-1} < X_i, f > tr(V_\varepsilon)X_i \]

\[= \frac{1}{N-1} tr(V_\varepsilon) \sum_{i=1}^{N-1} < X_i, f > \]

\[= \frac{1}{N-1} tr(V_\varepsilon) \hat{K}(f). \]

where \( V_\varepsilon = \mathbb{E}[\varepsilon_i \otimes \varepsilon_i |\mathcal{F}_{N-1}] \). Since \( \varepsilon_i \) 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,

\[\mathbb{E}[< \varepsilon_i, \varepsilon_i > |\mathcal{F}_{N-1}] = \sum_{j=1}^{+\infty} < V_\varepsilon(v_i), v_i > \]

\[= tr(V_\varepsilon). \]

Then,

\[\mathbb{E}[||A||_{HS}^2 |\mathcal{F}_{N-1}] \leq \frac{1}{N-1} tr \left[ \hat{K}_\delta^{-1} \hat{K}_\delta^{-1} \right] tr \left[ V_\varepsilon \right]. \]

For FT method:

\[tr \left[ \hat{K}_\delta^{-1} \hat{K}_\delta^{-1} \right] = \sum_{j=1}^{+\infty} \frac{\lambda_j}{(\lambda_j + \alpha)^2} \leq \int_{0}^{+\infty} \frac{x}{(x+\alpha)^2} dx \]

\[= \frac{1}{\alpha} \left[ \frac{-1}{x+\alpha} \right]_{0}^{+\infty} \leq \frac{C}{\alpha^2}. \]

where \( C \) is an arbitrary positive number.

For FSC method:
\[ \text{tr} \left[ \hat{K}^{-1}_\delta \hat{K} \hat{K}^{-1}_\delta \right] = \sum_{\lambda_j \geq \alpha} \frac{\lambda_j}{\hat{\lambda}_j^2} \leq \frac{C}{\alpha^2}. \]

The last line holds since \( \sum_{\lambda_j \geq \alpha} \hat{\lambda}_j < +\infty \) (\( \hat{K} \) is nuclear).

For FLF method:

\[ \text{tr} \left[ \hat{K}^{-1}_\delta \hat{K} \hat{K}^{-1}_\delta \right] = \sum_{j \geq 1} \left( 1 - \frac{(1 - d\hat{\lambda}_j)^{1/\alpha}}{\lambda_j} \right)^2 \leq \frac{C}{\alpha^2} \]

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 \( \mathbb{E} \left[ ||A||^2_{HS} | \mathcal{F}_{N-1} \right] = O_p \left( \frac{1}{\alpha^2 N} \right) \) for FT, FLF and FSC.

The convergence rate of \( ||C||^2_{HS}, ||B||^2_{HS} \) and \( \mathbb{E} \left[ ||A||^2_{HS} | \mathcal{F}_{N-1} \right] \) leads to

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

12.3 Proof of the proposition 3.

Following the same argument as in Proposition 1, we have

\[
\mathbb{E} \left[ \left| \left| \hat{\Psi}_\delta^* - \Psi^* \right| ^2_{HS} \right| \mathcal{F}_{N-1} \right] = \mathbb{E} \left[ \left| \left| \hat{K}^{-1}_\delta \hat{D}^* - K^{-1}_\delta D^* + K^{-1}_\delta D^* - \Psi^* \right| ^2_{HS} \right| \mathcal{F}_{N-1} \right] \leq \mathbb{E} \left[ ||A||^2_{HS} | \mathcal{F}_{N-1} \right] + 2 ||B||^2_{HS} + 2 ||C||^2_{HS}.
\]

where \( A = \hat{K}^{-1}_\delta \hat{C}_x \), with \( \hat{C}_x = \frac{1}{N-1} \sum_{i=1}^{N-1} X_i \otimes \varepsilon_{i+1} \), \( B = \hat{K}^{-1}_\delta \hat{K} \Psi^* - K^{-1}_\delta K \Psi^* \) and \( C = K^{-1}_\delta D^* - \Psi^* \).
For FPCA method:

Following Lemma 2, $||C||_{HS}^2 = O_p\left(\frac{\lambda_{m+1}^B}{mN}\right)$ and using Lemma 3, $||B||_{HS}^2 = O_p\left(\frac{1}{\lambda_mN}\right)$. Then, it remains to derive the convergence rate of $\mathbb{E}\left[||A||_{HS}^2 | \mathcal{F}_{N-1}\right]$

Following the same arguments as in Proposition 2,

$$\mathbb{E}[||A||_{HS}^2 | \mathcal{F}_{N-1}] \leq \frac{1}{N-1} \text{tr}\left[\hat{K}_\delta^{-1}\hat{K}_\delta^{-1}\right] \text{tr}\left[V_\varepsilon\right].$$

and

$$\text{tr}\left[\hat{K}_\delta^{-1}\hat{K}_\delta^{-1}\right] = \sum_{j=1}^{m} \frac{\hat{\lambda}_j}{\lambda_j^2} = \sum_{j=1}^{m} \frac{1}{\hat{\lambda}_j} \leq \frac{Cm}{\lambda_m}.$$ 

where $C > 0$ is an arbitrary constant. Then, $||A||_{HS}^2 = O_p\left(\frac{m}{\lambda_mN}\right)$. Combining the rate of convergence for $||A||_{HS}^2, ||B||_{HS}^2, ||C||_{HS}^2$ leads to

$$\mathbb{E}\left[||\hat{\Psi}_\delta^* - \Psi^*||_{HS}^2 | \mathcal{F}_{N-1}\right] = O_p\left(\frac{\lambda_{m+1}^B}{mN}\right) + O_p\left(\frac{m}{\lambda_mN}\right).$$

For the FPLS method:

Let us define

$$X = \left[X_1, X_2, ..., X_{N-1}\right]^\prime$$

$$Y = \left[X_2, X_3, ..., X_N\right]^\prime$$

and

$$\xi = \left[\varepsilon_2, \varepsilon_3, ..., \varepsilon_N\right]^\prime$$

Let us denote by

$$\Pi_a = \text{Span}\{v_j : \lambda_j \leq a\}$$
and

$$\hat{H}_a = \text{Span}\{\hat{v}_j : \lambda_j \leq a\}$$

the orthogonal projection onto the eigenvectors of the covariance operator $K$ (respectively $\hat{K}$) for which the corresponding eigenvalues $\lambda_j (\hat{\lambda}_j)$ 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

$$||\tilde{\Psi}_{PLS} - \Psi^*||_{HS} \leq ||\hat{H}_a(\tilde{\Psi}_{PLS} - \Psi^*)||_{HS} + ||(I - \hat{H}_a)(\tilde{\Psi}_{PLS} - \Psi^*)||_{HS}$$

$$\leq ||\hat{H}_a(\tilde{\Psi}_{PLS} - \Psi^*)||_{HS} + ||\hat{H}_a(\tilde{\Psi}_{PLS} - \Psi^*)||_{HS} + ||(I - \hat{H}_a)(\tilde{\Psi}_{PLS} - \Psi^*)||_{HS}$$

$$\leq (F1) + (F2) + (F3).$$

where $(I - \hat{H}_a)$ is the orthogonal complement of the space $\hat{H}_a$. Let us define $(F1)$, $(F2)$, and $(F3)$ by

$(F1) = ||\hat{H}_a(\tilde{\Psi}_{PLS} - \Psi^*)||_{HS}, (F2) = ||\hat{H}_a(\tilde{\Psi}_{PLS} - \Psi^*)||_{HS}$ and $(F3) = ||(I - \hat{H}_a)(\tilde{\Psi}_{PLS} - \Psi^*)||$ 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

$$(F1) = ||\hat{H}_a(\tilde{\Psi}_{PLS} - \Psi^*)||_{HS}$$

$$= ||\hat{H}_a\{\hat{P}_m(\hat{K})(T_n^*(\Sigma)) - \hat{P}_m(\hat{K})(\hat{K}(\Psi^*))\}||_{HS}$$

$$= ||\hat{H}_a\{\hat{P}_m(\hat{K})(T_n^*(\xi))\}||_{HS}$$

$$\leq \left\{ \sup_{\lambda \in [0,a]} |\hat{P}_m(\hat{\lambda})| \right\} ||\hat{H}_a(T_n^*(\xi))||_{HS}$$

$$\leq C|\hat{Q}_m'(0)||\hat{H}_a(T_n^*(\xi))||_{HS},$$

where $C > 0$ is an arbitrary constant. The last line holds because for $0 < \lambda < a$,

$$\hat{P}_m(\lambda) = \frac{1 - \hat{Q}_m(\lambda)}{\lambda} \leq |\hat{Q}_m'(0)| = \sum_{l=1}^{m} \frac{1}{\theta_l} \propto_p \left( \sum_{l=1}^{m} \frac{1}{\theta_l} \right).$$

(See Engl et al. (1996)). On the other hand, we have
\[ |\hat{Q}_m(0)| \asymp_p |Q_m(0)| \]

since \((\theta_j)_{j \geq 1}\) are distinct and are consistently estimated by \((\hat{\theta}_j)_{j \geq 1}\).

Moreover, \(T_n^*(\xi) = \hat{C}_{xe} = \frac{1}{N-1} \sum_{i=1}^{N-1} X_i \otimes \varepsilon_{i+1}\). Then,

\[
\mathbb{E} \left[ ||T_n^*(\xi)||_{HS}^2 |\mathcal{F}_{N-1} \right] = \mathbb{E} \left[ tr(\hat{C}_{xe}\hat{C}_{xe}) |\mathcal{F}_{N-1} \right] = \frac{1}{N-1} tr(V_{\varepsilon}) tr(\hat{K}) = O_p \left( \frac{1}{N} \right).
\]

Since \(\sum_{t=1}^{m} \frac{1}{\hat{\theta}_t} \leq C \frac{m}{\hat{\theta}_m}\) (with \(C > 0\) an arbitrary constant),

\[
(F1)^2 = O_p \left( \frac{m}{\hat{\theta}_m N} \right).
\]

**Upper bound rate of (F2)**:

\[
(F2) = ||\hat{\Pi}_a(\hat{\Psi}_{PLS} - \Psi^*)||_{HS}
\]

\[= ||\hat{\Pi}_a(\hat{P}_m(\hat{K})(\hat{K}(\Psi^*)) - \Psi^*)||_{HS}\]

\[= ||\hat{\Pi}_a(\hat{Q}_m(\hat{K})(\Psi^*))||_{HS}\]

\[\leq \left\{ \sup_{\lambda \in [0,a]} |\hat{Q}_m(\lambda)| \right\} ||\hat{\Pi}_a(\Psi^*)||_{HS}\]

\[\leq C ||\hat{\Pi}_a(\Psi^*)||_{HS}.
\]

where \(C\) is an arbitrary positive constant. This is possible since \(\{ \sup_{\lambda \in [0,a]} |\hat{Q}_m(\lambda)| \} \leq 1\) by definition of \(\hat{Q}_m\) (see Lemma 1 by Carrasco & Tsafack (2019)). Moreover, ||\(\hat{\Pi}_a(\Psi^*)||_{HS} \asymp_p ||\Pi_a(\Psi^*)||_{HS}\), since that \((\hat{v}_j)_{j \geq 1}\) and \((\hat{v}_j)_{j \geq 1}\) are consistently identified. Moreover,
We have

\[ \|\Pi_a(\Psi^*)\|_{HS}^2 = \sum_{\lambda_j < a} < \Psi^*(v_j), \Psi^*(v_j)>^2 \]

\[ = \sum_{\lambda_j < a} < K^{\beta/2} R(v_j), K^{\beta/2} R(v_j)>^2 \]

\[ = \sum_{\lambda_j < a} \lambda_j^\beta < R(v_j), R(v_j)>^2 \]

\[ = \left[ \text{Sup} \{ \lambda_j^\beta \} \right] \sum_{j \geq 1} < R(v_j), R(v_j)>^2 \]

\[ = O_p\left( \text{Sup} \{ \lambda_j^\beta \} \right). \]

Then

\[ (F2)^2 = O_p\left( \text{Sup} \{ \lambda_j^\beta \} \right). \]

Upper bound rate of \((F3):\)

\[ T^*_n(\mathbf{Y}) = \frac{1}{N-1} \sum_{i=1}^{N-1} X_i \otimes Y_{i+1} = \hat{D}^*. \]

Moreover, we have

\[ (F3) = \| (I - \hat{\Pi}_a)(\Psi_{FPLS}^* - \Psi^*) \|_{HS} \]

\[ = \| (I - \hat{\Pi}_a)(\hat{K}^{-1}_\delta \hat{K})(\hat{P}_m(\hat{K})(T^*_n(\mathbf{Y})) - \Psi^*) \|_{HS} \]

\[ \leq \| \hat{\Pi}_a(\hat{K}^{-1}_\delta) \| \| (I - \hat{\Pi}_a)(\hat{K})(\hat{P}_m(\hat{K})(T^*_n(\mathbf{Y})) - \Psi^*) \|_{HS} \]

\[ \leq \frac{1}{\sqrt{a}} \| (I - \hat{\Pi}_a) \{ \hat{K}(\hat{P}_m(\hat{K})(T^*_n(\mathbf{Y})) - \hat{K}(\Psi^*)) \} \|_{HS} \]

\[ \leq \frac{1}{\sqrt{a}} \| (I - \hat{\Pi}_a) \{ T^*_n \{ \hat{W}_n(\hat{P}_m(W_n)(\mathbf{Y})) - T_n(\Psi^*) \} \} \|_{HS} \]

\[ \leq \frac{1}{\sqrt{a}} \| (I - \hat{\Pi}_a) \{ T^*_n \{ \hat{W}_n(\hat{P}_m(W_n)(\mathbf{Y})) - \hat{Y} + \mathbf{Y} - T_n(\Psi^*) \} \} \|_{HS} \]

\[ \leq \frac{1}{\sqrt{a}} \| (I - \hat{\Pi}_a) \{ T^*_n \{ \hat{Q}_m(W_n)(\mathbf{Y}) \} \} \|_{HS} + \frac{1}{\sqrt{a}} \| (I - \hat{\Pi}_a) \{ T^*_n(\mathbf{Z}) \} \|_{HS} \]

\[ \leq \frac{1}{\sqrt{a}} \| (I - \hat{\Pi}_a) \{ \hat{Q}_m(\hat{K})(T^*_n(\mathbf{Y})) \} \|_{HS} + \frac{1}{\sqrt{a}} \| (I - \hat{\Pi}_a) \{ T^*_n(\mathbf{Z}) \} \|_{HS}. \]

where \(\hat{K}^{-1}_\delta\) is the generalized inverse of \(\hat{K}\) using the FPCA regularization. The fourth and fifth lines are possible given that \(\hat{K} = T^*_n T_n\) and \(W_n = T_n T^*_n\). The seventh line is possible because \(\hat{Q}_m(W_n)(\mathbf{Y}) = \)
\[ \hat{W}_n \{ \hat{P}_m(W_n)(Y) - Y \} - \xi = Y - T_n(\Psi^*) \]. The last line comes from the fact that \( \hat{Q}_m(\hat{K})(T_n^*(Y)) = T_n^*(\hat{Q}_m(W_n)(Y)) \) and this is because we have \( \hat{K} = T_n^*T_n \) and \( W_n = T_nT_n^* \).

Let us consider \( T_1 \) defined by \( T_1 = \hat{Q}_m(\hat{K})(T_n^*(Y)) \).

On the other hand, we have

\[
||T_1||_{HS} \leq ||\hat{\Pi}_{\hat{\theta}_m} \{ \varphi_m(W_n)(T_n^*(Y)) \}||_{HS} \\
\leq ||\hat{\Pi}_{\hat{\theta}_m} \{ \varphi_m(W_n)(\hat{K}(\Psi^*)) \}||_{HS} + ||\hat{\Pi}_{\hat{\theta}_m} \{ \varphi_m(W_n)(T_n^*\xi) \}||_{HS} \\
\leq \left\{ \sup_{t \in [0, \hat{\theta}_m]} |t^{1/2}\varphi_m(t)| \right\} ||\hat{\Pi}_{\hat{\theta}_m} (\Psi^*)||_{HS} + \left\{ \sup_{t \in [0, \hat{\theta}_m]} |\varphi_m(t)| \right\} ||\hat{\Pi}_{\hat{\theta}_m} (T_n^*\xi)||_{HS}.
\]

where,

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

Moreover, following lemma 1 by Carrasco & Tsafack (2019),

\[ \sup_{t \in [0, \hat{\theta}_m]} |t^{1/2}\varphi_m(t)| \leq |\hat{Q}_m(0)|^{-1/2} \]

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

and \( ||T_n^*\xi||_{HS}^2 = O_p(\frac{1}{N}) \). Then,

\[
||T_1||^2 \leq C_1 |\hat{Q}_m(0)|^{-1} ||\hat{\Pi}_{\hat{\theta}_m}(\Psi^*)||_{HS}^2 + C_2 \frac{1}{N}
\]

where \( C_1 \) and \( C_2 \) are arbitrary positive constants. Therefore, for each element \( a \) such that \( 0 < a \leq |\hat{Q}_m(0)|^{-1} \leq \theta_m \), the upper bound rate of \( (F3) \) is given by

\[ (F3)^2 = O_p \left( \frac{1}{a} |\hat{Q}_m(0)|^{-1} ||\Pi_a(\Psi^*)||_{HS}^2 \right) + O_p \left( \frac{1}{aN} \right) \]

By taking \( a = |\hat{Q}_m(0)|^{-1} \) and combining results from \( (F1), (F2) \) and \( (F3) \), yields.
\[ \|\hat{\Psi}_{FPLS} - \Psi^*\|_{HS}^2 = O_p\left(\|\Pi_{\theta_m}(\Psi^*)\|_{HS}^2\right) + O_p\left(\frac{1}{N}|Q'_m(0)|^{-1}\right) \]
\[ = O_p\left(\sup_{\lambda_j < \theta_m} \{\lambda_j^2\}\right) + O_p\left(\frac{1}{N} \sum_{i=1}^{m} \frac{1}{\theta_i}\right) \]
\[ = O_p\left(\sup_{\lambda_j < \theta_m} \{\lambda_j^2\}\right) + O_p\left(\frac{m}{N\theta_m}\right) \]
\[ = O_p\left(\lambda_{m+1}^\beta\right) + O_p\left(\frac{m}{N\theta_m}\right). \]

The first term is the bias term and the second one is the variance. The last line is possible since we have \[ \sum_{j=m+1}^{\infty} \Psi_j^2 = \|\Psi_{FPCA} - \Psi^*\|_{HS}^2 = O_p\left(\lambda_{m+1}\right), \] combined with the fact that \(\lambda_{m+1}\) is the highest eigenvalue of \(K\) that is lower than \(\theta_m\) (see lemma 1 by Carrasco & Tsafack (2019)). Therefore, \[ \|\hat{\Psi}_{PLS} - \Psi^*\|_{HS}^2 = O_p\left(\lambda_{m+1}\right). \]

### 12.4 Proof of the proposition 4.

For \(\delta\) fixed (\(m\) for FPCA and \(\alpha\) for FT, FSC and FLF),

\[ \hat{\Psi}_\delta = \hat{K}_\delta^{-1} \hat{D}^* \]
\[ = \hat{K}_\delta^{-1} \hat{K}(\Psi^*) + \hat{K}_\delta^{-1} \hat{C}_{x,\varepsilon}. \]

and \(\Psi_\delta^* = K_\delta^{-1} K(\Psi^*)\) Then,

\[ \hat{\Psi}_\delta^* - \Psi_\delta^* = \hat{K}_\delta^{-1} \hat{C}_{x,\varepsilon} + \hat{K}_\delta^{-1} \hat{K}(\Psi^*) - K_\delta^{-1} K(\Psi^*) \]
\[ = K_\delta^{-1} \hat{C}_{x,\varepsilon} + (\hat{K}_\delta^{-1} - K_\delta^{-1}) \hat{K}(\Psi^*) + K_\delta^{-1} \hat{K}(\Psi^*) - K_\delta^{-1} K(\Psi^*) \]
\[ = K_\delta^{-1} \hat{C}_{x,\varepsilon} + (\hat{K}_\delta^{-1} - K_\delta^{-1}) \hat{K}(\Psi^*) + (\hat{K}_\delta^{-1} - K_\delta^{-1}) \hat{K}(\Psi^*) + \hat{K}_\delta^{-1} (K_\delta - K_\delta)(\Psi^*) \]
\[ = K_\delta^{-1} \hat{C}_{x,\varepsilon} + \hat{K}_\delta^{-1} (K_\delta - K_\delta)(\Psi^*) + O_p\left(\frac{1}{N}\right) \]
\[ = K_\delta^{-1} \left[ \frac{1}{N} \sum_{i=1}^{N-1} X_i \otimes \varepsilon_{i+1} \right] + K_\delta^{-1} \left[ \frac{1}{N} \sum_{i=1}^{N-1} (X_i \otimes X_i - K) \Psi^* \right] + O_p\left(\frac{1}{N}\right). \]

Let us consider \(\tilde{X}_i = K_\delta^{-1}(X_i)\) and \(\tilde{X} = K_\delta^{-1}(X)\). Then,
\[
\dot{\Psi}_\delta^* - \Psi_\delta^* = \frac{1}{N} \sum_{i=1}^{N-1} \tilde{X}_i \otimes \varepsilon_{i+1} + \frac{1}{N} \sum_{i=1}^{N-1} \tilde{X}_i \otimes \Psi(X_i) - \mathbb{E}[\tilde{X} \otimes \Psi(X)] + O_p\left(\frac{1}{N}\right)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N-1} \tilde{X}_i \otimes \varepsilon_{i+1} + \frac{1}{N} \sum_{i=1}^{N-1} \Psi[\tilde{X}_i \otimes X_i - \mathbb{E}[\tilde{X} \otimes X]] + O_p\left(\frac{1}{N}\right).
\]

Under the assumptions that \(X_i \otimes \varepsilon_{i+1}\) are stationary and ergodic martingale difference sequences, \(\mathbb{E}[||X_i||^4] < +\infty\) and \(\mathbb{E}[||X_i||^2||\varepsilon_{i+1}||^2] < +\infty\),

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N-1} \begin{bmatrix} X_i \otimes \varepsilon_{i+1} \\ X_i \otimes X_i \end{bmatrix} \xrightarrow{d} N(0, \Omega_1)
\]

where \(\Omega_1 = \begin{pmatrix} G & 0 \\ 0 & K \end{pmatrix}\) is a \((2 \times 2)\) matrix of covariance operators. By the continuous mapping theorem with the transformation,

\[
\begin{bmatrix} A \\ B \end{bmatrix} \rightarrow K_\delta^{-1}A + \Psi K_\delta^{-1}(B - K)
\]

yields the asymptotic covariance operator of \(\sqrt{N}(\dot{\Psi}_\delta^* - \Psi_\delta^*)\)

\[
\Omega_\delta = \mathbb{E}\left[\tilde{X} \otimes (\tilde{X} + \varepsilon) - \mathbb{E}[\Psi X \otimes \tilde{X}]\right] \tilde{\otimes} \left[(\varepsilon + \tilde{X}) \otimes \tilde{X} - \mathbb{E}[\Psi X \otimes \tilde{X}]\right]
\]

\[
= \mathbb{E}\left[K_\delta^{-1}(X_i \otimes \varepsilon_{i+1}) \tilde{\otimes} (\varepsilon_{i+1} \otimes X_i)K_\delta^{-1}\right] + \mathbb{E}\left[\Psi K_\delta^{-1}(X_i \otimes X_i - K) \tilde{\otimes} [(X_i \otimes X_i - K)K_\delta^{-1}\Psi]\right]
\]

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

where \(\tilde{\otimes}\) is the tensor product of two operators. Then, for \((A, B) \in \mathbb{H}^H \times \mathbb{H}^H\), \(A \tilde{\otimes} B\) is an element of the Hilbert space of Hilbert-Schmidt operators from \(\mathbb{H}^H\) to \(\mathbb{H}^H\). If \(\delta\) is such that \(m \to +\infty\) for FPCA or \(\alpha \to 0\) for FT, FLF, and FSC, then \(\Omega_\delta \sim \Omega_0\), where
\begin{align*}
\Omega_0 &= \mathbb{E}\left[ (\tilde{X} \otimes \varepsilon_{i+1}) \tilde{\otimes} (\varepsilon_{i+1} \otimes \tilde{X}) \right] \\
&= \mathbb{E}\left[ K^{-1} (X \otimes \varepsilon_{i+1}) \tilde{\otimes} (\varepsilon_{i+1} \otimes X) K^{-1} \right] \\
&= \mathbb{E}\left[ K^{-1} G K^{-1} \right] \\
&= K^{-1} \mathbb{E}[G] K^{-1} \\
&= \sigma^2 \text{Id} \tilde{\otimes} K^{-1}.
\end{align*}

where \( G = \mathbb{E}[(X \otimes \varepsilon_{i+1}) \tilde{\otimes} (\varepsilon_{i+1} \otimes X)] \mathcal{F}_{N-1} \) The third line holds by the iterated expectation theory and the last line holds by the homoskedasticity assumption.

12.5 Proof of the proposition 5.

Under \( H_0 \), for each \( h \in \{r+1, \ldots, p\} \),

\[ \sqrt{N}(\hat{\Psi}_{h,\eta_h} - 0) \overset{d}{\rightarrow} N(0, \Omega_{\eta_h}) \quad \text{as} \quad N \rightarrow +\infty \]

where the asymptotic covariance operator admits the following eigen-decomposition

\[ \Omega_{\eta_h}^{-1} = \sum_{\ell=1}^{+\infty} \frac{[Q(\lambda_{\ell,h}, \eta_h)]^2}{\sigma^2 \lambda_{\ell,h}} v_{\ell} \otimes v_{\ell} \]

Then, for each \( \ell \geq 1 \) and \( h \in \{r+1, \ldots, p\} \),

\[ N \frac{[Q(\hat{\lambda}_{\ell,h}, \eta_h)]^2}{\hat{\sigma}^2 \hat{\lambda}_{\ell,h}} < \hat{\Psi}^*_{h,\beta_h}(\hat{v}_{\ell}), \hat{\Psi}^*_{h,\beta_h}(\hat{v}_{\ell}) > \overset{d}{\rightarrow} \chi^2_1(h, \ell) \quad \text{as} \quad N \rightarrow +\infty. \]

where \( \chi^2_1(h, \ell) \) is a random variable that follows a \( \chi^2(1) \) distribution for each \( h \) and \( \ell \). Therefore,

\[ N \sum_{h=r+1}^{p} \|\hat{\Omega}_{h,\beta_h}^{-1/2} \hat{\Psi}_{h,\eta_h}\|^2_{HS} \overset{d}{\rightarrow} \sum_{h=r+1}^{p} \sum_{\ell=1}^{+\infty} \lambda_{\ell,h} \chi^2_1(h, \ell) \quad \text{as} \quad N \rightarrow +\infty. \]

Under \( H_1 \), there exists \( h \in \{r+1, \ldots, p\} \) such that \( \|\hat{\Omega}_{h,\eta_h}^{-1/2} \hat{\Psi}_{h,\eta_h}\|^2_{HS} \neq 0 \)

\[ N \sum_{h=r+1}^{p} \|\hat{\Omega}_{h,\eta_h}^{-1/2} \hat{\Psi}_{h,\eta_h}\|^2_{HS} \overset{d}{\rightarrow} +\infty \quad \text{as} \quad N \rightarrow +\infty. \]
13 Graphics and Tables

13.1 Graphics

Figure 7: Comparison of the different estimation techniques. Model 1 with $N = 500$, $M = 1000$ and $\varepsilon^{(2)}$. 

![Graphs and Tables](image-url)
Figure 8: Comparison of the different estimation techniques. Model 2 with $N = 500$, $M = 1000$ and $\varepsilon^{(2)}$.

Figure 9: Comparison of the different estimation techniques with Model 1 with $N = 500$, and $\varepsilon^{(3)}$. 
Figure 10: Comparison of the different estimation techniques. Model 2 with $N = 500$, $M = 1000$ and $\varepsilon^{(3)}$. 

![Comparison of the different estimation techniques](image-url)
Figure 11: Comparison of the different estimation techniques. Model 2 with $N = 500$, $M = 1000$ and $\varepsilon^{(1)}$

Figure 12: The estimated Functional R-Squared. 2018
### Table 6: Comparison of the different estimation techniques. Sloping kernel with N = 500, M = 1000 replications, and $\varepsilon^{(1)}$

<table>
<thead>
<tr>
<th></th>
<th>Moments</th>
<th>FPCA</th>
<th>FPLS</th>
<th>FT</th>
<th>FLP</th>
<th>FSC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MSE</strong></td>
<td>Mean</td>
<td>0.7990</td>
<td>0.4074</td>
<td>0.5298</td>
<td>0.4599</td>
<td>0.8000</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0001</td>
<td>0.0153</td>
<td>0.0363</td>
<td>0.0278</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.7890</td>
<td>0.4182</td>
<td>0.5564</td>
<td>0.4798</td>
<td>0.7990</td>
</tr>
<tr>
<td><strong>RAD</strong></td>
<td>Mean</td>
<td>0.9990</td>
<td>2.0639</td>
<td>1.1268</td>
<td>0.9357</td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0001</td>
<td>0.2851</td>
<td>0.0384</td>
<td>0.0043</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.9800</td>
<td>2.2466</td>
<td>1.1519</td>
<td>0.9386</td>
<td>0.9880</td>
</tr>
<tr>
<td><strong>AD</strong></td>
<td>Mean</td>
<td>0.6928</td>
<td>0.3531</td>
<td>0.4245</td>
<td>0.3838</td>
<td>0.6928</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0001</td>
<td>0.0111</td>
<td>0.0261</td>
<td>0.0228</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.6930</td>
<td>0.3608</td>
<td>0.4430</td>
<td>0.3998</td>
<td>0.6930</td>
</tr>
<tr>
<td><strong>Rn</strong></td>
<td>Mean</td>
<td>0.7932</td>
<td>0.5468</td>
<td>0.6136</td>
<td>0.5333</td>
<td>0.7932</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0656</td>
<td>0.0154</td>
<td>0.0310</td>
<td>0.0149</td>
<td>0.0656</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.8337</td>
<td>0.5565</td>
<td>0.6347</td>
<td>0.5424</td>
<td>0.8337</td>
</tr>
<tr>
<td><strong>En</strong></td>
<td>Mean</td>
<td>0.9679</td>
<td>0.6981</td>
<td>0.7846</td>
<td>0.7000</td>
<td>0.9679</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0644</td>
<td>0.0155</td>
<td>0.0306</td>
<td>0.0169</td>
<td>0.0644</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>1.0063</td>
<td>0.7077</td>
<td>0.8040</td>
<td>0.7103</td>
<td>1.0063</td>
</tr>
<tr>
<td>$R^2_{oos}$</td>
<td>Mean</td>
<td>0.0016</td>
<td>0.6822</td>
<td>0.3336</td>
<td>0.6257</td>
<td>0.0077</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0002</td>
<td>0.0355</td>
<td>0.0482</td>
<td>0.3265</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.0017</td>
<td>0.7044</td>
<td>0.3646</td>
<td>0.8471</td>
<td>0.0083</td>
</tr>
<tr>
<td>$R^2_{oos}$</td>
<td>Mean</td>
<td>0.0002</td>
<td>0.0216</td>
<td>0.0226</td>
<td>0.0829</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0001</td>
<td>0.0050</td>
<td>0.0044</td>
<td>0.0651</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.0002</td>
<td>0.0250</td>
<td>0.0255</td>
<td>0.1376</td>
<td>0.0002</td>
</tr>
</tbody>
</table>
Table 7: Comparison of the different estimation techniques. Model 3 with M = 1000 replications, and ε(1)

<table>
<thead>
<tr>
<th></th>
<th>Var1</th>
<th>FPCA</th>
<th>FPLS</th>
<th>FT</th>
<th>FLF</th>
<th>FSC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MSE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.7990</td>
<td>0.6648</td>
<td>0.3670</td>
<td>0.2515</td>
<td>0.7999</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.0213</td>
<td>0.0287</td>
<td>0.0087</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.7980</td>
<td>0.6727</td>
<td>0.3852</td>
<td>0.2560</td>
<td>0.7999</td>
<td></td>
</tr>
<tr>
<td><strong>RAD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.9980</td>
<td>2.6554</td>
<td>1.0042</td>
<td>0.9879</td>
<td>0.9999</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.1526</td>
<td>0.0295</td>
<td>0.0143</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.9970</td>
<td>2.7577</td>
<td>1.0210</td>
<td>0.9937</td>
<td>0.9999</td>
<td></td>
</tr>
<tr>
<td><strong>AD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.6394</td>
<td>0.5246</td>
<td>0.2973</td>
<td>0.2028</td>
<td>0.6393</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.0052</td>
<td>0.0251</td>
<td>0.0081</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.6395</td>
<td>0.5287</td>
<td>0.3147</td>
<td>0.2069</td>
<td>0.6393</td>
<td></td>
</tr>
<tr>
<td><strong>Rn</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.7761</td>
<td>0.5497</td>
<td>0.5730</td>
<td>0.5097</td>
<td>0.7761</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0385</td>
<td>0.0143</td>
<td>0.0155</td>
<td>0.0107</td>
<td>0.0385</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.8007</td>
<td>0.5581</td>
<td>0.5783</td>
<td>0.5148</td>
<td>0.8006</td>
<td></td>
</tr>
<tr>
<td><strong>En</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.0051</td>
<td>0.7342</td>
<td>0.7667</td>
<td>0.6764</td>
<td>0.1050</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0443</td>
<td>0.0170</td>
<td>0.0232</td>
<td>0.0098</td>
<td>0.0443</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>1.0317</td>
<td>0.7414</td>
<td>0.7738</td>
<td>0.6830</td>
<td>0.1037</td>
<td></td>
</tr>
<tr>
<td><strong>R2is</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.0011</td>
<td>0.6532</td>
<td>0.3533</td>
<td>0.1096</td>
<td>0.0011</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.0353</td>
<td>0.0472</td>
<td>0.0479</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.0011</td>
<td>0.6752</td>
<td>0.3805</td>
<td>0.1437</td>
<td>0.0011</td>
<td></td>
</tr>
<tr>
<td><strong>R2oos</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.0001</td>
<td>0.0664</td>
<td>0.0252</td>
<td>0.0904</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.0905</td>
<td>0.0049</td>
<td>0.1037</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.0001</td>
<td>0.0519</td>
<td>0.0280</td>
<td>0.1150</td>
<td>0.0001</td>
<td></td>
</tr>
</tbody>
</table>

Table 8: Comparison of the different estimation techniques. Model 3 with M = 1000 replications, and ε(3)

<table>
<thead>
<tr>
<th></th>
<th>Moments</th>
<th>FPCA</th>
<th>FPLS</th>
<th>FT</th>
<th>FLF</th>
<th>FSC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MSE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.8000</td>
<td>1.5038</td>
<td>0.7626</td>
<td>3.8919</td>
<td>0.8000</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.0710</td>
<td>0.0060</td>
<td>0.0465</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.8000</td>
<td>1.5220</td>
<td>0.7647</td>
<td>3.8908</td>
<td>0.8000</td>
<td></td>
</tr>
<tr>
<td><strong>RAD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.0000</td>
<td>10.5133</td>
<td>1.2253</td>
<td>1.0002</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0002</td>
<td>0.5375</td>
<td>0.0116</td>
<td>0.0001</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>1.0000</td>
<td>10.8600</td>
<td>1.2328</td>
<td>1.0003</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td><strong>AD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.6365</td>
<td>1.0820</td>
<td>0.6157</td>
<td>2.7054</td>
<td>0.6365</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0003</td>
<td>0.0372</td>
<td>0.0039</td>
<td>0.0451</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.6365</td>
<td>1.0924</td>
<td>0.6171</td>
<td>2.6988</td>
<td>0.6365</td>
<td></td>
</tr>
<tr>
<td><strong>Rn</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.7078</td>
<td>0.7043</td>
<td>0.7014</td>
<td>0.7573</td>
<td>0.7078</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0076</td>
<td>0.0030</td>
<td>0.0053</td>
<td>0.0065</td>
<td>0.0076</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.7132</td>
<td>0.7058</td>
<td>0.7050</td>
<td>0.7605</td>
<td>0.7132</td>
<td></td>
</tr>
<tr>
<td><strong>En</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.9536</td>
<td>0.9481</td>
<td>0.9475</td>
<td>1.0318</td>
<td>0.9536</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0072</td>
<td>0.0033</td>
<td>0.0045</td>
<td>0.0075</td>
<td>0.0072</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.9581</td>
<td>0.9496</td>
<td>0.9490</td>
<td>1.0363</td>
<td>0.9581</td>
<td></td>
</tr>
<tr>
<td><strong>R2is</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.0003</td>
<td>0.1496</td>
<td>0.0738</td>
<td>0.0003</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.0049</td>
<td>0.0020</td>
<td>0.0001</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.0003</td>
<td>0.1522</td>
<td>0.0740</td>
<td>0.0003</td>
<td>0.0004</td>
<td></td>
</tr>
<tr>
<td><strong>R2oos</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.0001</td>
<td>0.0033</td>
<td>0.0004</td>
<td>0.0001</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>0.0001</td>
<td>0.0006</td>
<td>0.0000</td>
<td>0.0002</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.0001</td>
<td>0.0036</td>
<td>0.0004</td>
<td>0.0001</td>
<td>0.0001</td>
<td></td>
</tr>
</tbody>
</table>
References


