Financial Applications of Forecasting Models for Functional Time Series

Carlos Sanjuán Ruiz, *Master's Degree Thesis*, Master in Connected Industry, Universidad Pontificia Comillas - ICAI

Supervised by: Antonio Muñoz San Roque and Guillermo Mestre Marcos

***Resumen* - En el presente documento se analizan las capacidades de predicción de los modelos de series temporales funcionales cuando se aplican sobre las curvas obtenidas de los mercados financieros. Estas curvas se estudiarán utilizando técnicas de Análisis de Datos Funcionales (FDA, del inglés *Functional Data Analysis*), que han demostrado ser útiles cuando se trabaja con datos de alta dimensión. Las curvas originales se descompondrán en las primeras componentes principales funcionales, reduciendo de manera efectiva la dimensión de los datos. Una vez que los datos se han expresado en un espacio de baja dimensión, se ajustarán diversos modelos predictivos (como ARIMA, Función de Transferencia y Perceptrón Multi-capa) para estimar curvas futuras . Para el caso de estudio se utilizan las curvas de rendimiento del EURIBOR, generadas a partir de tipos de interés discretos con vencimientos de una, dos, tres semanas y de uno a doce meses, entre septiembre de 2016 y febrero de 2020. Se demostrará la utilidad de los modelos propuestos en un período de baja volatilidad de las curvas y en un período diferente, en el que la volatilidad de las curvas es alta. Se utilizarán variables exógenas para mejorar los modelos de previsión, obteniendo estimaciones más robustas.**

***Abstract* - This paper analyses the predictive capabilities of functional time series models when applied to curves obtained from financial markets. These curves will be studied using Functional Data Analysis (FDA) techniques, which have been proven to be useful when working with high-dimensional data. The original curves will be decomposed into their first functional principal components, effectively reducing the dimension of the data. Once the data has been represented in a low-dimensional space, several forecasting models (such as ARIMA, Transfer Function and Multi-Layer Perceptron) will be fitted to the data to estimate future curves. A case study using EURIBOR yield curves, generated from discrete interest data at maturities of one, two, three weeks and one to twelve months, between September 2016 and February 2020 is included, which illustrates the applicability of the proposed models in a period of low-volatility of the curves and a different period where the volatility of the curves is high. Exogenous variables will be used to improve the forecasting models, obtaining more robust forecasts.**

Key Words: Functional Data Analysis, EURIBOR, Principal Functional Components, ARIMA, Transfer Function, Multilayer Perceptron.

1 Introduction

Due to recent technological advances in science and industry, the collection and analysis of high-dimensional data has become an important topic in the field of applied statistics. The Functional Data Analysis (FDA) framework is a collection of statistical tools that enable the study of such high dimensional data, often recorded as a set of curves.

FDA techniques have gained importance in recent years due to the increase in applications in different fields, such as environmental sciences, mathematics, public health and medicine, geography or economics, among others [11].

These types of high dimensional data are present in several financial applications and can be categorized in two different classes. The first type of product are high volatility intraday indexes, as seen in [15], where the S&P500, DAX30, CAC40 and USD-Euro are analysed with one minute resolution. In [12] and [?] only the S&P500 is analysed with a five minute resolution. A variation of the S&P500 is the S&P100, analysed by [6] with a resolution of five minutes, or for example [17], whose authors model the volatility index VIX of the future PUT options of the S&P500 offered by the CBOE Chicago Board Options Exchange every fifteen seconds. Each of the curves is modelled as a curve defined by the discrete prices observed over the curse of one day, from opening to closing. In Figure 1, it can be seen how the time series of the S&P500 is divided into five functional data observations.

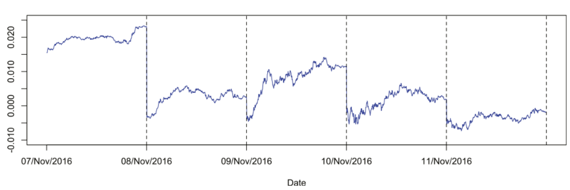

Figure 1: Division into functional data of the S&P500 stock exchange index, with one minute resolution during the month of November 2016. Source: [15]

The second type of financial product, commonly used for FDA, are *yield curves*, as illustrated in [4], which analyses month-end price curves (average supply and demand) for U.S. Treasury bonds with maturities of three to one hundred months. In [2], the author analyses EURIBOR curves sized for maturities of three months to fifteen years as a functional data set. [16] compares US interest rates (maturity period from three months to thirty years) and India (maturity periods from three months to fifteen years). And finally, the curves determining the value of Eurodollar futures contracts from three months to ten years of maturity are analysed in [10]. These types of curves are made up of interest rates or future values at different maturity periods, each of which forms the functional study data set. Figure 2 shows a typical data set for this type of financial product in 3D format.

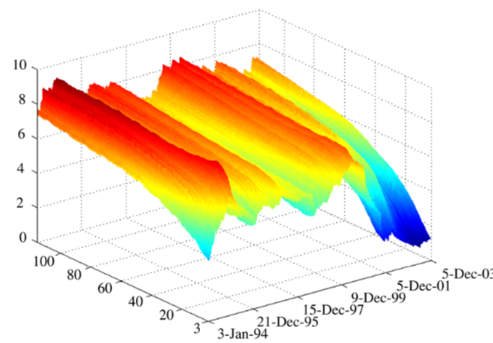

Figure 2: Yield curves for Eurodollar futures contracts. X-axis: maturation periods in months. Y-axis: date to which each of the curves belongs. z-axis: value of the contract. Source: [10]

One of the main disadvantages of having a functional data set is the difficulty of operating with a large dimensional space. Therefore, in order to reduce the dimension of the working space to a M dimensional space, dimensionality reduction techniques are usually applied. The most common technique used is the decomposition into principal components, where the dimension of the reduced workspace depends on how many principal components are needed to explain a great percentage of the variance of the original data. Thus, in [15] the authors reduce the space to four dimensions, in [5], [4] and [9] they set M to three principal functional components, while in [16] one principal component is sufficient. This methodology generates a reduced dimension functional base where the original data set can be expressed as a linear combination. Another technique widely used to reduce the size of the workspace, is based on an exponential approximation (see [4] and [2]), developed by Nelson and Siegel [13], which decomposes the set of curves into a three-dimensional space, characterizing the level, slope and curvature of the functions, generating a base in which the original curves are expressed as linear combinations. A variant of this methodology, which is mainly used for the analysis of *yield curves*, is the one developed by Diebold and Li [4], which offers a much more dynamic adjustment.

Once the dimension of the functional dataset has been reduced, a predictive model is trained to estimate future coefficients of the functional time series. The most commonly used models are those of an auto-regressive nature; these models use the historical data of the input variables to adjust the model with great efficiency, having a *conditional mean* nature. Examples of such models are: the *ARIMA* model [1], [2] and [7], been the most widely used model which includes autoregressive terms (AR), integral terms (I) and moving average terms (MA). A multivariate version of the AR model is the *Vector Auto Regressive model VAR*, used in [4] and [17]. But in [4] the authors argue that the VAR model has poor results when analysing financial products, due to the complexity and nature of the data. Combining the two models mentioned above together generates a new model called *VARMA*, presented in [16]. In [11] the authors propose a functional version of the AR model (the *FAR* model) that it is able to model functional data directly, thus not require the use of dimensionality reduction techniques. On the other hand, there are two other time series models that, instead of estimating the conditional mean of the data, model the volatility (i.e. changes in the variance) of the time series. These models are often referred to as *Conditional Variance models*, such as the *GARCH*, which, is used in several works like in [15] and [6], where they also perform a functional analysis generating a *FGARCH* model offering optimistic results in FDA. Finally, in [5] and [9] a non-linear model based on *Radial Basis Function Neural Network* is used for predicting future coefficients of the time series.

2 EURIBOR yield curves

The functional dataset to be analysed in this project consists of the EURIBOR (European Interbank Offered Rate) yield curves. Each curve represents the interest rate at which financial institutions in the Eurozone lend money to each other. This interest rate varies according to the maturity of the loan. The European Central Bank is responsible for calculating this rate on a daily basis, based on the estimated average number of operations that will be carried out by the most powerful entities that handle the Euro.

Every day the interest rate is marked at one to three weeks and one to twelve months, being the time to maturity of one year the most used by financial institutions. These indices will determine each of the curves to be analysed. An example of the type of curve to be studied is shown in the Figure 3.

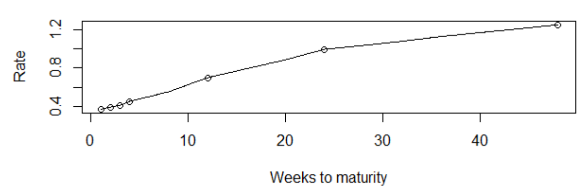

Figure 3: Euribor curve [2010-01-04]. Value of the interest y-axis. Time horizon in weeks x-axis.

Before starting the study and the training of the model, a visual inspection of the dataset has been carried out, with the intention of knowing a priori how the data evolves over time and cleaning the raw information.

For the first stage of this project, it was decided to choose a period in which the series would exhibit low variability, with the purpose of obtaining a simple predictive model that is capable of capturing the dynamic of the series. Figure 4 shows the data set that has been selected for this first phase of the project. As expected, a stable dataset with low dispersion over time can be observed.

Once the study dataset has been fixed, it is divided into two subsets. A first subset called the training period TR will be used as the known historical data in real application. Models will be identified and fitted using this data, characterizing the original functional data as accurately as possible. The second subset, called validation period VAL, is taken as a reference to determine the efficiency of the estimates generated by the FDA predictive model. For this project, the original set will be divided in a 70/30 proportion for the training and validation subset respectively.

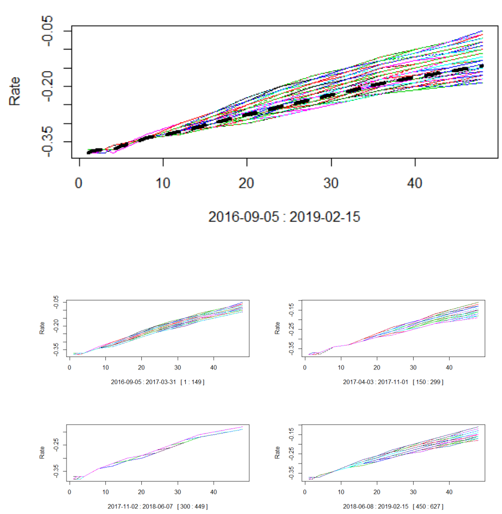

Figure 4: Low-dispersion curve section between 2016-2019.

3 Functional Forecasting models

This section will introduce the predictive models that have been developed for estimating future EURIBOR curves. The forecasting strategy of the models can be summarized as a three step process: Functional Principal Component decomposition, modelling of the principal component's scores and curve reconstruction. Subsection 3.1 will introduce the Functional Principal Component Analysis that will be used. Subsection 3.2 is devoted at explaining the forecasting models that will be used to estimate the scores of the principal components. Finally, the process of reconstruction of the estimated curves generated by the models will be explained in Subsection 3.1.1.

This Section will also include the development of the processes carried out on the set of exogenous variables in Subsection 3.4 and the period of high volatility in the Subsection 3.5.

3.1 Functional Principal Component Analysis

This FDA method consists in decomposing the set of study functions $Y_t(x)$ into a number M of Principal Functional Components [PFC] $\Phi(x)$ which are able to explain the rest of functions when multiplied by each of the coefficients $C_{t,j}$ associated to the original functions, being T the size of the function time series.

$$Y_t\ (x) \approx \widehat{\mu}(x)\ +\ \sum_{j=1}^{M} \widehat{c}_{t,j}\phi_j(x),\ for\ t=1,....,T \quad (1)$$

The function *pca.fd* (R language) can be used to decompose the set of curves, selected to train the model, into a number M of principal components. Each one of the new FPC is able to explain the variance of the original curves in a determined percentage. Being the first component the most representative in the system, as shown in Figure 5. When accumulating the information offered by each one of the components, it can explain better the original curves. From this concept, the number of M principal components to be used is determined by analysing the compromise between the *precision* and *complexity* of the model.

The explanations shown in this document will

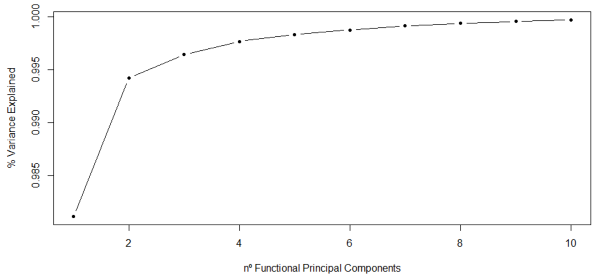

Figure 5: Percentage of information explained by the first 10 principal components.

be made with the first three principal components, which explain the 99.64% of the information of the original dataset. This way, the following principal functions are shown in Figure 6 with which this model will be trained.

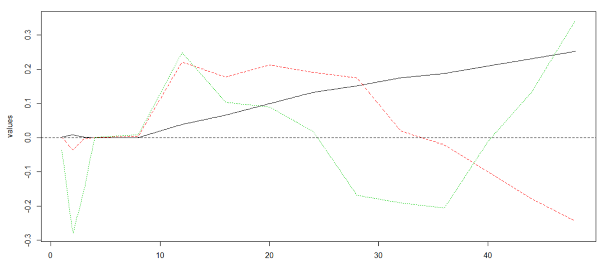

Figure 6: First 3 FPC (1º Black, 2º Red, 3º Green)

At the same time that the FPC basis are obtained, M coefficients are associated to each of the original curves. Generating M series of coefficient with the size of the number of curves of the dataset. These are represented in Figure 7.

Therefore, the training period is expressed in the base formed by the first three FPCs. In the same way, if the coefficients associated to the curves of the validation period are estimated and the curves expressed in this same base (generated with the TR set) are reconstructed, future EURIBOR curves can

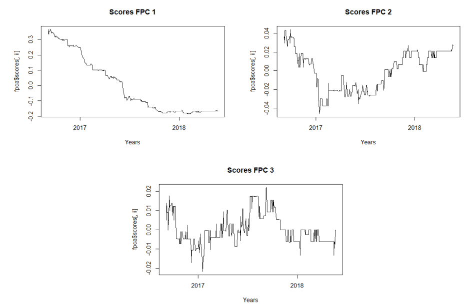

Figure 7: Scores curves of the first 3 FPCs

be estimated.

3.1.1 Reconstruction error

When breaking down a large number of curves to a few principal components, an intrinsic error appears, usually called *Reconstruction Error*. This value marks a limit in the efficiency of the models. On the other hand, these errors can be a reference to determining which will be the optimal value of functional principal components to be included. To calculate this error, the curves are decomposed into their first N_c principal components, and then reconstructed using the scores of the decomposition.

Finally, the error committed between the reconstructed set and the original set is calculated, allowing us to measure the *Reconstruction Error*, using several error metrics. Looking at Figure 8 it is possible to visualize the results and determine what would be the optimal number of principal components to use in the project.

Different error measures are determined to calculate the efficiency of the estimated results, which will be used throughout the project. These error metrics have the capacity to analyse functional data, and are: the Functional Mean Square Error (FRMSE) 2, the Functional Mean Absolute Error (FMAE) 3 and the Functional Mean Absolute Percentage Error

(FMAPE) 4.

$$FRMSE = \sqrt{T^{-1} \sum_{t=1}^{T} \int \left(Y_t(u) - \widehat{Y}_t(u) \right)^2 du} \quad (2)$$

$$FMAE = T^{-1} \sum_{t=1}^{T} \int \left| Y_t(u) - \widehat{Y}_t(u) \right| du \quad (3)$$

$$FMAPE = T^{-1} \sum_{t=1}^{T} \frac{\int \left| Y_t(u) - \widehat{Y}_t(u) \right| du}{\int |Y_t(u)| du} \quad (4)$$

Figure 8 compares the results of the *Naive* model and the *Reconstruction Error* which determines the prediction limit of the model. Both in the TR and VAL periods, visually it could be determined that for the selected low volatility data set, the optimal number will be around six principal components.

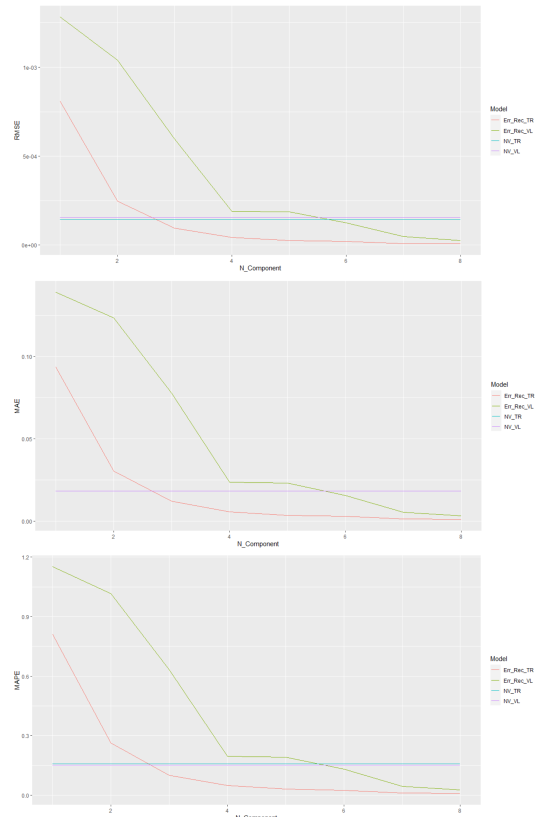

Figure 8: Reconstruction error FRMSE, FMAE and FMAPE of the period TR and VAL (Y axis) and the first 8 principal components (x axis). Including the errors made by the *Naive* model in the TR and VAL period.

The errors mentioned above represent different precision metrics (or imprecision) of the functional data calculated in the project with respect to the original. As they are functional, they use integrals in one way or another to determine the area along all points of the curve. For the specific application of the analysis models used on the EURIBOR curves, it would also be appropriate to determine the precision of each of the observations with which the curves are constructed, since the ultimate objective is to determine what the interest rates will be in the different maturity periods for a given prediction horizon. And to be able to contrast the efficiency of the models for each of the points used to construct the original set of functional data. In this way, the error *MAE in basic points* will be calculated, from the following expression 5, in the Figure 9, the graph of the reconstruction errors is shown under the mentioned metric for the first eight principal components.

$$MAE(u_p) = 100 \left(T^{-1} \sum_{t=1}^{T} \left| Y_t(u_p) - \widehat{Y}_t(u_p) \right| \right) \quad (5)$$

The result will generate a vector with dimension P, being the number of original observations, where the average error for each of these points will be determined. In addition, if we want to know the global error along the whole curve, the $MAEPB$ is calculated from the *MEA in basic points* as shown in the following expression 6.

$$MAEPB = P^{-1} \sum_{p=1}^{P} MAE(u_p) \quad (6)$$

As expected, the error produced by the model is different for each of the interest rates. In addition, depending on the number of principal components se-

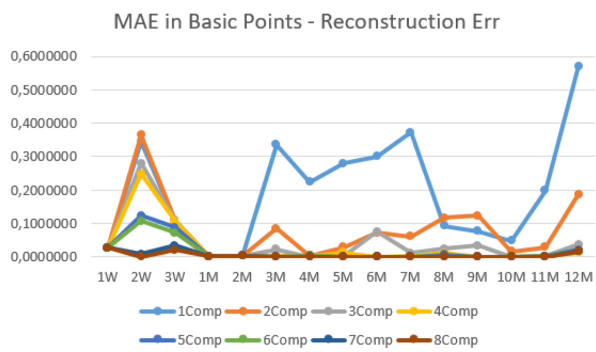

Figure 9: MAE error in basic points, for the reconstruction errors with the first eight principal components.

lected, the error varies progressively, which was also expected. Thanks to these graphs, it is visually determined which of the basic points will be the optimal one for financial applications, once the adjustment of the predictive models is completed. That is to say, the interest rates which are most reliable in this period of time will be those with a maturity time of one, two, four, ten and eleven months.

3.2 Forecasting models

In this project the following predictive models will be used according to the needs of the analysed datasets, so that the functional data sets can be modelled and new EURIBOR curves can be predicted in the future, as set out in the objectives of this document.

3.2.1 ARIMA model

The Auto Regressive Moving Average model, ARMA is the most common model to analyse a stationary time series y_t. This model describes the time series y_z in terms of two polynomials: the first one accounts for the autoregressive term of the series, and the second one accounts for the moving average terms of the series. The equation for the ARMA(p,q) model is 7. Where ψ_i are the autoregressive coefficients, θ_i are the moving average coefficients and ε_t are the innovations of the process.

$$y_t = \sum_{i=1}^{p} \psi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \qquad (7)$$

If the time series is not stationary in the mean, it is necessary to apply a differentiation process for which this mean is stabilised. A variant of the ARMA model that contains this new methodology is the *Auto Regressive Integrated Moving Average* (ARIMA). This model consists of three variables to be controlled, d determines the number of differentiations, p controls the coefficients of the AR effect and q the effects number of MA terms.

When fitting the ARIMA model, it is necessary to identify the autoregressive and moving average orders of the model. The *Auto Correlation Function (ACF)* and the *Partial Auto Correlation Function (PACF)* are used to identify these orders, as explained in the document [3].

3.2.2 Transfer Function, TF

The Transfer Function model is a multiple dynamic regression model with ARIMA noise. It works by regression methodologies applied to the output y_T and to different input variables $x^i_{t,h}$. Once the regression terms have been characterised, the regression residuals are modelled with an ARIMA model [14]. This model, unlike the ARIMA model which only analyses the historical data of the time series, by complementing the information of the series using multiple input variables, generates more robust estimates.

3.2.3 Multilayer Perceptron, MLP

The neural network *Multilayer Perceptron* is a non-linear model capable of modeling non-linear relations between the data from a deep analysis of the input data.

The main structure of the network consists of three type of layers: input, output and hidden layers. The input and output layers are used to manage the *inputs* and *outputs* respectively. In this case study a single hidden layer will be used, which includes a network of *perceptrons*. This *perceptrons* are processing units or neurons, which analyse the non-linear relationships between their inputs through an iterative learning process.

3.2.4 Naive, NV

The following model is commonly called the *dummy* model, due to its simplicity and because it is not very intuitive. For the training of the model the hypothesis is established in which the value of the prediction in $t+1$ is equal to the value in t, in other words $\widehat{y}_t = y_{t-1}$ [8].

The results provided by this model serve as a reference for a first approximation of the training of other models.

3.3 Reconstruction of the estimated curves

Once the principal functional components are applied to the original data set and the forecasting models have estimated the time series of coefficients. The estimated EURIBOR yield curves are then reconstructed so that the model can be validated by comparing them with the original curves. The error is calculated in both the VAL and TR period to compare the orders of magnitude.

The European Central Bank publishes the values of interest rates at different periods of time, adjusting them to two decimal places. In other words, the interval between two values of different days and the same maturity is at least 0.01. Thus, once the curves expressed in the base of the FPC have been reconstructed, a rounding stage is carried out, in which the functional data which have just been calculated are given the original format of the curves published by the BCE. The following Figure 10 shows the set of reconstructed functional data before and after the rounding stage. It should be mentioned that this process helps to improve the accuracy and robustness of the forecasting model

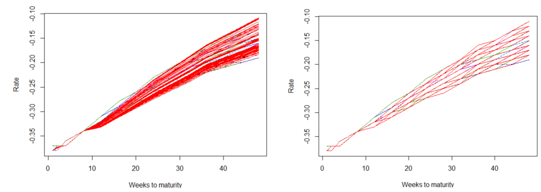

Figure 10: Original data set (colour curves) and reconstructed curves (red curves) before rounding (left) and after rounding (right)

In the low volatility period the ARIMA model is trained and the curves are reconstructed with the estimates of the coefficients. To determine the efficiency of the model, the error metrics mentioned in Section 3.1.1 are obtained. Being these the Functional Mean Square Error (FRMSE) (2), the Functional Mean Absolute Error (FMAE) (3) and the Functional Mean Absolute Percentage Error (FMAPE) (4) of the period TR and VAL. Figure 11 shows how the mean error of the model decreases as more principal components are incorporated into the model.

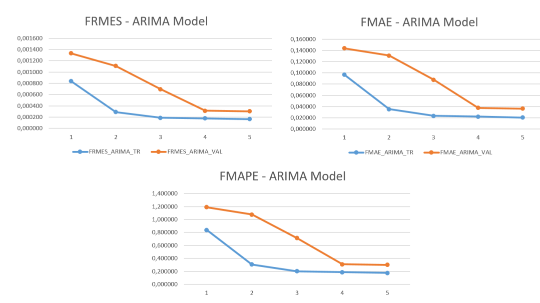

Figure 11: FRMSE, FMAE and FMAPE error of the ARIMA model in the TR and VAL period using a different number of principal functional components.

3.4 Exogenous variables

The information provided by the EURIBOR curves are the reference interest rates for interbank loans, generated daily by the European Central Bank, as an estimate of the number of transactions to be carried out by the main financial institutions. That is why, it could be interesting to include certain exogenous variables, which characterise aspects of these main financial institutions in the Eurozone and the socio-economic environment.

The exogenous variables to be analysed in this project, shown in Figure 12, are the values of the shares of certain European financial institutions: Santander (Spain), BNP (France), Intesa (Italy), ING (Netherlands), BBVA (Spain), Credit Agricole (France), Deutsche Bank (Germany), HSBC (United Kingdom), BPCE Group (France), UBS (Switzerland), Barclays (United Kingdom), Société Générale (France), Crédit Suisse (Switzerland), The Royal Bank of Scotland (United Kingdom), Nordea Bank (Finland), Commerzbank (Germany), Danske Bank (Denmark), Unicredit (Italy), Allianze (France), Zurich Secures (Switzerland), Danske Bank (Denmark) And the values of the stock exchange of indexes of different countries, being these: IBEX 35 (Spain), DAX30 (Germany), FTSE100 (United Kingdom), CAC40 (France), AEX25 (Netherlands), S&P500 and S&P100 (United States) and SHSZ300 (China).

In this way, it is proposed to analyse in a first instance the correlation that these variables have with the time series of the coefficients, associated with the different principal components used by the model. This analysis aims at determining whether the information provided by these stock market indices is linearly related to any aspect of the EURIBOR data set, in order to include them in the analysis model and improve its robustness. A correlation matrix is obtained for this purpose, where the linear relationship between each of the variables and the rest of them is determined. Analysing the column or row associated to the output of the system, being y the temporal series of the FPC coefficients, the percentage of the information of the exogenous variables is related to the variable of interest y is obtained.

Figure 13 shows the correlation matrix of the set of exogenous variables selected for the model and the coefficients associated with the first FPC. This ma-

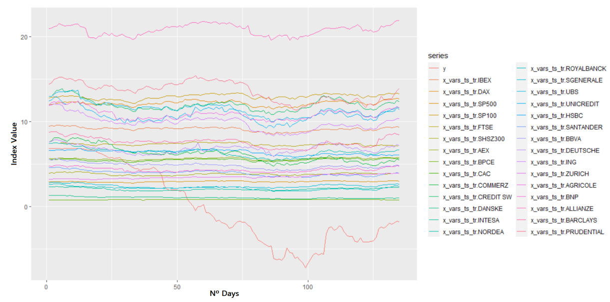

Figure 12: Display of all the exogenous variables to be analysed, calibrated in the same order of magnitude.

trix indicates that the variations of the selected stock market indices are related to the level of the EURIBOR curves, whether the average of each curve rises or falls (since this is the information that characterises the first principal component).

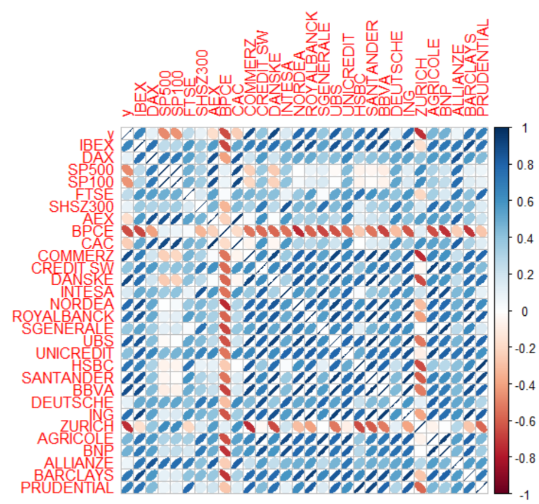

Figure 13: Correlation matrix between exogenous variables and the time series of coefficients associated with the first principal components, denoted by y

The following principal components provide in-depth information on the dynamics of the functional data set, making it more unlikely that exogenous variables have a strong correlation with them. This hypothesis is tested by obtaining correlation matrices in which all exogenous variables and the coefficients for the second and third principal component are included. In Figure 14 shows the correlation matrix for the second, third and fourth FPCs. As expected, the correlation percentages are not very high, so the exogenous variables are not very representative in the analysis models of the coefficient series, with the exception of the second principal component in which certain variables could explain part of the information of this time series.

In this way, exogenous variables (those that have a certain correlation with y) will be used in the analysis models of the first two FPCs. The rest of the time series of coefficients will be modelled with ARIMA autoregressive techniques, without including the exogenous variables.

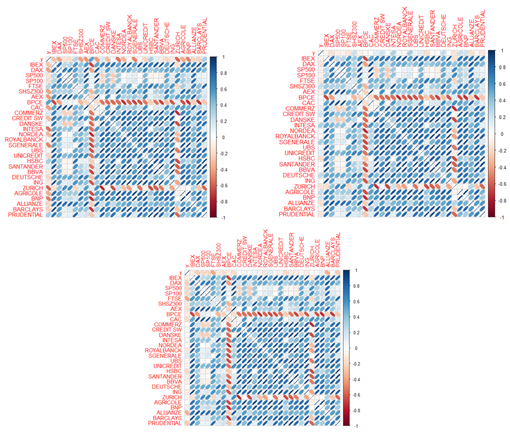

Figure 14: Correlation matrix between exogenous variables and the time series of coefficients associated with the second, third and fourth principal components.

3.5 Period of high volatility

The period used for the previous analysis, from November 2016 to February 2019, is very stable. This first assumption was made to simplify the modelling of the functional data. By applying the models to high-volatility period, the real-life effectiveness of the models can be measured. The new dataset falls within the range of *15-05-2019* and *17-4-2020*.

As expected, the *Naive* model has a higher error, this improvement is due to the fact that in this new set there is a greater distance between the t and $t+1$ interest rates. In addition, the models used work much more efficiently by increasing the volatility of the analysis data set. Figures 15 and 16 show the results of the error rates using as references the FRMSE, FMAE and FMAPE, respectively in the TR and VAL period. The graphs show the error value on the y-axis and the number of selected principal components on the x-axis. It is worth mentioning that in these results only six functional principal components have been used, as in Section 3.1.1 it is determined that from the sixth component the *Naive* model is exceeded therefore, it is checked that the models operate according to the expected limits.

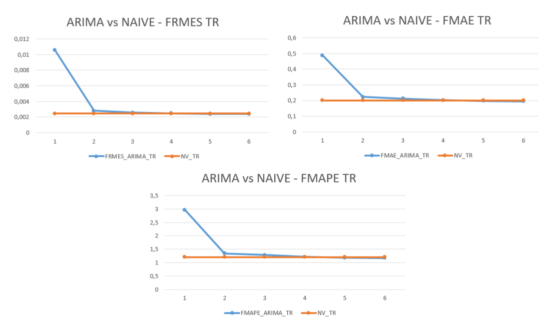

Figure 15: Comparison of the efficiency of the ARIMA and Naive models in the TR period, showing the errors FRMSE, FMAE and FMAPE respectively, in the period of high volatility.

In order to analyse in greater depth the development of the ARIMA model when estimating the values of the different interest rates for the necessary maturity times, the MAE error in basic points is shown in the following Figures 17 and 18.

For this period, the above-mentioned models TF 3.2.2 and MLP 3.2.3 have been trained, in which the exogenous variables of the 3.4 are included, for the

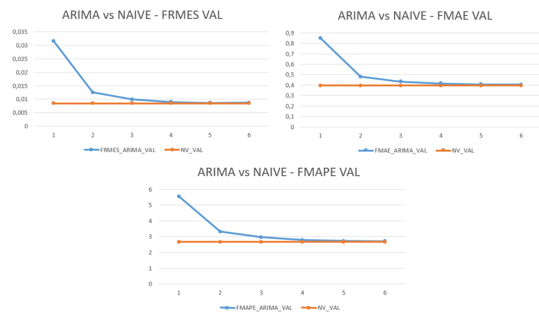

Figure 16: Comparison of the efficiency of the ARIMA and Naive models in the VAL period, showing the errors FRMSE, FMAE and FMAPE respectively, in the period of high volatility.

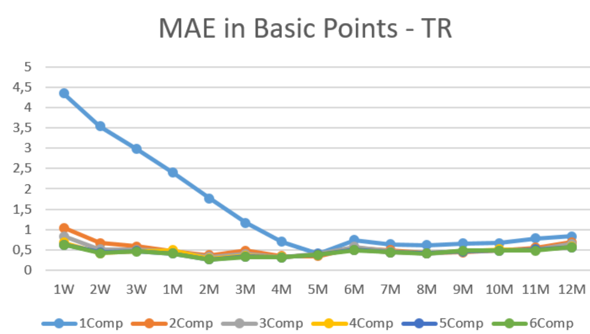

Figure 17: MAE error in basic points of the high volatility period TR, of the estimated results of the ARIMA model for the high volatility period, using up to six principal components

adjustment and estimation of the data set. In this way, the Figures 19 and 20 show the graphs of the error rates selected in this project when training the ARIMA, TF and MLP models. These graphs compare the efficiency of the models in the TR and VAL validation and training period, where the improvement of these multivariate models can be seen (despite not having carried out an exhaustive study).

Because there is a greater margin between functional data from one day to the next, the intrinsic benefit that can be obtained on the purchase and sale

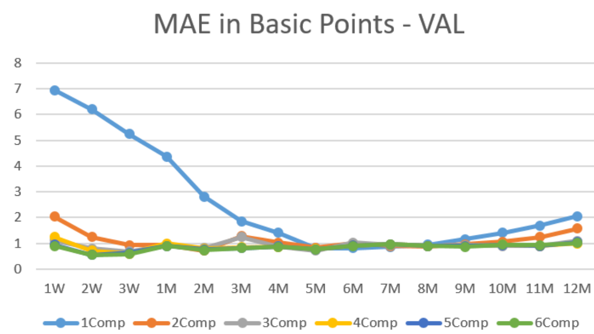

Figure 18: MAE error in basic points of the high volatility period VAL, of the estimated results of the ARIMA model for the high volatility period, using up to six principal components

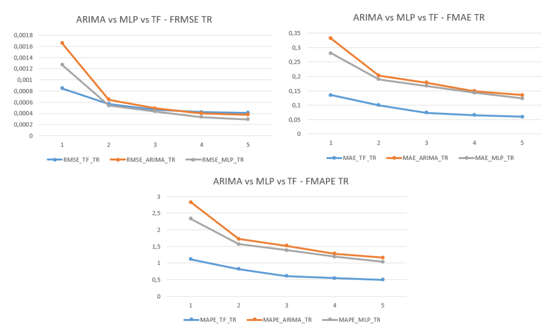

Figure 19: Comparison of the efficiency of the ARIMA, TF and MLP models in the TR period, showing the errors FRMSE, FMAE and FMAPE respectively, in the period of high volatility.

operations in financial products referenced to the EURIBOR is much greater, also providing a greater risk in the investment due to the natural rules of the stock market. However, as has been seen in this section, the more complex models produce a lower forecasting error.

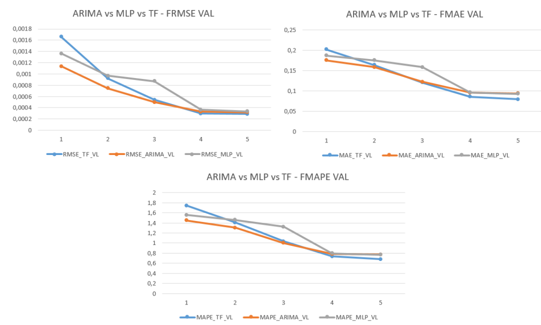

Figure 20: Comparison of the efficiency of the ARIMA, TF and MLP models in the VAL period, showing the errors FRMSE, FMAE and FMAPE respectively, in the period of high volatility.

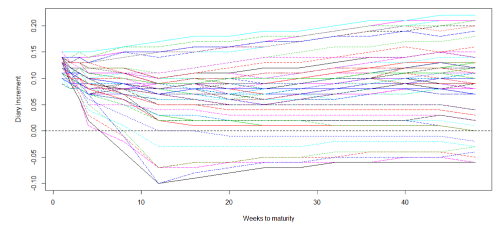

Figure 21: A set of functional data, which explains the margin between successive days. Weeks of interest rate maturation (x-axis) and operating margin (y-axis).

4 Applications to financial markets

Once the Forecasting models have been introduced, it is necessary to illustrate the contribution of the model to a real world application. One of the most interesting areas that could benefit from the results obtained by the proposed models is to serve as a support for financial *traders*. Who work with derivatives directly related to the EURIBOR curves or the interest rate at a specific maturity period. In this way, offering an estimate for the future can serve as a powerful reference when carrying out purchase and sale operations.

In this way, the efficiency of the forecasts has been studied in terms of the profit margin with which a daily trader could work. To this end, the variation of each of the interest rates is determined in the first instance, for the different periods of maturity observed with the real data. This process generates a new set of data, which represents the margins between consecutive curves. Figure 21 shows a visual example of this new set.

Another set of data is then generated in a similar way by calculating the difference between the original curves in t and the curves estimated by the predictive model in $t+1$. In this way, by comparing the two sets of data, it is possible to determine the effectiveness of the models. It has been decided to use two metrics to characterize the efficiency of the predictions as a function of the operating margin. The first one, reflected in the graphs of the Figure 22, the average hit rate $\left(\sum_{t=1}^{T} 100 * \frac{|P_i(t+1)|}{|R_i(t)|}\right) * T^{-1}$ is determined for each of the observed i maturity periods, where $P(t)$ is the set that characterizes the margin between model estimates and actual data and $R(t)$ is the set calculated with actual data only. This information is shown in color-coded graphs, with red being the lowest percentage and yellow the highest percentage.

The second metric used to determine the efficiency of this application on trader operations, is through the average absolute error in basic points, comparing the set formed by the original data and the estimated data. In the figure 23 the graphs that relate the error MAE with the total margin for each one of the observed points are shown.

The results shown in these graphs will be used to determine which period of maturity is the most suitable to carry out a buy-sell *trader* operation, taking into account the compromise between the highest profit margin (y-axis) and the efficiency of the predictive modeller (colour code of the legend), in the case of the percentage of success the highest percentage and in the error MAE the one with less error. It is worth mentioning that as it can be seen in the graphs the higher maturity times have a higher variability, so

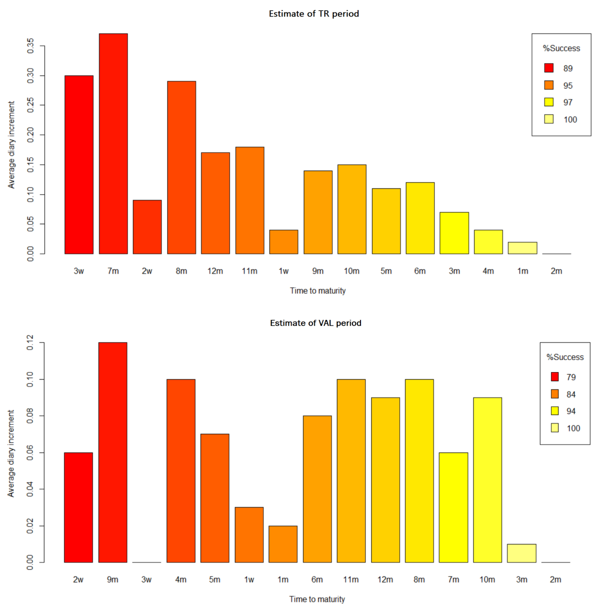

Figure 22: Graphs (upper TR period and lower VAL period) showing: the average interval, the variation of the interest rate from one day to another (y axis), the observed maturity period (x axis) and the success rate of the predictive model denoted by the colour code shown in the legend.

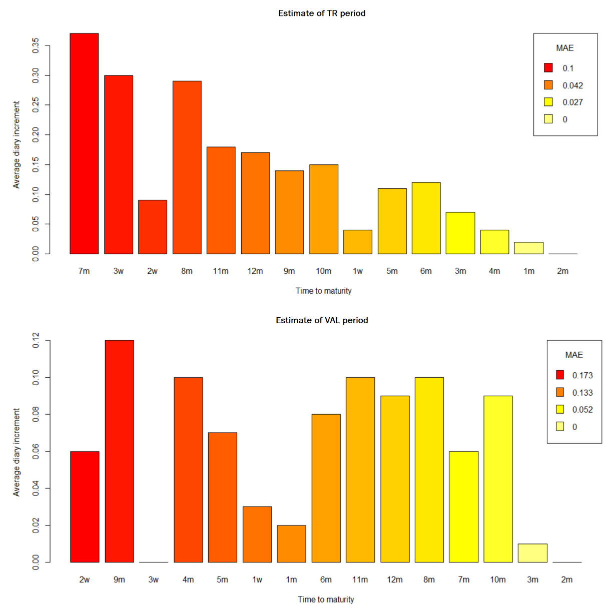

Figure 23: Graphs (upper TR period and lower VAL) showing: the average interval of the variation of the interest rate from one day to another (y-axis), the observed maturity period (x-axis) and the MAE error made by the predictive model when estimating the variation interval (arranged on the y-axis) denoted by the colour code shown in the legend.

they will offer a bigger margin between one day and another (being the objective of the trader operations) but also due to this volatility it makes the estimations less reliable, so the efficiency metrics used decrease. In any case the results shown are very favourable, due to the high performance of the models, so it is concluded that this project is very beneficial as support for *trading* operations.

5 Conclusions

The feasibility of using functional models on financial data sets has been tested in this project and some optimistic conclusions have been reached.

By applying the decomposition into functional principal components, the reduction of work space is effectively generated, in order to work in a comfortable scenario. This methodology generates an error called *Reconstruction Error*, which marks the limit of the model's improvement. It also offers a reference to determine the number of FPC to be included in the optimal model.

In the bibliography, it is observed that one of the most used methods once the decomposition in functional principal components has been made is the ARIMA model. In this way, we proceed to reproduce this process by training this autoregressive model, generating a starting point in the estimates of the financial data set.

As an innovation for the project, stock market values related to the EURIBOR curves have been used. For this purpose, the correlation of financial products with the series of coefficients of the principal components is analysed and the most correlated series are selected. The methodologies applied with the set of exogenous variables have been carried out with the intention of studying the viability of this proposal, so a deep study will be necessary in future works.

The models used to adjust the exogenous variables and the different series of coefficient have been the *Transfer Function* and *Multi-Layer Perceptron* models. The results of the errors obtained show that these models generate more robust estimates. Therefore, it is concluded that the FDA models on financial products work more efficiently by including exogenous variables and also perform better in a high volatility period.

Finally, a possible application of this project in the real world of finance has been developed. Being a support for the operations *traders* of purchase-sale. Very optimistic results are shown in the predictions of the EURIBOR curves.

References

[1] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo. Stock price prediction using the ARIMA model. *Proceedings - UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014*, pages 106–112, 2014.

[2] O. Blaskowitz and H. Herwartz. Adaptive forecasting of the EURIBOR swap term structure. *Journal of Forecasting*, 28(7):575–594, 2009.

[3] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time series analysis: forecasting and control*, volume 734. John Wiley & Sons, 2011.

[4] F. X. Diebold and C. Li. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364, 2006.

[5] Y. Guan, S. Wang, and Z. Zhou. Research on the prediction model of coal and gas outburst based on principal component analysis and radial basis function neural network. *Proceedings - 2014 4th International Conference on Instrumentation and Measurement, Computer, Communication and Control, IMCCC 2014*, pages 260–264, 2014.

[6] S. Hörmann, L. Horváth, and R. Reeder. A functional version of the arch model. *Econometric Theory*, 29(2):267–288, 2013.

[7] L. Horváth, Z. Liu, G. Rice, and S. Wang. A functional time series analysis of forward curves derived from commodity futures. *International Journal of Forecasting*, 36(2):646–665, 2020.

[8] R. J. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts: Melbourne, Australia., 2nd edition, 2018.

[9] W. Jiesheng. Intelligent data pretreatment based on principal component analysis and fuzzy C-means clustering in flotation process. *Proceedings of the 26th Chinese Control Conference, CCC 2007*, pages 409–412, 2007.

[10] V. Kargin and A. Onatski. Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, 99(10):2508–2526, 2008.

[11] P. Kokoszka and M. Reimherr. *Introduction to Functional Data Analiysis*, volume 3. Taylor & Francis Group, 2017.

[12] H. G. Müller, R. Sen, and U. Stadtmüller. Functional data analysis for volatility. *Journal of Econometrics*, 165(2):233–245, 2011.

[13] C. R. Nelson and A. F. Siegel. Parsimonious Modeling of Yield Curves. *The Journal of Business*, 60(4):473, 1987.

[14] A. Pankratz. *Forecasting with dynamic regression models*, volume 8. John Wiley & Sons, 1992.

[15] G. Rice, T. Wirjanto, and Y. Zhao. Forecasting value at risk with intra-day return curves. *International Journal of Forecasting*, 36(3):1023–1038, 2020.

[16] R. Sen and C. Klüppelberg. Time series of functional data with application to yield curves. *Applied Stochastic Models in Business and Industry*, 35(4):1028–1043, 2019.

[17] H. L. Shang, Y. Yang, and F. Kearney. Intraday forecasts of a volatility index: functional time series methods with dynamic updating. *Annals of Operations Research*, 282(1-2):331–354, 2019.