Outlier Detection in Functional Time Series

Catarina Padrela Loureiro
catarinapadrela@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

December 2019

Abstract

In recent years methods for representing data through functions or curves have gained attention. Such data are known as functional data. Typical data sets consist of time series and cross-sectional data, such as time series of stock prices. However, the presence of outliers has adverse effects on the modelling and forecasting of functional data. In this context, outliers detection is not an easy task since the whole set of functional data is not always possible to visualise. Nevertheless, procedures for detecting functional outliers have been proposed over recent years. Some of these procedures are based on Functional Principal Components Analysis (FPCA), which assumes each sample curve is drawn from an independent and identical distribution (i.i.d.). This assumption is inconsistent with financial data, where samples are often interlinked by an underlying temporal dynamic process. These dynamics should be taken into account to detect outliers in functional time series. To overcome this situation, a dynamic extension of FPCA that takes into account the dependence structure of the functional data has been proposed. Taking this into consideration, this work introduces an outlier detection method for functional time series based on these dynamic principal components. This technique is applied to anomaly detection in financial data.

Keywords: Functional data, functional time series, outlier detection, dynamic functional principal components, financial data

1. Introduction

With the current technological evolution, it has become easier to obtain uninterrupted records of numerous types of data. This kind of data, which is observed over a continuous measure, is called functional data. The label "functional" arises from the fact that this data seem to have an underlying function or smooth curve that originated them. The term Functional Data Analysis (FDA) was first introduced by Ramsay (1982) and Ramsay and Dalzell (1991). Ramsay and Dalzell (1991) also stated some of the advantages of FDA: a set of finite observations can be put through a smoothing process in order to obtain a functional approximation; some problems can be more naturally modelled if looked at from a functional perspective; derivatives of the observed functions could provide additional information, regarding for instance data visualisation. Therefore, time series data are a natural candidate to be treated as functional. In particular, this work will focus on financial time series consisting of stock prices.

Ramsay and Silverman (2005) comprise a detailed study on FDA, with the application of several multivariate data analysis methods adapted to the functional scenario. Moreover, Ramsay and Silverman (2002) show how FDA can be applied to several data sets from different fields of study. One of the first multivariate data analysis techniques to be adjusted to FDA was Principal Component Analysis (PCA) by Dauxois et al. (1982). Wang et al. (2016) offers a thorough review of FDA, including some fundamental methods, while Shang (2011) surveys Functional Principal Component Analysis (FPCA). On the computational front, Ramsay et al. (2009) present several practical applications using R and MATLAB, which is accompanied by the FDA website (Ramsay, 2013).

As with any kind of data the presence of outlying observations has severe effects on forecasting and modelling of functional data. Thus, the first step on any descriptive analysis of the data set should be outliers detection, before any modelling or prediction method. In the functional context, however, this can be a very complex task due to the nature of the data and the difficulty in visualising the entire data (curves, images or functions). Nonetheless, their analysis in functional data has been seldom addressed with several methods based on distances and FPCA. In fact, the first efforts consisted of distance based methods, using depth measures (Febrero et al., 2007, 2008). Hyndman and Shang

(2009); Arribas-Gil and Romo (2014); Tarabelloni (2017), on the other hand, put forward outlier detection methods based on graphical tools. Furthermore, Sawant et al. (2012) approaches the detection of outliers by applying robust PCA techniques to a coefficient matrix constructed from the coefficients in the expansion given by (2). Another way to deal with outlying observations is through the robustification of the methods. In that sense, Hyndman and Ullah (2007) used a projection-pursuit approach to obtain a robust FPCA, while Gervini (2008) proposed spherical principal components. Robust functional principal components estimators with a projection-pursuit approach were also proposed by Bali et al. (2011). Boente and Barrera (2015) suggested S-estimators for functional principal components. Vilar et al. (2016) proposes two methods, one based on projections and another based on the residual of the original curve with an uncontaminated version of each curve, using robust FPCA and then applying time series outlier detection.

However, one handicap of FPCA is that it assumes the sample curves are independent with each other. In financial data, samples are most likely interlinked by an underlying temporal dynamic process, leading to an inconsistency with the assumption of FPCA. Additionally, financial and economic data can usually be seen as a Functional Time Series (FTS), considering they comprise a set of curves registered continuously over time. Therefore, in order to properly detect outliers in FTS, the dynamics and dependence structure in the data must be taken into account. Dynamic Functional Principal Component Analysis (DFPCA) was recently proposed by Hörmann et al. (2015) and takes these dependencies into consideration. Nonetheless, although the DFPCA is a useful tool it has not been used for outliers detection yet. Gao et al. (2018) employed DFPCA in forecasting high-dimensional functional time series.

Furthermore, intraday stock price curves are one of the most natural and obvious functional data. The functional perspective enables the analysis of the information present in the shapes of these curves, which reflects the reactions and expectations of intraday investors. Forecasting of the stock prices is also very important for the investors in the stock exchange markets. On that account, outlier detection becomes of the utmost importance in dealing with financial data.

This work aims to analyse a financial data set from a functional perspective. The main objective is to compare several outliers detection methods among each other and with external social-economic events that could have caused the anomalies. Moreover, since most outliers detection methods are based on FPCA, an important aspect of the analysis is obtaining the principal components. Nevertheless, since DFPCA is better suited to financial data sets, an adapted version of the outlier detection method based on projections introduced by Vilar et al. (2016) is proposed in this work.

2. Background

2.1. Functional Data

The one constant characteristic underlying functional data is that they are comprised of functions or there is a function that originates the observed data. Thus, the observed data functions are considered as one entity. Moreover, functional data are continuously defined, regardless of in practice the observed data being typically discrete. The FDA approach is useful in a parsimonious representation of the data by taking advantage of their smoothness. Instead of looking at a function as a dense vector of values, it can often be represented in an linear combination of a handful of (well–chosen) basis functions. The first step is to define the space where FDA operates, which is not $\mathbb{R}^n$ as in multivariate analysis. The Hilbert space of square integrable absolutely continuous functions defined on a compact domain, $[a, b]$, called $L^2_{[a,b]}$, is considered instead. This choice allows to take advantage of many properties of projection and distance inherent to a Hilbert space. From the construction of the space, it follows that a random variable X is a function such that $X : (\Omega, \Sigma, \mathbb{P}) \to L^2_{[a,b]}$, where $(\Omega, \Sigma, \mathbb{P})$ is a probability space. A functional random variable X is a real-valued function $X(t)$ considered as an element of the Hilbert space $L^2_{[a,b]}$, satisfying $\int_a^b X^2(t)dt < \infty$. From now on, it is considered that an integral sign without the limits of integration denotes the integral over the whole interval of the domain where the functions are defined.

Being able to measure distances between functional data is of most importance. In the L^2 Hilbert space, the following inner product $\langle \cdot, \cdot \rangle$ is defined for the functional variables X, Y, as $\langle X, Y \rangle = \int X(v)Y(v)dv$. The inner product generates the following L^2 norm $\|X\| = \langle X, X \rangle$. Assume that X has mean function $\mu(t) = \mathbb{E}(X(t))$ and covariance function $\mathrm{Cov}(X(t), X(s)) = \gamma(t, s) = \mathbb{E}\left[(X(t) - \mu(t))(X(s) - \mu(s))\right]$. The corresponding positive and semi-defined covariance operator $\Gamma(.) = \mathbb{E}[\langle (X - \mu), . \rangle (X - \mu)]$, $\Gamma : L^2_{[a,b]} \to L^2_{[a,b]}$, is such that $\Gamma(x)(t) = \int \gamma(t, s)x(s)ds$ and satisfies $\int\int \gamma^2(t, s)dtds < \infty$. Consequently, Γ is a Hilbert–Schmidt operator. The operator Γ has a countable number of positive real eigenvalues. The Mercer's theorem (e.g. Kokoszka and Reimherr, 2017, Chapter 10) implies that the spectral decomposition of Γ, or equivalently of $\gamma(t, s)$, leads to $\gamma(t, s) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(t)\varphi_i(s)$, with uniform convergence, where λ_i are the eigenvalues (in descending order) and φ_i the corresponding orthogonal eigenfunctions.

In practice the observed data is discrete, which is then used to estimate the function and some of its derivatives. Consider the observed data vector as $\boldsymbol{y} = (y_1, \dots, y_n)$, where each y_j is a "snapshot" of the underlying function x. These can be expressed as,

$$y_j = x(t_j) + \epsilon_j, \tag{1}$$

where ϵ_j is the noise or error contributing with roughness to the data, and $t_j \in \mathcal{T}$ the interval over which data is collected. The standard model considers ϵ_j are assumed to be independently distributed with mean zero and constant variance σ^2.

Functions can be represented as a linear combination of basis functions. The use of these is a computational trick that allows to fit hundreds of thousands of data points, providing flexibility and the possibility of performing calculations resorting to the well-known matrix algebra. In practice, working with an infinite basis is not feasible and the usual approach is to truncate the number of basis functions to a finite K. So, provided that K is sufficiently large, a function x can be represented by a basis-expansion,

$$x(t) \approx \sum_{k=1}^{K} c_k \phi_k(t), \tag{2}$$

where ϕ_k are known basis functions that are mathematically independent of each other and c_k are the coefficients or coordinates that represent each observation in the selected basis.

Accordingly, smoothing can be accomplished either using data-driven basis functions or known basis functions. The latter are, for example, the Fourier basis functions which are better suited for periodic data, or even B-splines for data without any strong cyclic variation.

In fact, smoothing with a roughness penalty is an efficient method for fitting a function to discrete data. The goal is reducing noise in the measurements by choosing c_k's that minimise,

$$\text{PENSSE} = \sum_{j=1}^{n} (y_j - x(t_j))^2 + \lambda \int [x''(t)]^2 dt, \tag{3}$$

where λ is a smoothing parameter that controls the trading ratio between the fit to the data (1st term) and the function's variability (2nd term). λ is usually selected using Generalized Cross Validaion (GCV).

Assuming that x has a second derivative and that t_j, $j = 1, \dots, n$, are distinct, the function x that minimizes PENSSE is a cubic spline with a knot at each data point t_j. The knots' placement will lead to an exploitation of areas where there is a great amount of observations and to substantial smoothness where there are few. Hereupon, cubic spline smoothing is the most usual computational method for spline smoothing.

2.2. Functional Principal Components Analysis

The FPCA is used to reduce the dimension of infinitely dimensional functional data to a small finite dimension and it can be very helpful at retrieving information from the covariance structure. The FPCA can be seen in two ways, as coordinates maximizing variability, or as an optimal orthonormal basis (empirical bases). Assuming that the sample mean function values have been subtracted from the observed $x_i(t)$, the sample covariance function is defined by,

$$\upsilon(t, s) = N^{-1} \sum_{i=1}^{N} x_i(t) x_i(s), \tag{4}$$

which leads to the sample covariance operator,

$$V(x)(t) = \int \upsilon(t, s) x(s) ds. \tag{5}$$

The eigenanalysis of V results in the pairs of eigenvalues/eigenfunction, $(\rho_i, \xi_i(\cdot))$, by the Mercer's theorem. The eigenvalues are order descending, $\rho_i \geqslant \rho_{i+1}$, and the eigenfunctions, $\xi_i(\cdot)$s, correspond to the sample functional principal components. Moreover, the sample principal component scores are,

$$f_i = \langle \xi, x_i \rangle = \int \xi(s) x_i(s) ds. \tag{6}$$

Hence, x_i can be approximated through the Karhunen-Loève expansion (e.g. Kokoszka and Reimherr, 2017, Chapter 11),

$$x_i(t) \approx \sum_{k=1}^{K} f_{ik} \xi_k(t). \tag{7}$$

2.3. Functional Time Series

In many practical situations functions are naturally ordered in time. For example, when dealing with daily observations of the stock market. A Functional Time Series (FTS) is a sequence of curves $(X_t(u) : t \in \mathbb{Z}, u \in \mathcal{T})$, where t is a discrete parameter and u is a continuous one. In fact, t is the time index, which typically refers to day or year, and u is the time within that unit. For example, when dealing with daily observations, t refers to the day and u is the intra-day time parameter. Thus, the main idea behind FTS is that the time record can be split into natural intervals, treating the curve within each interval as a unit.

Many procedures assume the stationarity of FTS. In that sense, the null hypothesis that the series is stationary can be written as, for any h and any t,

$$(X_{1+h}, \ldots, X_{t+h}) \overset{d}{=} (X_1, \ldots, X_t). \tag{8}$$

Horváth et al. (2014) introduces tests which are nontrivial extensions of the KPSS family tests. Two classes of tests are considered, one based on the curves and another based on the finite dimensional projections of the curves on the functional principal components.

A common example of FTS are intraday price curves. In that setting, the price of a financial asset at time t_j on day n is defined as $P_n(t_j)$, $n = 1, \ldots, N$, $j = 1, \ldots, m$. The tests when applied to this kind of data and to sufficiently long periods of time reject the stationarity. For shorter periods of time, the rejection of stationarity does not always happen. Therefore, a transformation should be employed to deal with the non-stationarity, resulting in the Cumulative Intraday Returns (CIDR), which were first introduced by Gabrys et al. (2010) and can be defined as,

$$R_n(t_j) = 100 \left[\ln P_n(t_j) - \ln P_n(t_1) \right]. \tag{9}$$

CIDR, however, are not directly comparable to daily returns, since they disregard the overnight price change. CIDR do enable the statistical analysis of the intraday price curves' shapes.

2.4. Dynamic Functional Principal Components Analysis

The FPCA presented above is considered "static", since it does not take into account possible dependencies between the several observations or repetitions of the function. In order to deal with this shortcoming, Hörmann et al. (2015) proposed a "dynamic" alternative to FPCA. Considering a stationary FTS, $(X_t(u), t \in \mathbb{Z})$, let the auto-covariance function at lag h and the long-run covariance function be given respectively by,

$$\gamma_h(u, s) = \operatorname{Cov}(X_t(u), X_{t+h}(s)), \tag{10}$$

$$\gamma(u, s) = \sum_{h=-\infty}^{\infty} \gamma_h(u, s). \tag{11}$$

This leads to the definition of the operator,

$$\Gamma(x)(u) = \int \gamma(u, s) x(s) ds, \tag{12}$$

which admits an eigendecomposition,

$$\Gamma(x) = \sum_{k=1}^{\infty} \lambda_k \left(\int x(u) \varphi_k(u) du \right) \varphi_k, \tag{13}$$

where λ_k, $k \geqslant 1$, are the eigenvalues (in descending order) and φ_k, $k \geqslant 1$, its corresponding eigenfunctions.

$X_t(u)$ can be approximated using the Karhunen-Loève expansion, making use of the first K principal components obtained,

$$X_t(u) \approx \sum_{k=1}^{K} \beta_{t,k} \varphi_k(u), \tag{14}$$

where the k-th principal component score at time t is given by,

$$\beta_{t,k} = \int X_t(u) \varphi_k(u) du. \tag{15}$$

Therefore, the observed FTS, $x_t(u)$, can be approximated using the empirical principal components. The DFPCA is implemented in the R package `freqdom.fda`.

2.5. Robust Functional Principal Components Analysis

Hyndman and Ullah (2007) introduce a two-step algorithm for Robust Functional Principal Components Analysis (RFPCA), which combines the weighted principal component method and the RAPCA projection pursuit algorithm (Hubert et al., 2002). The point is to find the functions $\varphi_k(u)$ which maximize the variance of the new scores,

$$z_{t,k} = w_t \int \varphi_k(u) X_t(u) du, \tag{16}$$

subject to,

$$\int \varphi_k^2(u) du = 1, \quad \int \varphi_k(u) \varphi_{k-1}(u) du = 0 \text{ if } k \geqslant 2.$$

The weights w_t are given by,

$$w_t = \begin{cases} 1 & \text{if } \upsilon_t < s + \lambda \sqrt{s} \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

where s is the median of $\{\upsilon_1, \ldots, \upsilon_t\}$ and $\lambda > 0$ is a tuning parameter that controls the level of robustness (the smaller λ is, the greater the number of curves deemed as outliers). Additionally, υ_t is the integrated squared error given by,

$$\upsilon_t(u) = \int \left(X_t(u) - \sum_{k=1}^{K} \beta_{t,k} \varphi_k(u) \right)^2 du, \tag{18}$$

where K is the number of components, φ_k are the principal component functions, and $\beta_{i,k}$ their scores. Thus, the two-step algorithm follows,

1. Obtain initial values for $\beta_{t,k}$ and $\varphi_k(u)$, using the RAPCA algorithm and assuming equal weights w_t;
2. Update the values of w_t according to (17), and use them to update $\beta_{t,k}$ and $\varphi_k(u)$ through the weighted principal component method (Hyndman and Ullah, 2007, Section 3.1).

This algorithm has been implemented in the R package `ftsa`.

3. Outlier Detection in Functional Data

Data exploration which aims to reveal some aspects of the dynamics of the underlying process is an important task. In that sense, outlier detection plays a crucial part, considering outliers can have severe adverse effects on the modelling and forecasting of the data. There are two types of outlying curves: "magnitude outliers", curves that lie outside the range of the rest of the data; "shape outliers", curves that have a different shape of the rest of the data even though they are within their range; they can also be a combination of these two types.

3.1. Methods Based on Distances

The earliest attempt at outlier detection was based on the idea of distance, with Febrero et al. (2007) using a likelihood ratio test statistic. Febrero et al. (2008) proposes another method based on functional depths. Functional depths measure the centrality of a curve within a sample of curves. In that sense, depths allow to order the sample curves from the centre outwards. Thus, depth and outlyingness are inverse notions, which means magnitude outliers are expected to have a significantly low depth.

Hyndman and Ullah (2007) advance another alternative which is a product of the robust functional principal components algorithm introduced above. It is built on the integrated squared error given by (18), which is used to update the weights w_t (17). Consequently, a curve is considered an outlier if its associated weight w_t is zero. The higher the integrated squared error is, the most likely a curve can be considered an outlier.

Another distance to be taken into account is the familiar robust Mahalanobis distance. If the functional data are recorded on equally spaced points, it can be looked at as multivariate data and then applied the robust Mahalanobis distance.

All these distance based methods are implemented in the R package `rainbow`.

3.2. Methods Based on Graphical Tools

The difficulty in visualising the whole set of functional curves can complicate the detection of outliers. As such, Hyndman and Shang (2009) proposed two graphical methods that can identify outliers: the functional bagplot and the Highest Density Region (HDR) boxplot. These methods are available in the R package `rainbow`.

The functional bagplot shows the median curve (curve with the greatest depth), inner region (region enclosed by all curves corresponding to points in the bivariate bag and containing 50% of the curves) and outer region (region confined by all curves corresponding to points within the bivariate fence region). Thus, the functional bagplot is obtained by mapping the bagplot of the first two robust principal component scores to the functional curves. Although this method can detect outliers when they are distant from the median, it can misidentify outliers near the median. The functional HDR is more suitable to deal with those.

The functional HDR is obtained by computing the bivariate kernel density estimate on the first two robust principal component scores, and then applying the bivariate HDR boxplot. The functional HDR shows the modal curve (curve with the highest density), inner region (region limited by all curves corresponding to points inside the 50% bivariate HDR, containing 50% of the curves) and outer region (region limited by all curves corresponding to points within the outer bivariate HDR). Consequently, curves outside the outer HDR are considered outliers.

Still within the graphical domain, Arribas-Gil and Romo (2014) introduce the outliergram, which helps identify shape outliers by taking advantage of the relation between two measures of depth for functional data. These two depth measures consist of the Modified Band Depth (MBD) and the Modified Epigraph Index (MEI). They offer an idea of how central or deep a curve is with respect to a sample of curves. Hence, the outliergram consists of mapping the (MEI, MBD) points.

Tarabelloni (2017) also proposes a robust adjusted functional boxplot, this one better suited to the detection of magnitude outliers. To construct the functional boxplot it is necessary to use a depth measure to rank the functions from the centre of the distribution outwards, and then defining a central region. Both the outliergram and this functional boxplot are available in the R package `roahd`.

3.3. Method Based on Projections

More recently, Vilar et al. (2016) introduced two algorithms that detect outliers in Functional Time Series, which take advantage of FPCA. The following algorithm consists of an adapted version of the method based on projections,

1. Apply FPCA and obtain the time series of principal components scores $\{\beta_{t,1} \ldots, \beta_{t,K}\}_{t=1}^{n}$.
2. Apply a time series outlier detection method on the series obtained.
3. Define the set of outliers as, $\mathcal{O} = \{X_t : t \in \mathcal{I}\}$, where $\mathcal{I} = \{t : (\beta_{t,1} \ldots, \beta_{t,K})$ was identified as outlier$\}$.

This algorithm allows to combine the power of FPCA with methods that detect outliers in time series. The projections $\beta_{t,k}$ display the most prominent characteristics of the data. As such, one of these projections is considered an outlier, only if its originating curve is also an outlier.

In the implementation of this algorithm it is necessary to specify the type of FPCA and a time series outliers detection method. In this work, it will be tested the use of static, dynamic and robust FPCA. It is also required to choose the threshold parameter K (number of principal components to retain). According to the sensitivity studies performed by Vilar et al. (2016), it is recommended to select K so that at least 98%/99% of the variability is explained. This value is so high because this method uses the scores and not the original curves. Moreover, the first scores are related to the presence of magnitude outliers, while the scores of higher order are related to the presence of shape outliers.

4. Application to a Financial Data Set

This data set is composed by the closing price of Banco Comercial Português (BCP) stocks, from 1989 to 2018, as represented in Figure 1. BCP trades on Euronext Lisbon, being the largest contributor to the PSI-20 index. The objective here is to detect outliers using the methods introduced for FTS. In order to accomplish that, it is necessary to split the time series and create a FTS, as Figure 2 shows. This way, the observed FTS, $\{x_t(u)\}_{t=1}^{30}$, is composed by thirty curves each corresponding to a year of daily measurements (366 points per curve). Additionally, the stationarity tests proposed by Horváth et al. (2014), (R package `ftsa`), were performed on the time series. With a p-value of 0.23, the series can be considered as stationary for the usual significant levels (1%, 5%, 10%).

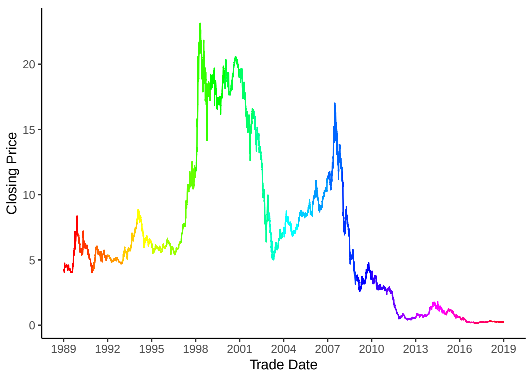

Figure 1: Time series of BCP stocks price.

Figure 3 shows the variance-covariance surface and Figure 4 the cross-correlation surface of the BCP data set. These were obtained using R package `fda`. It is possible to see that the highest variability occurs around April, while the lowest happens around October. From the cross-correlation plot it is possible to discern that correlation among closest days is higher, which makes sense considering these data are sequential in time.

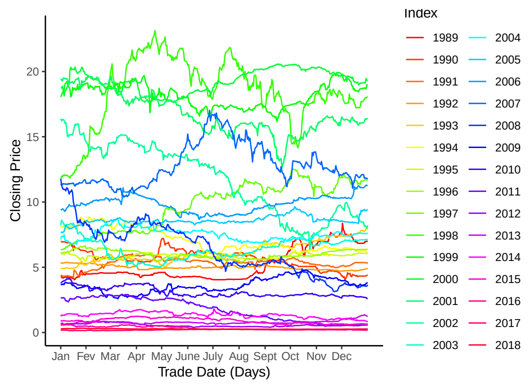

Figure 2: Yearly curves of BCP stocks price.

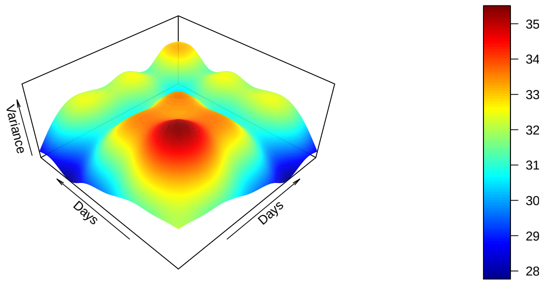

Figure 3: Variance-covariance surface.

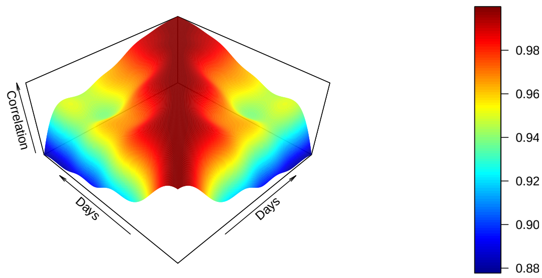

Figure 4: Cross-correlation surface.

Smoothing using a roughness penalty with cubic splines and knots at each day was performed on the data. The smoothing parameter was chosen by minimizing the GCV, resulting in $\lambda = 10^5$. The resulting smoothed data can be seen in Figure 5.

The graphical methods functional bagplot and HDR boxplot were obtained as shown in Figure 6 and 7 (R package `rainbow`). Moreover, the outliergram and functional boxplot were also computed and are represented in Figure 8 and 9 (R package `roahd`). The detected outliers by each of these methods are summarized in Table 1. The years 2002 and 2008 were identified as outliers by three of the methods. Economically speaking these years were marked by recession, so it is not unexpected that they are marked as outliers. The year 1997 was detected as an outlier by two of the methods. This year, on the other hand, saw some economic growth and development. It is also worth recalling that the functional boxplot is better suited to

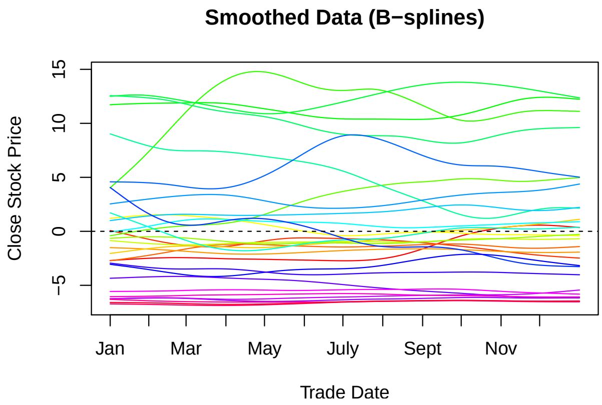

Figure 5: Smoothed data obtained using B-spline basis with the smoothing parameter $\lambda = 10^5$ (R package `fda`).

detect magnitude outliers. Additionally, the years 1998 − 2001 identified by this method were precisely years marked by economic growth, so it makes sense that the stock price curves are above the rest of the curves. The outliergram, however, is more apt in identifying shape outliers, being that the years selected correspond to periods of recession.

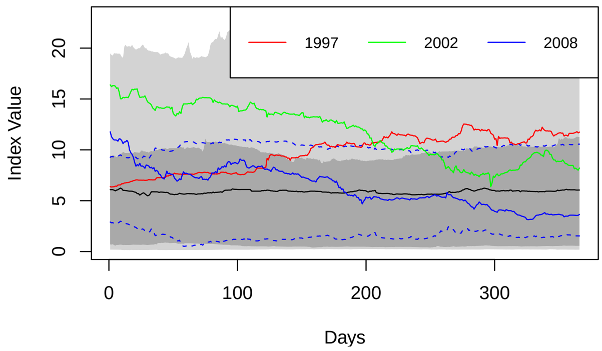

Figure 6: Functional Bagplot. The dark and light grey regions correspond to the bag and fence regions, respectively. The black line is the median curve, while the dotted blue lines correspond to 95% pointwise confidence intervals. The colored curves are outliers.

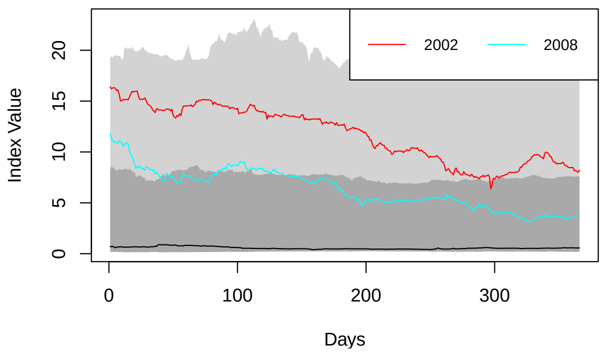

Figure 7: Functional HDR Boxplot. The dark and light grey regions represent the 50% HDR and outer HDR, respectively. The black line is the modal curve. The colored curves are outliers.

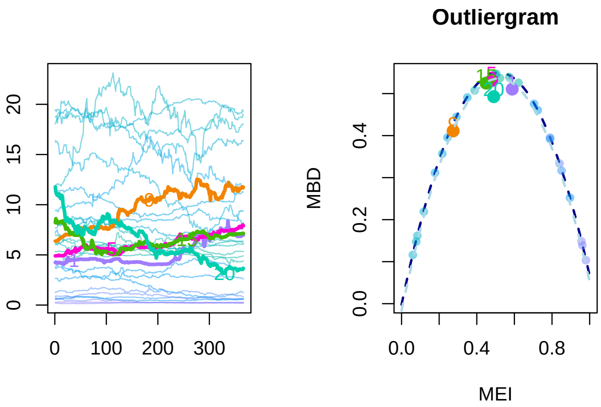

Figure 8: Outliergram. On the left: original data set with outliers in colored lines. On the right: outliergram with the outliers' IDs.

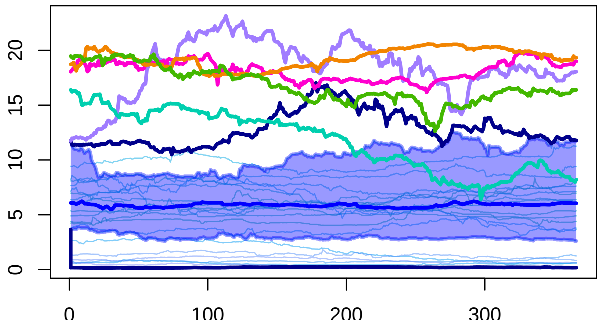

Figure 9: Robust adjusted functional boxplot. The fences are represented in dark blue and the outliers in colored lines.

Method	Detected Outliers
Functional Bagplot	1997, 2002, 2008
Functional HDR Boxplot	2002, 2008
Outliergram	1989, 1993, 1997, 2003, 2008
Functional Boxplot	1998, 1999, 2000, 2001, 2002

Table 1: Outliers detected by the different graphical tools.

The methods introduced in Section 3.1 based on distances were also applied to the data set, leading to the results in Table 2 (R package `rainbow`). The trimming percentage used was $\alpha = 0.1$.

Method	Detected Outliers
Likelihood Ratio Test	None
Functional Depth	1998, 1999, 2000, 2002, 2004, 2006, 2007, 2010, 2013
Integrated Square Error	1989, 1990, 1993, 1994, 1997, 1998, 1999, 2001, 2002, 2003, 2006, 2007, 2008, 2011
Robust Mahalanobis Distance	2002

Table 2: Outliers detected using methods based on distances.

The versions of static, dynamic and robust FPCA were applied on the smoothed data. In any version of FPCA, the number of principal components, K, was selected so that at least 99% of the variability was explained. The resulting first static principal component is represented in Figure 10, which shows the mean function and the consequences of adding and subtracting small amounts of each component. The first principal component, which accounts for 96.9% of the variability, appears to represent a constant vertical shift in the mean.

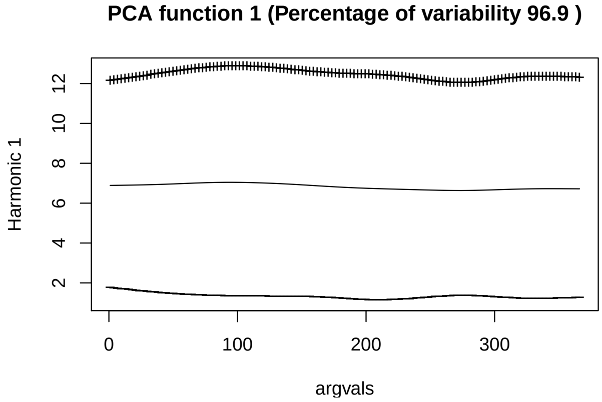

Figure 10: First principal component, represented by adding (+) and subtracting (-) the component from the mean function (line).

Likewise, the dynamic functional principal components resulted on the Karhunen-Loève expansion represented in Figure 11 with $K = 2$ ($q = \lfloor \sqrt{N} \rfloor = 5$ (Hörmann et al., 2015) was taken as the window size for the kernel estimator). These two dynamic principal components explain 99.15% of the variability in the data.

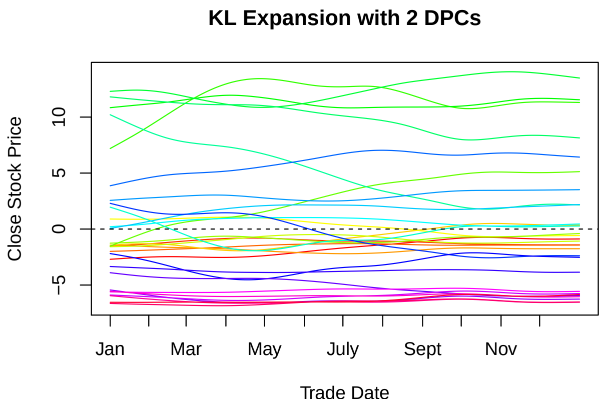

Figure 11: Karhunen-Loève expansion with 2 dynamic principal components (`freqdom.fda`). Smoothed data using an empirical base.

Moreover, the scores obtained from the first dynamic principal component are represented in Figure 12. These are the scores used to construct the new time series in the method based on projections (Section 3.3) and over which are detected outliers.

In the method based on projections introduced in Section 3.3, robust FPCA (Hyndman and Ullah, 2007) was computed, with a tuning parameter $\lambda = 3$. Therefore, the detected outliers using

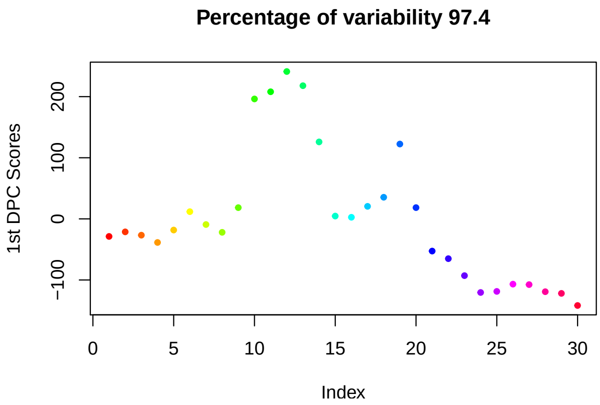

Figure 12: Scores obtained from the first dynamic functional principal component.

that method are summarized in Table 3. As mentioned before, the aim is to compare the results of using static, dynamic and robust FPCA as dimension reduction tools in FTS. In addition, two different methods of time series outliers detection methods were experimented on the principal components scores. These methods are implemented in the R packages `tsoutliers` and `anomalize`. The `tsoutliers` uses Chen and Liu (1993)'s outlier detection method, while `anomalize` implements a method called generalized extreme studentized deviation ("gesd").

Type of FPCA	TS Method	PC	Outliers
Static FPCA	`tsouliers`	1st	1997, 1998, 2000, 2002, 2003, 2007
		2nd	1997, 2002, 2008
		3rd	1998, 1999
	`anomalize`	1st	None
		2nd	1997, 1999, 2001, 2002, 2011
		3rd	1989, 1998, 1999, 2000, 2001, 2003
Dynamic FPCA	`tsouliers`	1st	1998, 2003
		2nd	1997, 2002
	`anomalize`	1st	1998, 1999, 2000, 2001
		2nd	1997, 2002
Robust FPCA	`tsouliers`	1st	1997, 1998, 2000, 2002, 2003, 2007
	`anomalize`	1st	None

Table 3: Outliers detected using the method based on projections.

A summary of the results is reported in Figure 13, plotting the outlying frequency of each curve using the three types of detection methods (graphical, distance and FPCA). Graphical Tools refers to Table 1, Distance Based Methods to Table 2, and FPCA Based Methods to Table 3. As the barplot suggests, the years 1997 – 2003 and 2007 – 2008 are the most frequently detected as outliers. In fact, 1997 – 2000 is seen as a period of social-economic development, while 2001 – 2003 and 2007 – 2008 as periods of recession. Thus, this results are not surprising and somewhat correspond to what was expected based on social-economic events.

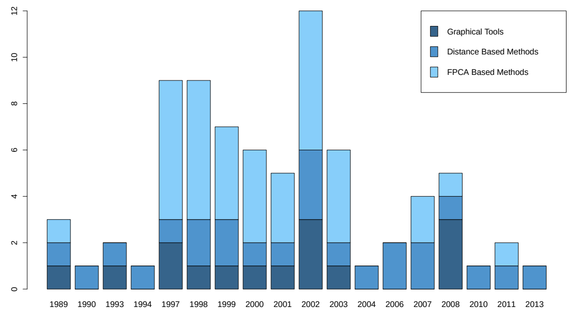

Figure 13: Outlying frequency of each curve using the three types of detection methods (graphical, distance and FPCA).

5. Conclusions and Future Work

In this work, a financial data set was analysed from a functional perspective. In fact, studying time series from a functional perspective is a still not very common. Most existing literature treats time series as scalars or vectors, and functional data as i.i.d. observations. Thus, FTS represents mostly uncharted territory. However, this approach allowed to detected anomalous curves with several outlying procedures. In this work, the dynamic FPCA was employed in an attempt to deal with the temporal dependencies inherent to FTS, which is a novelty in the literature.

For the BCP data set, modelling the closing stock prices as a functional time series allowed to search for curves whose behaviour differed from the rest. In this analysis, as each curve corresponded to a year, the aim was to search for years marked by recession or economic development. This pursuit begun with outliers detection methods based on graphical tools, on distances, and finally using a method based on projections. Also, taking into consideration the serial dependence structure was fundamental in order to avoid misleading conclusions. Consequently, when using dynamic FPCA two components were enough to explain 99% of the variability in the data, while with static FPCA three components were necessary. Therefore, DFPCA does seem to be better suited in representing this type of data. In addition, as reported in Table 3, there were less detected outliers over the dynamic functional principal components. This might be the result of these components accounting for the serial dependencies in the data, that otherwise would have been seen as outlying curves.

As for the methods based on distances they seemed less realistic in terms of results. Considering these methods were the first attempt at outliers detection in functional data, this is not surprising. The methods based on graphical tools and on projections seemed to have much more plausible results. This methods use principal component analysis in order to simplify the data representation and only then attempt at selecting outlying observations. Hence, performing FPCA seems to provide a clear advantage in the detection of outliers. For the BCP data set, the results obtained appear to correspond fairly well to what was expected based on external social-economic events.

Regarding future work suggestions, there are some lines of research that were left unexplored in the scope of this work. The most pressing suggestion is performing a simulation study to confront the outlying efficiency of the different methods. Then it would also be easy to compare the simulation output with the results obtained for the BCP data set. For example, check if the tendency for detecting less outliers with dynamic FPCA in the method based on projections holds.

Another suggestion is using the robust functional principal components estimators proposed by Bali et al. (2011) as another type of FPCA in Step 1 of the algorithm based on projections presented in Section 3.3. Some other time series outliers detection methods could also be experimented in Step 2 of this algorithm. Perhaps a robust time series outlying detector might be more suitable to the task at hand.

An additional topic would be to apply some changes to the methods based on graphical tools, in particular, to both the functional bagplot and the functional HDR (Hyndman and Shang, 2009). Considering that both of these methods first obtain the robust principal components scores, it would be interesting to obtain new plots based on the dynamic functional principal components scores instead.

Acknowledgements

The author would like to thank the financial support of Fundação para a Ciência e Tecnologia (FCT Portugal), through the research project PTDC/EGE-ECO/30535/2017 (Finance Analytics Using Robust Statistics).

References

Arribas-Gil, A. and Romo, J. (2014). Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics*, 15(4):603–619.

Bali, J. L., Boente, G., Tyler, D. E., and Wang, J.-L. (2011). Robust functional principal components: a projection-pursuit approach. *The Annals of Statistics*, 39(6):2852–2882.

Boente, G. and Barrera, M. S. (2015). S-estimators for functional principal component analysis. *Journal of the American Statistical Association*, 110(511):1100–1111.

Chen, C. and Liu, L.-M. (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88(421):284–297.

Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, 12:136–154.

Febrero, M., Galeano, P., and González-Manteiga, W. (2007). A functional analysis of nox levels: location and scale estimation and outlier detection. *Computational Statistics*, 22(3):411–427.

Febrero, M., Galeano, P., and González-Manteiga, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. *Environmetrics*, 19(4):331–345.

Gabrys, R., Horváth, L., and Kokoszka, P. (2010). Tests for error correlation in the functional linear model. *Journal of the American Statistical Association*, 105(491):1113–1125.

Gao, Y., Shang, H. L., and Yang, Y. (2018). High-dimensional functional time series forecasting: An application to age-specific mortality rates. *Journal of Multivariate Analysis*, 170:232–243.

Gervini, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika*, 95(3):587–600.

Horváth, L., Kokoszka, P., and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, 179(1):66–82.

Hubert, M., Rousseeuw, P. J., and Verboven, S. (2002). A fast method for robust principal components with applications to chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 60(1-2):101–111.

Hyndman, R. and Ullah, M. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics and Data Analysis*, 51(10):4942–4956.

Hyndman, R. J. and Shang, H. L. (2009). Rainbow plots, bagplots and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1):29–45.

Hörmann, S., Kidziński, L., and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):319–348.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. Taylor & Francis.

Ramsay, J. O. (1982). When the data are functions. *Psychometrika*, 47(4):379–396.

Ramsay, J. O. (2013). Functional data analysis. McGill University, Canada. Viewed 18 April 2019,`http://www.psych.mcgill.ca/misc/fda`.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society*, 53(3):539–572.

Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Use R! Springer.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer, second edition.

Sawant, P., Billor, N., and Shin, H. (2012). Functional outlier detection with robust functional principal component analysis. *Computational Statistics*, 27(1):83–102.

Shang, H. L. (2011). A survey of functional principal component analysis. Department of Econometrics & Business Statistics, Monash University, Australia. Working Paper.

Tarabelloni, N. (2017). *Robust Statistical Methods in Functional Data Analysis*. PhD thesis, Politecnico di Milano.

Vilar, J. M., Raña, P., and Aneiros, G. (2016). Using robust fpca to identify outliers in functional time series, with applications to the electricity market. *SORT: Statistics and Operations Research Transactions*, 40(2):321–348.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Review of functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295.