A study of functional data analysis on yield curves

Voort, I. van der

Leiden Univeristy

Bachelor Thesis

A study of functional data analysis on yield curves

Author
Indy van der Voort
s2036487

Supervisor
prof. Valentina Masarotto

September 14, 2021

Abstract

We introduce functional data and a basis function systems to represent it. The B-spline basis system is explored further as we will apply it in our application. We explore principal component analysis and functional principal component analysis. We apply functional principal component analysis on yield curves of the United States. We validate that functional principal component analysis gives similar results as principal component analysis. We create two different scenarios namely a crisis scenario and a growth scenario and apply functional principal component analysis on them. Here we see that in a extreme scenario the impact of the second and the third principal component both significantly increase. In the end we introduce forecasting of functional time series using a weighted functional principal component regression. We apply a simplified model of this forecast on the yield curves.

Contents

1 Introduction

Advancement in data collection technology has introduced a new type of data, namely functional data. New types of data require new tools for analyses. In this paper we will discuss functional data analyses. Data collection technology has made it possible to permit heavily densely observations sampled over time, space or other continua. We then assume that there are underlying functions which generate the observations, hence the name functional data. We observe this function on discrete time points which are our data points. From a mathematical perspective, they are taken to be random elements of an infinite dimensional Hilbert space. The interested reader should consult the book Horváth and Kokoszka (2012). Some examples of functional data are growth spurts, weather data and data of a person's handwriting.

These underlying functions and their derivatives hold additional information which could be very interesting. In multivariate statistics when we analyse data we have data points and treat all of them as individual observations. With functional data we treat a curve or surface as a single observation. The main problem with functional data is the Curse of Dimensionality which refers to the correlation between the amount of data points and the amount of space for these data points. With functional data we have a continuum hence there is a infinite amount of space but we only have a finite amount of data points. As a result Classical Multivariate data analysis is not fully able to analyse functional data. Thus there is a need for a alternative way to analyse the functional data, which is called functional data analysis (FDA). We will mainly consult the book of Ramsay and Silverman (2005)

As we then have the recorded curve the questions then arises how to interpret it. This curve can contain outliers and measure errors. Hence we want to find some kind of system which fits this curve but also keeps these disturbances in mind. Therefore with FDA we want to approximate the underlying function x which generates the data y. We use a system of basis functions to estimate x. A system of basis functions is a set of K functions which are independent of each other. There are various basis functions which can be used but for the sake of this paper we will only consider spline functions. FDA of course greatly depends on how we choose our basis system. Which basis functions, how many basis functions? Hence we view K as a parameter on itself. These basic functions should ideally have the same characteristics as the data. For example if we look at periodic data we would like to use functions which are themselves periodic. Because then we can capture this feature of data more naturally. The choice of basis system is particularly important for the estimates of the derivatives. We should also consider the smoothness of our basis system because if we estimate to precisely we might include too much noise in our system. This results in high-frequency oscillations and poor estimates of the derivatives. So there is a trade-off between the fit of our basis system and the noise of the measurement.

As stated before we are going to focus on spline functions for our basis system. Spline functions are the most common used functions for the approximation of systems for non-periodic functional data. This system combines fast computation of polynomials with flexibility. We divide our data in intervals and we define over each of these intervals a polynomial of order m. At the knots which is where the adjacent intervals the polynomials join up "smoothly". Placement of knots can be at intervals, equally spacing, at data the points or dependable of density of the data. The splines are then fully characterised by the location of the knots and their order. Increasing the order does not always comes with a better estimation. Because if we increase the order we also increase the influence of noise in our system. There are multiple ways to construct such systems but the most common used is the B-spline basis developed by de Boor De Boor (2001).

We are going to apply our basis system to yield curves of the United States. Yield curves represent annual interest rates of government-issued stocks. The cure is a graph of yield versus time to maturity. We use the package "FDA" and "Yieldcurve" in R to get our data. We will first look at principal component analysis in a multivariate context. After we will transcribe this to an functional context, functional PCA (fPCA). The main goal of PCA is to identify patterns in the data. PCA aims to detect the correlation between variables. If a strong correlation between variables exists, the attempt to reduce the dimensionality only makes sense. Functional PCA (fPCA) uses the same thinking as with PCA but only in an functional context. PCA has been applied to Yield curves in infinite dimension and we want to validate those results.

The structure of the thesis is as follows. In chapter 2 we will first show some examples of functional data. We explain why we call them functional data and discuss the underlying function which we assume exists. We than will explain more about functional data and how to approach it. Chapter 3 we introduce the system of basis functions. Especially the method we will use later with the yield curves, namely b-splines. Now we have a way represent the data we will discuss principal component analyses and functional principal component analyses in chapter 4. This is largely based on Ramsay and Silverman (2005). Now we can apply this technique on yield curves using R. At first we will introduce yield curves and explain that they are often used as health indicators of the economy. We will discuss our strategy and how we did our fPCA. Then we discuss our results and see if we get similar results as using PCA. At the end we investigate how much impact specific economical states have on our fPCA and present our conclusion.

2 Introduction to Functional Data Analysis

2.1 Introduction

We call Functional Data a set of observations along a continuum which are taken together to form a single curve or surface. Time is a common used continuum but any continuous domain is sufficient. Other continuum's are spatial position, frequency, weight, and so forth.

An example of functional data is the temperature and precipitation at 35 different locations in Canada over 1960 to 1994. The data is collected from 35 different weather stations in Canada. The temperature curves are displayed in figure 1. Even though the cities are not all close to each other they do show similar trends. For example Vancouver and Toronto are both in this data set and still they have the same trend. This could be explained by seasonal weather. Furthermore the curves are reasonable smooth as there are no extreme outliers or high oscillations. Therefore we can assume that there exists an underlying function which generates the data. So we then view the temperature of a city over a year as one entity captured in a curve and treat it as such. The

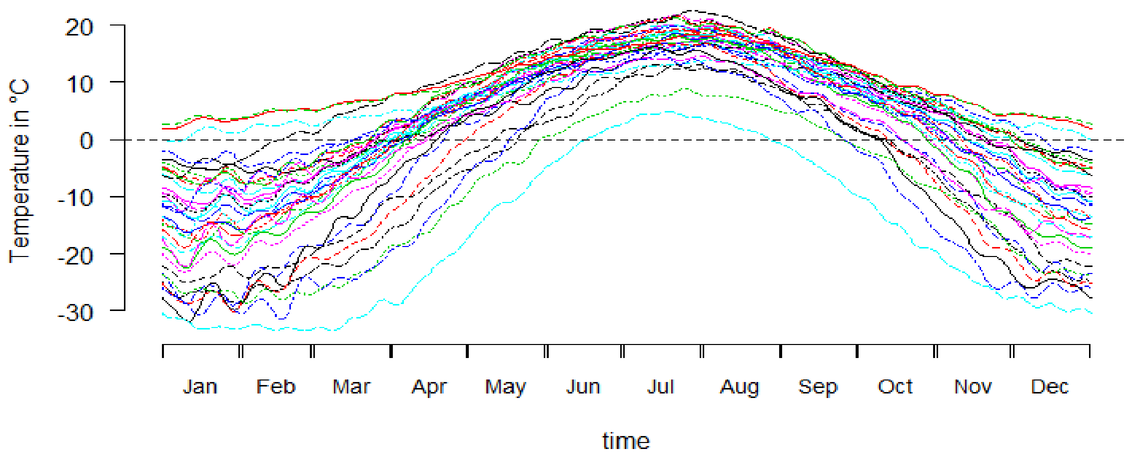

Figure 1: Weather data of different cities.

main application that we will explore later in the thesis are yield curves. The term yield curve can be used on various curves in a economic setting. We will look at the most common used setting which is in the bond market. Here it refers to a plot made by the yields of government-issued stocks. Where the yield curves represent effective annual interest rates of these bonds. The curve is a graph of yield versus time to maturity. Here we also assume there is an underlying function which we will try to estimate and use to analyse the data.

2.2 Functional data

We consider a observation to be functional if the data implies the existence a function x which generates the data. This does not mean that x is fully recorded because that would mean that we would have to store an infinite number of values. Furthermore we want x to be smooth. We mean by smooth that the function x can be differentiated one or multiple times. In practice we only assume this property holds and we do not prove if directly. We will use the discrete data $\mathbf{Y} = \{y_1, y_2, ..., y_n\}$ to estimate the function x and some of its derivatives. Denoted $D^m x$, where m is the order of the derivative. For example if we track the position of an object the underlying function x would then represent the position of the object in space and time. Its first derivative Dx would be its velocity and D^2x is the acceleration. Both of these derivatives store a lot of information and hence are important for the analysis of the data. The actual collected data may not be smooth at all due noise and measurement error. Hence we smooth our data before we further analyse it. The interval $\mathcal{T}$ over which the data is recorded may vary between the curves. Also the interval can be periodic for example if we look at yearly data. The function x in the beginning of the interval will smoothly pick up the values of the end of the interval. We call those functions periodic and non-periodic if the function does not have clear periodic behaviour. Finally, the data may be distributed over a multidimensional domain. For example brain photographs where the intensity and possibly colour composition is a function of spatial location.

The smoothness of the function x is an trade-off between additional information and the possibility of including more noise. If we have our data vector $\mathbf{Y}$ we use the following notation to express y_i

$$y_i = x(t_i) + \epsilon_i, \tag{1}$$

where ϵ_i is the error term regarding y_i. This holds for $i = 1, 2, ..., n$. If we use vector notation we get a simpler and cleaner expression

$$\mathbf{Y} = x(\mathbf{t}) + \mathbf{e}, \tag{2}$$

Where $\mathbf{Y}$, $x(\mathbf{t})$ and $\mathbf{e}$ are column vectors of length n. In the standard multivariate statistics model for the distribution of the ϵ_i's we assume that they are independently distributed with mean zero and constant variance σ^2. Then in this model we have the following formula for the variance

$$Var(\mathbf{Y}) = \sigma^2 \mathbf{I} \tag{3}$$

where I is the identity matrix of order n. However if we want to use this assumption of the independently distributed errors with mean zero and constant variance σ^2 for functional data it oversimplifies the situation. Which is not viable. Rather we can assume that the variance of the residuals will vary over the continuum. Also we have to consider correlation between ϵ_i's. In fact the concept of independently distributed errors which is made in the standard model is not realistic in nature because it would require infinite dimension to achieve. Here the meaning of dimension is that the observed data captures all of the information.For example, fluctuations in a large stock market are often treated as having white noise properties, but in reality only a limited number of stocks can be traded within a short time interval such as a millisecond, and consequently scale stock prices will exhibit some structure within a time scale that is small enough. However we do not always necessarily have to model the variable variance or auto-correlation structure in the residuals. If we do use a complicated model for the variance and auto-correlation structure of the residuals this could slow down computation significantly. However the estimates of functions could be approximately the same as when we assume independence in residuals. Hence the payoff might not always be as significant and so you should consider the importance of your residuals model. Nevertheless a model which only specialises in the variance and/or auto-correlation can be very interesting itself and may pay off with better estimation. We should also keep in mind that errors or disturbances might multiply rather than add when the data are intrinsically positive, in which case it is more sensible to work with the logarithms of the data. For more details about the structure of the model of residuals and its importance one should consult Ramsay and Silverman (2005).

2.3 Functional means and variances

The classical multivariate statistics definitions apply equally to functional data. The mean function with values

$$\bar{x}(t) = N^{-1} \sum_{i=1}^{N} x_i(t) \tag{4}$$

is the average of the functions point-wise across "replications". Similarly the variance function **var** has values

$$var_x(t) = (N-1)^{-1} \sum_{i=1}^{N} |x_i(t) - \bar{x}(t)|^2, \tag{5}$$

and the standard deviation function is the square root of the variance function.

2.4 Curve registration

Besides dealing with discrete measurements, often we have to deal with perturbed data as well. Intuitively speaking, such a perturbation implies that the "peaks" and the "valleys" of given functions are not aligned properly. We have two variations of this misalignment of the "peaks" and "valleys". One with respect to the x-axis and the other to the y-axis.

It is visible from figure 2 that the curves show similar behaviour but with a difference in the magnitude of the peaks as well as a misalignment with respect to the x-axes or y-axis. These horizontal perturbations might come from an uncertainty in the data sampling process or they can represent an inherent variability of the process itself that needs to be separated from the variability along the y-axes. Mathematically we talk about amplitude variation and phase variation.

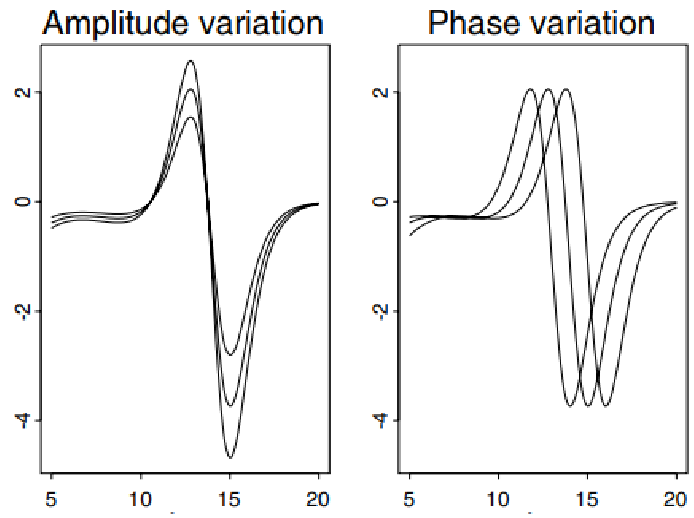

Figure 2: Perturbed data displaying amplitude variation and phase variation.

Whether phase variation is a problem for the statistical inference depends on the specific application. In general, the Karhunen–Loeve expansion of X is very different from the one of X and an unaccounted-for warping can introduce errors and artefacts in the statistical analysis. There is no canonical way to solve the curve registration problem. Liu and Müller (2004), Panaretos and Zemel (2016), Kneip and Gasser (1992), Tang and Müller (2008) among others offer a variety of different registration methods. We will not further discuss these methods in this paper. Amplitude variation however can be easier to solve. A method to solve this problem is multiplying the curves with respect to their mean.

3 Representation of functional data

In this chapter we will discuss how to represent the data when handling functional data. In multivariate statistics it is more straightforward. We have a set of points and this can be easily represented as there are no unseen variables which you have to consider. With functional data we assume that there exists an underlying function and this is what we want to represent. If we succeed in doing so we can use strategies from the multivariate statistics and transform them to be applied on the given representation.

3.1 Basis functions

A basis function system is a set of know functions ϕ_k that are independent of each other in a Hilbert space. Again consult Horváth and Kokoszka (2012) for a more detailed explanation. This system can approximate arbitrarily well any function by taking the weighted sum or a linear combination of a number K of these functions. Basis functions give a linear expression of a function x

$$x(t) = \sum_{k=1}^{K} c_k \phi_k \tag{6}$$

where there are K known basis functions ϕ_k. Let $\mathbf{c}$ be the vector of length K of the coefficients c_k and Φ as the functional vector whose elements are the basis functions ϕ_k. If we express (6) in matrix notation we get

$$\mathbf{X} = \mathbf{c}'\Phi = \Phi'\mathbf{c}. \tag{7}$$

In reality, basis expansion methods represent the potentially infinite dimensional world of functions within the finite-dimensional framework of vectors like $\mathbf{c}$. The dimension of the expansion is therefore K which is the amount of basis functions. However this is not a simple reduction to multivariate data analysis. FDA also greatly depends on how the basis system, Φ, is chosen. An exact representation or interpolation is achieved if $K = n$. Then we can choose the coefficients c_k to hold $x(t_j) = y_j$ for each j. Hence the smoothing of the data as opposed to interpolation of the data is determined by K. As a result we see K not as a fixed number but as a parameter itself which we choose according to the characteristics of the data. The basis functions should ideally have the same features as the functions that are being estimated. This would make it easier to estimate those functions while using a smaller number of K. If we have a small K we obtained a set of basis functions which captures the characteristics of the data. In general, the smaller K is we have more degrees of freedom we have to test hypotheses and compute accurate confidence intervals the less computation is required. Moreover the coefficients c_k might even become interesting descriptors of the data and contain valuable information. Hence there is no such thing as a universally good basis it all depends on the characteristics of the data. The choice of system is particularly important for the estimates of the derivatives

$$D\hat{x}(t) = \sum_{k}^{K} \hat{c}_k D\phi_k(t) = \hat{\mathbf{c}}' D\Phi(t). \tag{8}$$

Do note that some bases may work well for function estimation but are poor for the derivatives estimates. This is because accurate representation of the observations may result in small but high-frequency oscillations in $\hat{x}$. With functional data we can use various basis system however in this paper we will only discuss B-spline basis. We will explore this basis system in chapter 3.4.

3.2 Introduction to splines

Spline functions are the most common used functions for the approximation systems for non-periodic functional data. This system combines fast computation of polynomials with flexibility with the spacing of the intervals. This can often be realized with only a modest number of basis functions. We will first discuss the structure of spline functions, and then we will describe the usual basis system used to construct it, the B-spline system.

3.3 Spline functions

A spline is a string of piecewise polynomials on intervals which are "clicked" together. The first step in defining a spline is to divide the domain over which a function is estimated in L intervals separated by values τ_l, with $l = 1, 2, ..., L-1$ which are called knots or breakpoints. Note that the endpoints are τ_0 and τ_L. Over each of these intervals a spline is a polynomial of specified order m. The order of the polynomial is the number of constant needed to define it. The order is one more than its degree, its highest power.

We will give an example of a spline function used to approximate the sinus function on the interval $[0 - 2\pi]$. On the left side op the panel we see the spline basis, on the other side we can see how this interpolates to the points. In the top row the order of the spline function is three and we increase it by one each row. The number of basis functions increases then with the order. Adjacent polynomials join up smoothly at the breakpoints, so that the function values are bounded to the interval where they are defined on. So in the top row the number of basis functions is five and it increases by one each row. Here we can see that if we increase the order of the spline function the fit of the spline increases. But of course as there is no measuring error increasing the order stops having a effect after order 3. When we look at the derivatives of the spline functions they also must meet a condition. Namely the derivatives order up to order $m - 2$ must also be one more than its degree. For the total number of degrees of freedom we have the following rule

The total number of degrees of freedom in the fit equals the order of the polynomials plus the the number of interior breakpoints.

If there are no interior knots, hence the interval is the full domain, then the spline reverts to a simple polynomial. Increasing the order does not always come with a better estimation. Mainly because if we increase the order we also increase the influence of noise in our system. The main way to increase flexibility is by increasing the number of breakpoints. We can equally space the intervals but we can also increase the number of breakpoints where the data has the most complex variation. A notable remark is that we do not want an interval that does not have data. Which is reasonable because when there is no data we can not capture function's features. Moreover note that breakpoints are not the same as knots. This is because we can have that two or more breakpoints are coalescent or be coincident. With this we can engineer abrupt changes in a derivative or even a function value at pre-specified breakpoints. The interested reader should consult De Boor (2001) for further details on this matter. Thus the term breakpoint refers to the number of unique knots, while knots refers to the values at the breakpoints. In conclusion, a spline function is determined by two things: the order of the polynomials within the intervals and the knot sequence τ. There is also a rule for the required amount of parameters to define a spline function,

$$\#parameters = m + L - 1. \tag{9}$$

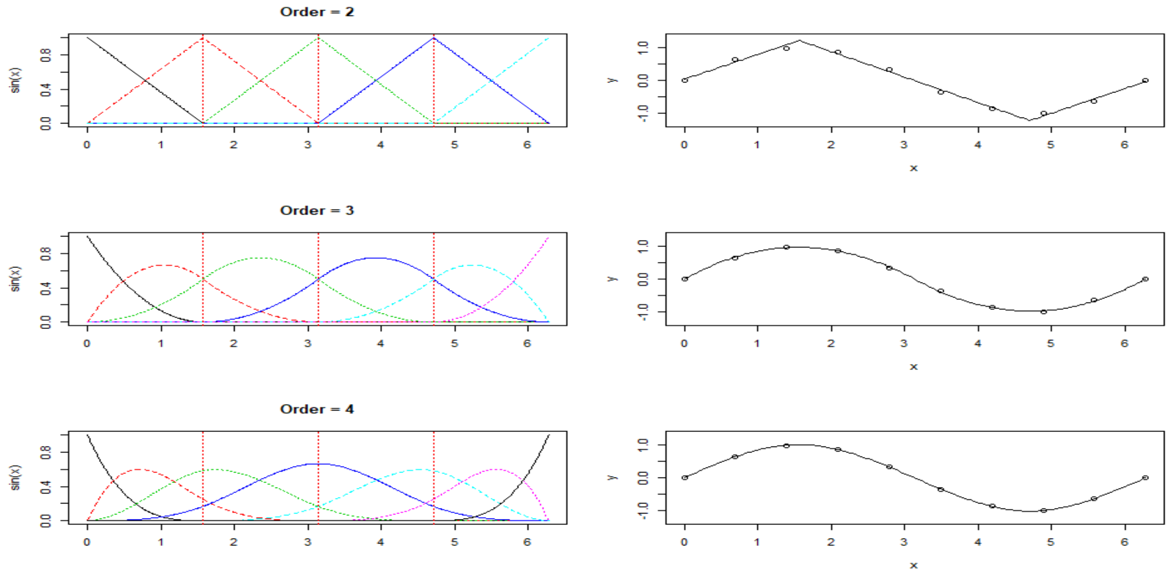

Figure 3: Sinus function.

3.4 The B-spline basis

As mentioned before we will mainly discuss the B-spline basis for the system of functions. We will use this basis later when we examine yield curves in chapter 4. In the previous sub-chapter we discussed splines but we have not yet talked about how to actually construct them. For this we will define a system of basis functions $\phi_k(t)$ with the following properties:

- each basis function $\phi_k(t)$ is itself a spline function as defined by an order m and a knot sequence τ
- since a multiple of a spline function is still a spline function, and since sums and differences of splines are also splines, any linear combination of these basis functions is a spline function
- any spline function defined by m and τ can be expressed as a linear combination of these basis functions

There are multiple ways to construct such systems but the most common used is the B-spline basis developed by De Boor (2001). Which is also widely available in programming languages such as *R* and Matlab. Other spline basis systems are discussed in De Boor (2001) and Schumacher et al. (1981).

An order m B-spline basis function has the property that it is positive over no more than m intervals. Furthermore these intervals are adjacent. This property is called the *compact support* property and is very important for efficient computation. When we have K B-spline basis functions, then the order K matrix of inner products of these functions will be band-structured. This matrix will only have $m-1$ sub-diagonals above and below the main diagonal containing non-zero values. Hence no matter how large K is the computation of spline functions can be organised such that in only increases linearly with K. Thus splines share the computational advantages of potentially orthogonal basis systems. Do note that we lose differentiability of the functions that we

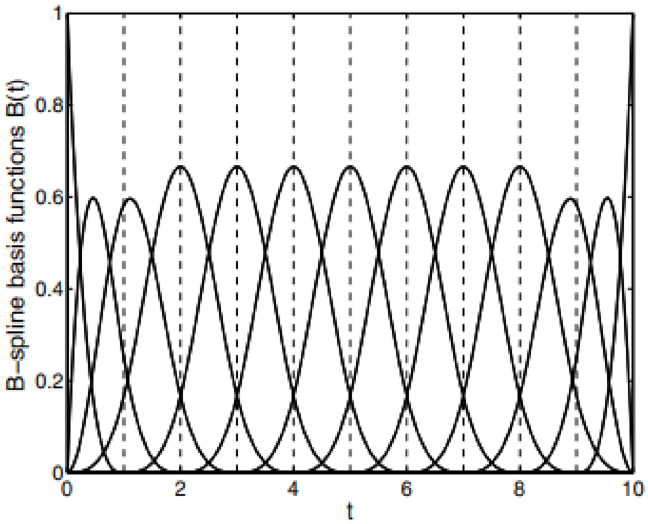

Figure 4: Thirteen basis functions defining an order four spline with nine interior knots, shown as vertical dashed lines, Ramsay and Silverman (2005)

try to estimate at the boundaries of the intervals. This is displayed in figure 4 where we lose differentiability at the boundaries. Which makes sense because we do not know what happens beyond the domain of the data. Hence we do allow the possibility that our functions may be discontinuous beyond those boundaries. Note that each of the seven basis functions complies to the *compact support* property. This behaviour of B-slpine basis functions is orchestrated by placing m knots at the boundaries. When B-splines are actually computed the knot sequence τ is extended at each end to add a additional $m-1$ replicates of the boundary knot value. We have discussed before that it is not always preferable to have $m-2$ derivatives at certain fixed points in the interior of the domain. With B-splines this is can be readily accommodated. We place a knot at such fixed points and every time we differentiate an additional knot is placed at that location as well. For example, if we were working with order four splines, and wanted the derivative to be able to change abruptly at a certain value of t but still wanted the fitted function to be continuous, we would place three knots at that value. We use the notation $B_k(t,\tau)$ to indicate the value at t of the B-spline basis function defined by the breakpoint sequence τ. The number k refers to the number of largest knot or the value of the point immediate left of the value t. This notation gives us $m+L-1$ basis functions, as required in the usual case where all interior knots are discrete. We can now define a spline function $S(t)$ with discrete interior knots as

$$S(t) = \sum_{k=1}^{m+L-1} c_k B_k(t,\tau). \tag{10}$$

The question arises where to place the interior breakpoints or knots. The most common used technique is equal spacing. This strategy works as long as the data is equally spaced. If this is not the case one should consider placing a knot at every jth data point, were j is a fixed number in advance. A special case of this is *smoothing splines* which we use in the Canadian weather example. The data is then smoothened by a certain roughness penalty. In spline smoothing the mean squared error is one way of capturing what we usually mean by the quality of estimate. In the Canadian weather example we look at the square of the second derivative $[D^2x(t)]^2$ of a function at t. For more about *smoothing* one should consult chapter 5 Ramsay and Silverman (2005).

Lastly, we can place more knots in regions of the data containing high curvature, and fewer in regions where there is less curvature.

An important note regarding features of spline bases is that increasing K does not naturally result in a better fit to the data. We fix the order of the spline however the function space defined by K b-splines is not necessarily contained within that $K+1$ B-splines. Complicated effects due to knot spacing relative to sampling points can result in a lower-dimensional B-spline system actually producing better results than a higher-dimensional system. However, if we increase K by adding a new breakpoint to τ or by increasing the order and leaving τ unchanged then the K-space is contained within the $(K+1)$-space.

4 Functional Principal Component Analysis

In this chapter we will discuss Functional Principal Component Analysis (fPCA) which we will apply on yield curves of the United States.

4.1 Principal component analysis

When doing analyses on any data there are multiple obstacles we need to overcome. Data sets can be very big, how do we read the data? And what to do with it? Principal component analysis (PCA) is a common technique for dimensionality reduction/extracting features/finding patterns in high dimensional data. It does so by focusing on the main direction of variation and to find features that summarise within them as much variation as possible. The interested reader should consult Jolliffe and Cadima (2016) for more details about PCA. Before we introduce fPCA we will first discuss PCA.

In classical multivariate analysis the variance-covariance and correlation functions can be difficult to interpret. And so does not always give a direct comprehensible presentation of the structure of the variability in the observed data. PCA provides a reasonable easy way to look at the covariance structure that is very informative. The central concept exploited many times in multivariate statistics is that of taking a linear combination of variable values,

$$f_i = \sum_{j=1}^{p} \beta_j x_{ij},\ i = 1, ..., N, \tag{11}$$

or in matrix notation

$$f_i = \beta' x_i,\ i = 1, ..., N. \tag{12}$$

where β is the vector of weighting coefficients applied to the observed values in the vector $x_i = x_{i1,...,x_{ip}}$. In the multivariate situation, we choose the weights to display types of variation that are strongly represented in the data. We can define PCA by the following step-wise procedure. Here we define sets of normalised weights that maximise variation in the f_i's

1. Find the weight vector $\xi_1 = (\xi_{11}, ..., \xi_{p1})'$ for which the linear combination values

$$f_{i1} = \sum_j \xi_{j1} x_{ij} = \xi_1 x_i$$

Have the largest possible mean square $N^{-1} \sum_i f_{i1}^2$ subject to the constraint

$$\sum_j \xi_{j1}^2 = ||\xi_1||^2 = 1.$$

2. Carry out second and subsequent steps, possibly up to a limit of the number of variables p. On the mth step, compute a new weight vector ξ_m with components ξ_{jm} and new values $f_{im} = \xi'_m x_i$. Thus, the values f_{im} have maximum mean square, subject to the constraint $||\xi_m||^2 = 1$ and the $m-1$ additional constraint(s)

$$\sum_j \xi_{jk}\xi_{jm} = \xi'_k \xi_m = 0,\ k < m.$$

In the first step we maximise the mean square to identify the strongest and most important mode of variation in the variables. To be sure that the problem is well-defined we need the constraint of the unit sum of squares on the weights. Otherwise the mean squares of the linear combination values could be made arbitrarily large. On second and subsequent steps, we again seek the most important modes of variation, but require the weights defining them to be orthogonal to those identified previously, so that they are indicating something new. Of course, the amount of variation measured in terms of $N^{-1}\sum_i f_{im}^2$ will decrease on each step. At some point, usually well short of the maximum index p, we expect to lose interest in modes of variation thus defined. The definition of principal components analysis does not actually specify the weights uniquely; for example, it is always possible to change the signs of all the values in any vector ξ_m without changing the value of the variance that it defines.

The values of the linear combinations f_{im} are called *principal component scores* and are often of great help in describing what these important components of variation mean in terms of the characteristics of specific cases or replicates. To be sure, the mean is a very important aspect of the data, but we already have an easy technique for identifying it. While doing PCA we look at a number of dimensions that are smaller than the dimension of our data set but still as interesting as possible. With interesting we mean how much the observations vary along each dimension. We do not care how far away from the origin our data are, therefore we center data removing their mean. Also, this distance from the origin can be unequal across dimensions and skew what we want to see.

4.2 PCA for functional data

We have seen in short how PCA works for multivariate data now we are going to look how it works in a functional context. Because we now work with curves as our observations and the function values $x_i(s)$ are our variables. So we replace the discrete index j by a continuous index s. We combine the weight vector β_0 with the data vector x by calculating the inner product

$$\beta' x = \sum_j \beta_j x_j.$$

When β and x are functions $\beta(s)$ and $x(s)$, summations over j are replaced by integrating over s to define the inner product

$$\int \beta x = \int \beta(s)x(s)\,ds. \tag{13}$$

So the weight β_j become functions with values $\beta_j(s)$. Using the notation (13), the principal component scores with respect to weight β are now

$$f_i = \int \beta x_i = \int \beta(s)x_i(s)\,ds. \tag{14}$$

As discussed in the previous section PCA gives a way to explain our data. In the functional context PCA gives a way to approximate/explain the curves which we have observed. We will first need to find the appropriate weight functions. Again we will explain this by a step-by-step procedure.

1. We first have to find the *principal component weight* function $\xi_1(s)$ for which the *principal components scores*
$$f_{i1} = \int \xi_1(s) x_i(s)\, ds$$
maximise $\sum_i f_{i1}^2$ subject to
$$\int \xi_1^2(s)\, ds = ||\xi_1|| = 1.$$
2. Next, compute weight function $\xi_2(s)$ and principal component scores maximising $\sum_i f_{i2}^2$ subject to the constraint $||\xi_2||^2$ with the additional constraint
$$\int \xi_2(s)\xi_1(s)\, ds = 0$$
3. And so on as required where the weight function ξ_m must satisfy the mentioned additional constraint $\int \xi_k \xi_m = 0$, $k < m$ on subsequent steps, known as the orthogonality constraint(s)

Each weight function has to define the most important mode of variation in the curves subject to each mode being orthogonal to all modes defined in previous steps. Again we note that the weight functions are defined within a sign change.

Before we further investigate the computational methods for fPCA we first must look to a characterisation of fPCA. To do this we are going to discuss the characterisation of fPCA in terms of the eigenanalysis of the variance-covariance function or operator. As we have seen before we are in a functional context hence we view $x_i(t)$ as our observed values. Which we assume results from subtracting the mean function values, so that their cross-sectional means $N^{-1}\sum_i x_i(t)$ are zero. We define the covariance function $v(s,t)$ by

$$v(s,t) = N^{-1}\sum_{i=1}^{N} x_i(s)x_s(t). \tag{15}$$

One may prefer to use a divisor of $N-1$ to N since the means have been estimated, but it makes no significant difference to the PCA. We can find the principal component weight functions $\xi_j(s)$ using results discussed in Section A.5.2 of Ramsay and Silverman (2005). The weight functions satisfy the equation

$$\int v(s,t)\xi(t)\, ds = \rho\xi(s). \tag{16}$$

With respect to an appropriate eigenvalue ρ. The left side of (16) is an *integral transform* V of the weight function ξ defined by

$$V\xi = \int v(\cdot,t)\xi(t)\, dt. \tag{17}$$

We call this integral transform the *covariance operator* V. Hence we can express the eigenfunction as follows

$$V\xi = \rho\xi \tag{18}$$

where ξ is a eigenfunction rather than a eigenvector. So fPCA can be translated to the problem of the eigenanalysis of the covariance operator V.

As we have discussed the characterisation of fPCA we can return to some computational methods for fPCA. Suppose we have a set of N curves x_i. Before hand we have subtracted the mean curve. Let $v(s,t)$ be the sample covariance function of the observed data. We have setup the eigenanalysis problem in (16) and we will consider some possible strategies for approaching this.

While there are several strategies we will always convert the continuous functional eigenanalysis problem to an approximately equivalent matrix eigenanalysis task.

A simple but efficient strategy that we will discuss in this paper is to discretize the observed functions x_i to a fine grid of n equally spaced values s_j that span the interval $\mathcal{T}$. This approach is used by Rao (1958), Rao (1987) and Tucker (1958), who applied multivariate principal components analysis without modification to observed function values.

We obtain an $N \times n$ data matrix $\mathbf{X}$ which we can put into a standard multivariate PCA program. From this we can derive the eigenvalues and eigenvectors satisfying

$$\mathbf{V}\mathbf{u} = \lambda \mathbf{u} \tag{19}$$

for n-vectors $\mathbf{u}$. Do note that we may have a n which is much larger than N.

A possible strategy is to work with the $n \times n$ matrix $\mathbf{V}$ to find the eigenvalues of the eigenequation (19). The variance matrix satisfies $N\mathbf{N} = \mathbf{W}\mathbf{D}^2\mathbf{W}'$, and hence the non-zero eigenvalues of $\mathbf{V}$ are the squares of the singular values of $\mathbf{X}$. Furthermore the corresponding eigenvector are the columns of $\mathbf{U}$. If we use a standard FDA package, which we do, these steps are carried out automatically.

To transform these vector principal components back into functional terms there are some steps needed. The elements of the sample variance-covariance matrix $\mathbf{V} = N^{-1}\mathbf{X}'\mathbf{X}$ are $v(s_j, s_k)$ where $v(s,t)$ is the sample covariance function. Let $\tilde{\xi}$ be the n-vector of values $\xi(s_j)$. Let $w = T/n$ where T is the length of $\mathcal{T}$. Then, for each s_j

$$V\xi(s_j) = \int v(s_j, s)\xi(s)\, ds \approx w \sum v(s_j, s_k)\tilde{\xi},$$

so the functional eigenequation $V\xi = \rho\xi$ has the approximate discrete form

$$w\mathbf{V}\tilde{\xi} = \rho\tilde{\xi}.$$

The solutions of this equation will be the same as the solutions of (19), with eigenvalues $\rho = w\lambda$. The discrete approximation to the normalisation $\int \xi(s)^2\, ds = 1$ is $w||\tilde{\xi}||^2 = 1$, so that we set $\tilde{\xi} = w^{-\frac{1}{2}}\mathbf{u}$ if $\mathbf{u}$ is a normalised vector of $\mathbf{U}$. Finally to derive the estimate of the eigenfunction ξ from the discrete values $\tilde{\xi}$ we use a interpolation method. When the values s_j are closely spaced, the difference between interpolation method will be very small.

To discretize the integral is the first approach to fPCA used by Rao (1958), Rao (1987) and Tucker (1958), who applied multivariate principal components analysis without modification to observed function values.

There are multiple ways to motivate PCA, and one is to define the following problem: We want to find a set of exactly K orthonormal functions ξ_m so that the expansion of each curve in terms of these basis functions approximates the curve as closely as possible. Since these basis functions are orthonormal, it follows that the expansion will be of the form

$$\hat{x}_i(t) = \sum_{k=1}^{K} f_{ik}\xi_K(t),$$

where f_{ik} is the principal component value $\int x_i\xi_k$. As a fitting criterion for an individual curve, consider the integrated squared error

$$||x_i\hat{x}_i||^2 = \int [x(s) - \hat{x}(s)]^2\, ds$$

and as a global measure of approximation,

$$PCASSE = \sum_{i=1}^{N} ||x_i - \hat{x}_i||^2. \quad (20)$$

The problem is then, more precisely, what choice of basis will minimise the error criterion (20). The answer is exactly the same set of principal component weight functions that maximise variance components as defined above. For this reason, these functions ξ_m are referred to in some fields as empirical orthonormal functions, because they are determined by the data they are used to expand.

5 Analyses on yield curves

In this chapter we will discuss our analyses on yield curves. First we will explain what yield curves are, how we got our data and which packages we used. Next we will discuss and compare our results using fPCA with the results of a paper which used PCA. We consult Choudhry (2019) and European Central Bank (2021) to give accurate descriptions of our economical terms.

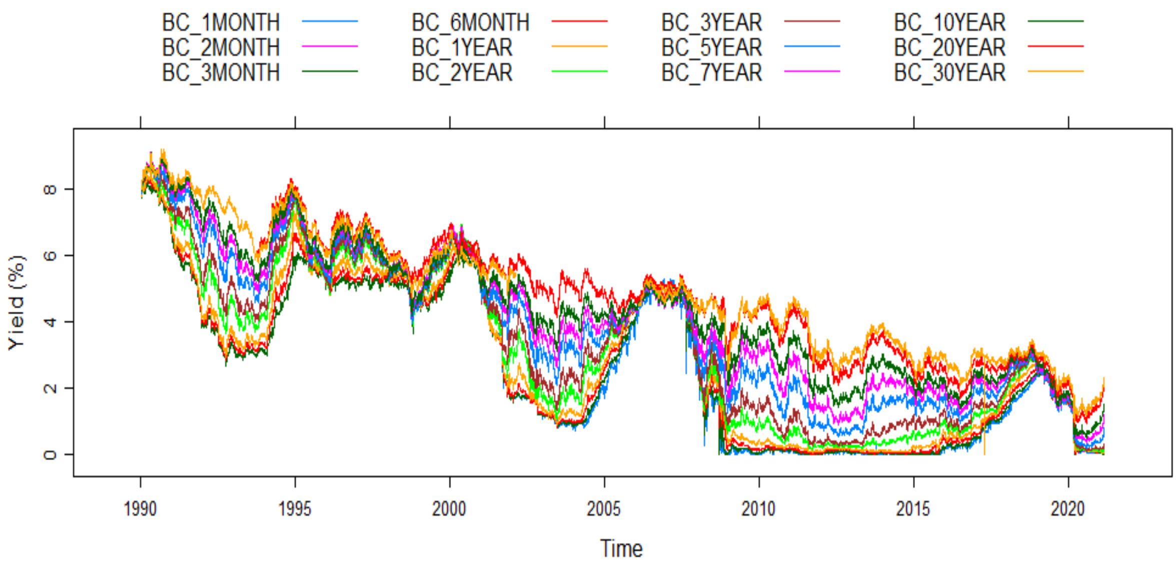

Figure 5: Interest curves of the United States bonds dating back to 1990.

5.1 Yield curves

Before we can discuss yield curves we first need to introduce bonds. A bond is a fixed income instrument that represents a loan made by an investor to a borrower (typically corporate or governmental). An individual bond is a piece of a massive loan. That's because the size of these entities requires them to borrow money from more than one source. Bonds are used by companies, municipalities, states, and sovereign governments to finance projects and operations. Owners of bonds are debt holders, or creditors, of the issuer. You have different type of bonds with various maturity dates.

A yield curve is a representation of the relationship between market interest rates and the remaining time to maturity of debt securities. Easier said governments or companies want to borrow money and to do so they sell bonds. These are then paid back over time with interest. These interest rates are then displayed in yield curves. A yield curve can also be described as the term structure of interest rates. Yield Curves date are published every trading day on the Treasury website (Federal Reserve, European central bank, ...). The yield curves that we shall analyse represents the interest rates of U.S. government bonds. From observing yield curves in different markets at any time, we notice that a yield curve can adopt one of four basic shapes, which are;

1. *Normal* or *conventional* in which yields are at “average” levels and the curve slopes gently upwards as maturity increases, all the way to the longest maturity;

2. *Upward-sloping* or *positive or rising* in which yields are at historically low levels, with long rates substantially greater than short rates;

3. *Downward-sloping* or *inverted or negative* in which yield levels are very high by historical standards, but long-term yields are significantly lower than short rates;

4. *Humped* where yields are high with the curve rising to a peak in the medium-term maturity area, and then sloping downwards at longer maturities.

We very rarely observe a somewhat flat yield curves which suggests that investors require different rates of return depending on the maturity of the instrument they are holding. It is also possible that yield curves show some combination of the above-mentioned features. For instance, a commonly observed curve in developed economies exhibits a positive sloping shape up to the penultimate maturity bond, and then a declining yield for the longest maturity.

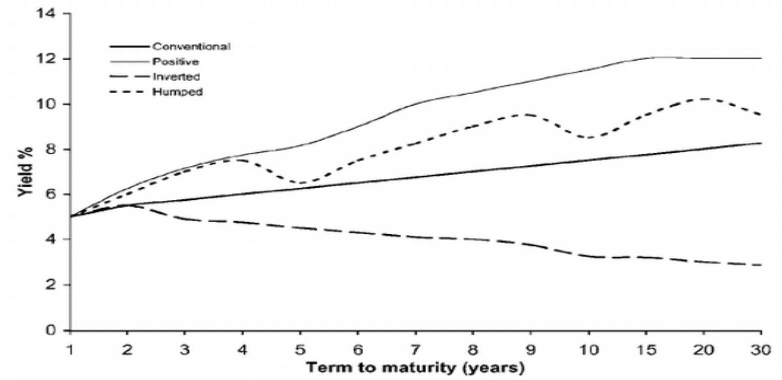

Figure 6: Observed yield curve shapes, Choudhry (2019)

When buying and selling bonds, investors include their expectations of future inflation, real interest rates and their assessment of risks. An investor calculates the price of a bond by discounting the expected future cash flows. The information content of a yield curve reflects the asset pricing process on financial markets. Asset pricing theory tries to understand the pries values of claims to uncertain payments. The interested reader should consult Cochrane (2009). The role of asset prices including interest rates, stock returns, dividend yields and exchange rates as predictors of inflation and growth is extensive and the literature is studied with great interest by policymakers[1].

[1] Mehl (2009)

Interest rates are normally viewed upon as always positive. The unconventional situation has come to pass that we have negative interest rates. Yields on bonds and Treasury securities (government bonds) have dipped below zero. Interest rates depend on supply and demand and this is indicated by the neutral interest rate. Which is the rate where the inflation is steady and the economy is at full employment. This neutral interest has been decreasing over the past decades and as a result it is at an all-time low. All over the world countries and some big companies show negative interest[2]. Negative interest rates operate in an upside-down world of banking. Instead of a bank paying you to hold your money so you can save you have to pay them.

The main reason of implementation of negative interest rates is reserving them for the most desperate of times, economically speaking. They have been promoted as methods to spur intense borrowing and spending, to push inflation back up to target and to devalue a country's currency, boosting trade. Hence to increase the growth of a country's economy.

But there is also reason for this negative interest which can not be solved that easily. The main reason is oversupply and a shortage of demand. The oversupply is caused by various developments. One of them is that we have developed in what is called an ageing society. Life expectancy has increased and therefore our pension is longer and we need more money, so we save more money. The demand is also a lot lower as the biggest companies are tech companies like Amazon, Google and Facebook. These large tech companies mainly offer services and do not need expensive machinery so there is less need of big investments. Furthermore if you loan money it must be profitable else you would only lose money on the interest rates. One should consult Cœuré (2016) and Rognlie (2016) for more detailed economical reports about negative interest.

Another concerning development is the inverted yield curve. When the interest rates on long-term government bonds are lower than the interest rates on short-term government bonds, the yield curve is inverted. For many economists, this means one thing: an economic recession is on the way. That's because, for the past six decades, three months of an inverted yield curve was followed by a recession. The yield curve has inverted in March and has inverted for over a quarter so are we going in to a future recession? This is not certain because the yield curve is distorted by the FED. And policymakers do know about the inverted yield curve and act accordingly. The interested reader should consult Wright (2006).

5.2 Data

We got our data from de YieldCurve package in R via command data(FedYieldCurve). These are the yield curves with maturity dates 3 months, 6 months, 1 year, 3 years, 5 years, 7 and 10 years. Our yield curves start in 12-31-1981 and end in 30-11-3012. Our time periods are months and we look at monthly mean. We convert our data to a matrix and from there we can compute our b-spline basis functions. We consult the book J. Ramsay, Hooker, and Graves (2009) for most of the packages and functions.

5.3 Fitting basis to Yield curves

We are going to use the package "FDA" to analyse the data. We plotted a sample of the yield curves which can be seen in figure 7. We want to compute the basis functions and for this we are going to use B-spline basis functions. Using the function smooth.basis we compute these functions in R. Now we can interpolate these basis functions to our data and then we can do our analysis on this approximation.

[2]PWC (2019), European Central Bank (2021)

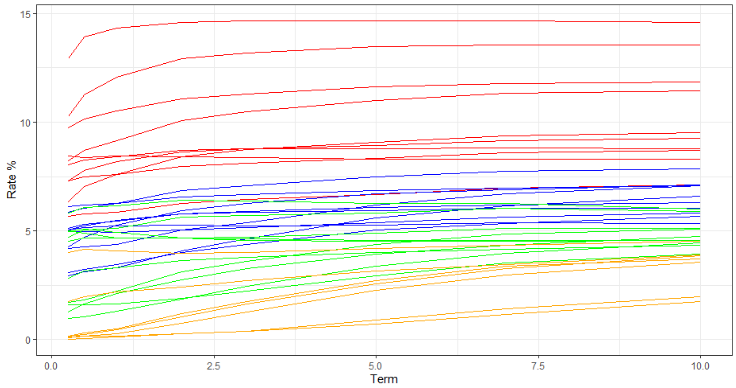

Figure 7: Sample of yield curves from our data.

5.4 Analyses of Yield curves

We have fitted a system of b-spline basis functions which approximates the underlying function. We will use this system to perform fPCA on the yield curve data. To check if FDA is a possible tool to analyse functional data we want to validate statements made by researchers and economist about yield curves who used PCA. If we obtain comparable results it would be an argument to use FDA more often when analysing functional data.

We will directly compare our results with Ramin Nakisa (2021). They applied PCA on U.S. yield curve data which dates from 1950-01-01 up to the present day. The maturity dates are 1Y, 2Y, 5Y, 7Y and 10Y. Note that we have more maturity dates. The results of their PCA is illustrated in figure 8. The red flat line is the first principal component, and corresponds to a parallel move up

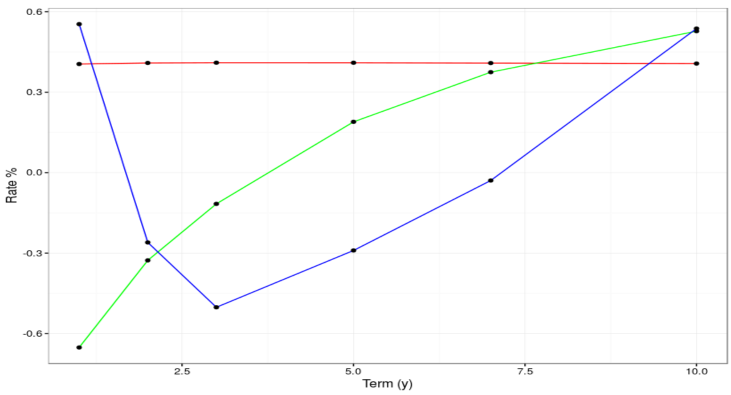

Figure 8: Red = PC1, Green = PC2, Blue = PC3.

and down in the level of the entire yield curve. The green line is the second principal component and is a steepening and flattening of the curve. The third component is a twist. For this data the first principal component captures 98.98%. The second component manages to capture 0.96% and the third component 0.05%, beyond that only 0.01% of variance remains.

5.5 Results of fPCA

We will break this subsection in three parts. In the first part we will talk about what we did and why we used certain strategies. The second part will be about our results and in the third and last part we will compare our results with the results of the earlier mentioned paper which used PCA.

5.5.1 Strategy

For all our programming we used R. We used the packages "YieldCurve" and "fDA". For the approximation of basis system of functions we use b-splines as this is the most common used for non-periodical data. For conveniency we equally space our intervals so we have eight knots which results in seven intervals and six breakpoints. The number of basis functions is eight so we have cubic splines as nbasis = nbreaks + order −2 hence the order is four. We chose for cubic splines as again it is most commonly used. Using the package "FDA" we could easily apply fPCA and we only computed the first four principal component as the other principal components are negligible against the first three.

5.5.2 Results of fPCA

First we look at the variance explained by the first four principal components.

PC1	PC2	PC3	PC4
0.9843	0.0147	0.00065	0.00016

The principal components that are computed using fPCA are similar to the principal component computed using PCA by Ramin Nakisa (2021). The first principal component explains almost all the variance, and after the third principal component there is little variance not explained. In figure 9 there is a plot of the eigenfunctions to check if the interpretation of the first principal component is also the same. We can than use this to check the statements with respect to the interpretation of the second and third principal component. We will only plot the first three principal component as after that there is not much variance unexplained.

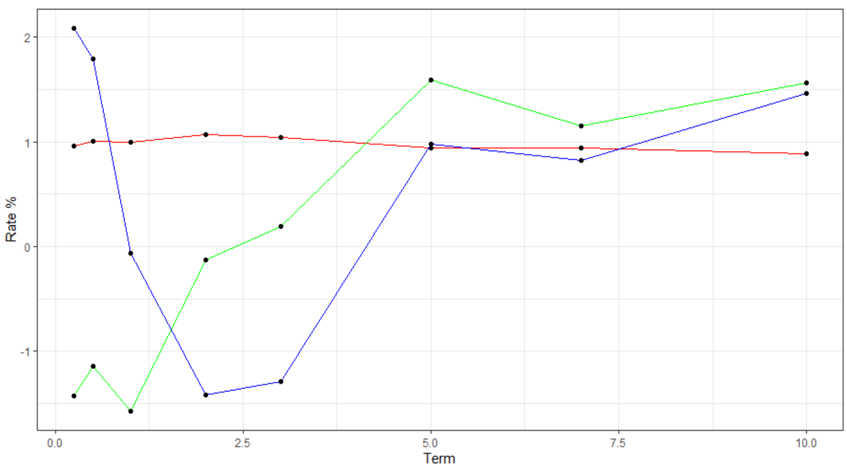

Figure 9: Red = PC1, Green = PC2, Blue = PC3

All the yield curves can be constructed as linear combinations of these first three principal components. We can derive from this that the first principal component corresponds with a feature of the yield curves which is constant over time, the intercept. Combining this with the variance explained by this principal component we know that the intercept of the yield curve has a big impact on the rest of the curve. The second principal component is orthogonal to the first principal component. Moreover it is increasing and hence this hints that it corresponds with the slope of the yield curves. We will see in other plots this is indeed the case. We can not say much about the third principal component yet but as its slope does change this might imply that it has something to do with the curvature of the curve. Furthermore, we will investigate this further in the same manner as we will do with the second principal component. As we know that the first principal component corresponds with the intercept we can use this to determine the corresponding features of the second and third principal component. We do this by using scatter plots where we plot the first principal component to the second or third principal component. We will first discuss the second principal component.

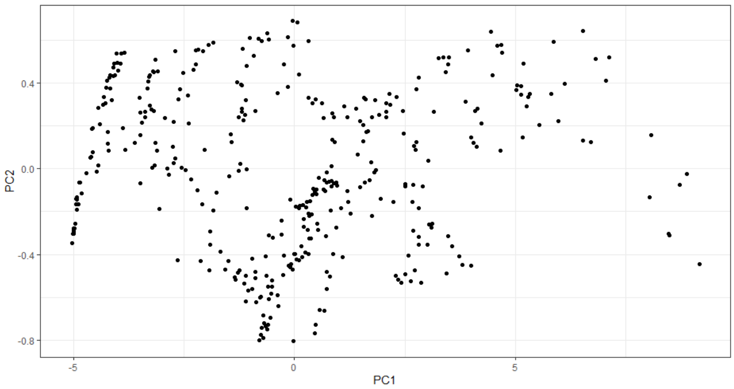

Figure 10: PC1 to PC2 scatter plot.

Because all the principal component are orthogonal to each other we can rotate the data in the direction of the principal components. This way we can observe the relation between the second principal component with a certain feature of the data as we know what the first principal component relates to. Note that the yield curves are highly correlated hence we do not see clear clusters (each point represents a yield curve). So the idea is that we take two curves with a similar value of the first principal component, which corresponds with the intercept, and we have opposite values for the second principal component. If we then plot these two curves in a graph we can see what the second principal component corresponds with. This is displayed in the following plot

Note that the curves have opposite slope and using the information from figure 10 we can determine that the second principal component corresponds with the slope of the yield curve.

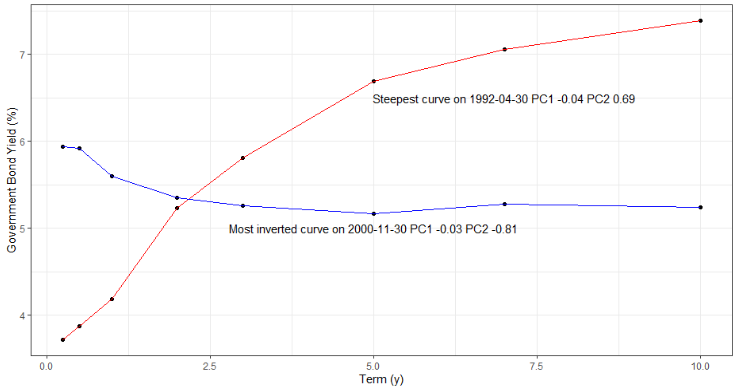

Figure 11: Two curves with similar first principal component but are differentiated by the second principal component

We can do the same for the third principal component and we observe that is corresponds with the flexing of the slope of the yield curve. Again we use the first principal component to

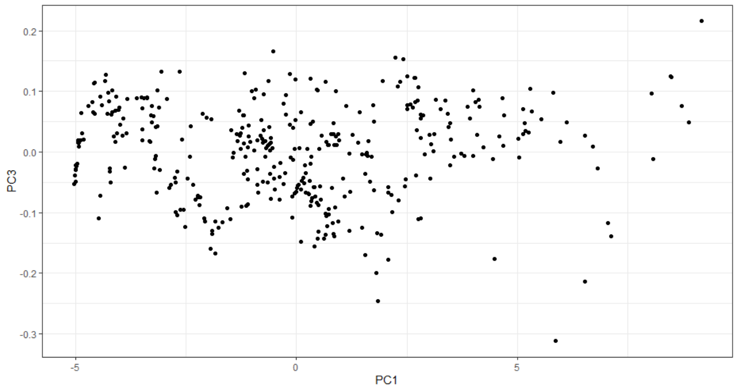

Figure 12: Scatter plot PC1 to PC3

differentiate the third principal component.

Again we take two curves with a similar intercept but have very different values for the third principal component. We can derive then from this plot that the third principal component corresponds with the flexing of the yield curve.

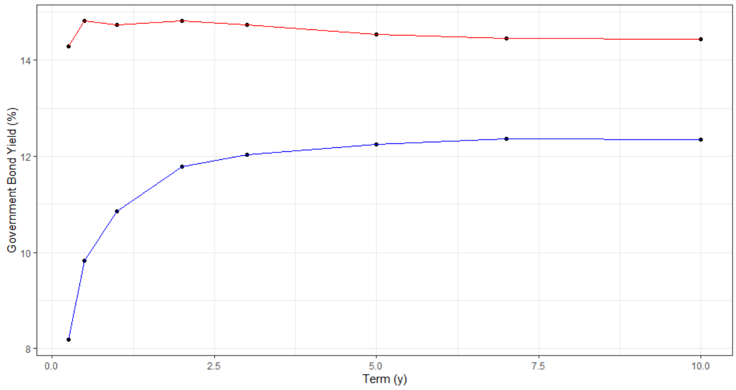

Figure 13: Curves which are differentiated by the third principal component

5.5.3 fPCA versus PCA

As stated before we are going to compare our results with the PCA displayed by Ramin Nakisa (2021).

	PC1	PC2	PC3
FPCA	98.43%	1.47%	0.065%
PCA	98.98%	0.96%	0.05%

The differences are very small moreover as we can see in the plots of respectively the eigenvectors and eigenfunctions in figure 8 and figure 9 are also very similar. From multiple papers and analyses of applying PCA on yield curves it followed that the first principal component corresponds with the intercept, the second principal component with the slope and the third principal component with the curvature of the yield curve. As we have seen in chapter 5.5.2 we can validate these statements using fPCA.

5.6 Extreme situations and its influences on principal components

We know that fPCA is a way to handle big data sets and give a better interpretation of the variance structure. We have seen we can extract certain features of the data set corresponding to the first three principal components. However we have a big data set containing multiple scenario's of economical "stages". Our data encompasses both economic prosperity and also a huge financial crisis. Furthermore the economy has overall grown during the time period of our data. The question is that we asked ourselves is whether these scenarios have a big impact on our fPCA. Does the relation between the principal components and the features of the yield curves still hold if we look at extreme situations? Is the first principal component still the most significant? Do these scenarios have any influence on the variance explained by these principal components?

Our data starts in 1981 and ends in 2012 during this time period there have been a period of great economic growth and a financial banking crisis. We made two scenarios namely; The growth scenario which is the period of 1998-04-31 to 30-04-2002 as according to Forbes (2020) that in this period there was great economic growth. The crisis scenario is set in the period 30-06-2008 until 30-11-2011. As in this period the financial crisis occurred, The Guadrian (2011). Note that these periods are not equally big. We then applied the fPCA on this data sets in the same manner as we did before. We will compare these result per principal component starting with the first one.

PC1

Variance explained by PC1:

All	Growth	Crisis
0.984	0.942	0.840

We can immediately see that the first component significantly explains less variance in our scenarios. Especially in the crisis scenario which indicates that the intercept is less important for our scenarios then in the overall situation. Before we continue our analyses we must note that the overall scenario is in fact also a growth scenario. We know this because the economy has grown in the period 1981-2012. So it can be expected that the growth scenario is similar to the overall scenario.

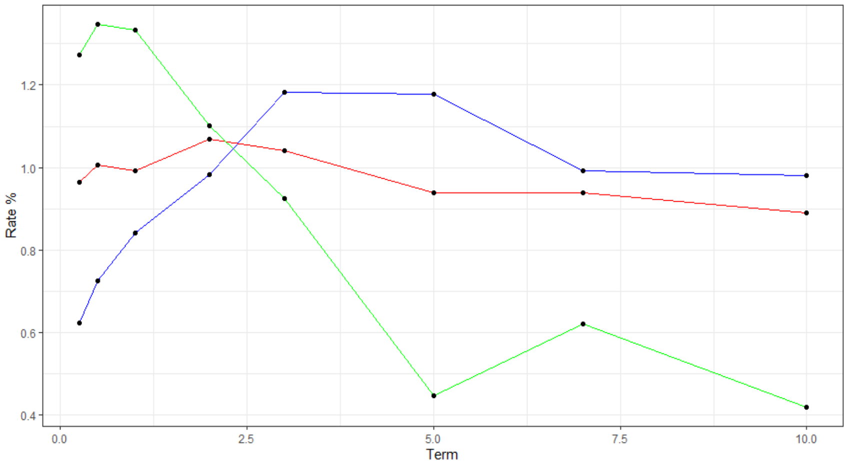

Figure 14: Red = All , Blue = Crisis, Green = Growth

First we must note that this a zoomed in picture hence there is more oscillation than in other plots. So in the crisis scenario we can still assume that the first principal component corresponds with the intercept. Whereas the second principal component in the growth scenario is no longer constant but decreasing. We can explain this by looking at the data set that we have for that period. The intercept during this period is constantly decreasing which could explain the decreasing component. Which means that it could also be that the first principal component relates to the the slope of the yield curve. Which is more probable hence we assume it does and we will try to validate this by looking at the second principal component. Note that this does not work in a similar way for the crisis scenario. This could be explained by the fact that the economy of the U.S. is very stable and strong and can withstand a crisis. There was always faith in the recovery of U.S. economy hence the component flattens when it crosses a certain threshold. Whereas the trust in the growing economy is as one that never stops growing.

PC2

Variance explained by PC2:

All	Growth	Crisis
0.0148	0.0557	0.1502

This result is very interesting as the second principal component is way more significant in our extreme scenarios, especially for the crisis situation. Before we make any statement about the interpretation of the second principal component of these scenarios we first have some notes.

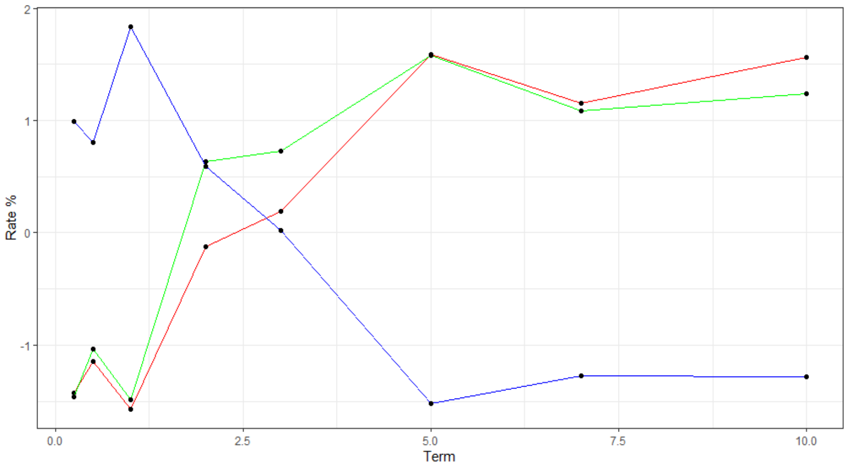

Figure 15: Red = All , Blue = Crisis, Green = Growth

For the crisis scenario we saw that the first principal component still relates to the intercept. Hence if this second principal component is similar to the second principal component in the overall situation we can make similar statements. However for the growth scenario we shall need to do some extra work. We will then again make the scatterplots and use this to determine the interpretation of the second principal component in the growth scenario. But first we will show the second principal components with respect to each other in the different scenarios.

In figure 15 the principal components have similar trends in all the scenarios. The component in the crisis situation is mirrored to the overall scenario but principal components are indifferent under sign changes. The growth scenario is very similar to the overall result only it has a sharper slope in the beginning but it flattens the same as in the overall scenario. So we can still claim that the second principal component in the crisis scenario relates to the slope of the yield curve. Thus particularly in the crisis scenario the slope of the curve does explain a significant amount of variance. This could be justified by the following; the crisis happened unexpected and there was a sudden fast drop. So even when the intercept was relatively high the drop of yield rate was enormous. So the slope has a much bigger impact on the explained variance of the yield curves. As stated before we need to do some extra work to make any statement about the second principal component in the growth scenario. In figure 16 we have the scatterplot.

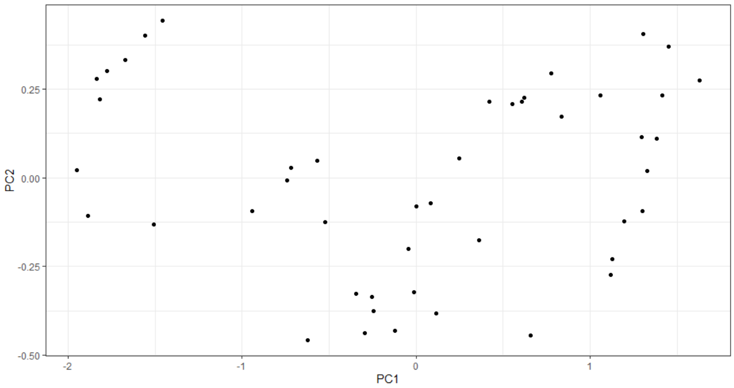

Figure 16: PC1 to PC2 in the growth scenario

We apply the same method as we did in the overall scenario and we pick two yield curves with similar values for the first principal component but different values for the second principal component.

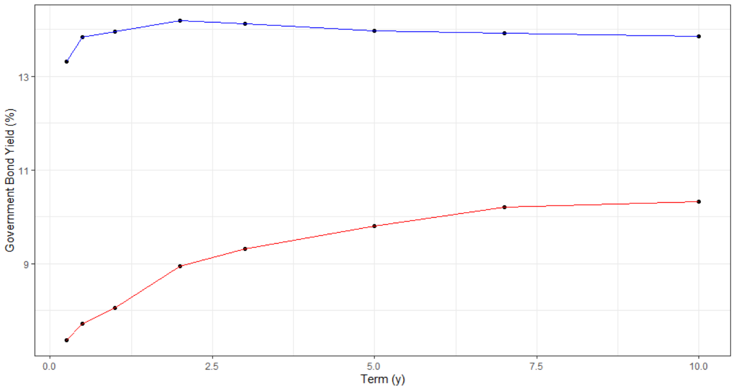

Figure 17: Yield curves differentiated by PC2

If we assume that the first principal component relates to slope than we it follow from these plots that the second principal component is related to the intercept. However it is not very clear and we do not have a strong result from these plots. Nevertheless if we keep our assumption this is a very shocking interchange of positions of components. As they are orthogonal on each other they swap places which means that on this particular time period the intercept is no longer the biggest principal component as before it was significantly the biggest. The reason of this could be sought in the overwhelming trust in the economy during a period of growth. As there would never come an end to the growth hence one would only look at the slope of the curve instead of the intercept.

PC3

Variance explained by PC3:

All	Growth	Crisis
0.0006	0.0019	0.0071

As in the overall scenario the third principal component is negligible in comparison with the other two principal components. The plots are very similar and as the third principal component is not

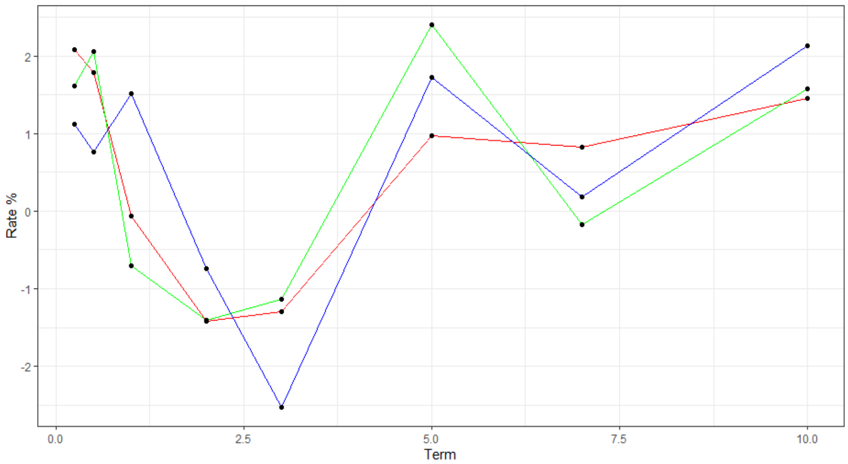

Figure 18: Red = All , Blue = Crisis, Green = Growth

significant in comparison with the other components the plot is not very interesting. Again we can still claim that the third principal component in the crisis situation is related to the curvature of the curve. For the growth scenario we can not make any statements as we have no strong claim about the interpretations of the first and second principal component.

6 Forecasting functional time series

In this chapter we will further explore applications of functional data analysis. We are going to apply forecasting on an functional time series, namely the before discussed U.S. yield curves. Before we can do this we need a technique to expand the multivariate method of forecasting to a functional context. We are going to use the method proposed in the paper Hyndman and Shang (2009). Here they propose a method of forecasting of functional time series using weighted functional principal component regression.

6.1 Forecasting

First we will briefly discuss the subject of forecasting. Forecasting is the prediction of the future of the data based on previous and present data. A common example is the prediction of future interest at some specific future time point. We will only discuss the forecasting of time series as we have time series of yield curves. Furthermore we have functional data hence we have functional time series. We observe our functions at times $t = 1, 2, ..., n$ and we want to forecast at times $t = n + 1, n + 2, ..., n + h$. We have our representation of our functions of the data which is displayed in equation (1). Here we assumed that the functions $x_i(t)$ are sufficiently smooth. With the yield curves we than have for the i'th maturity date the underlying function $x_i(t)$ at date t.

There are many other applications involving functional time series including those studied in Erbas, Hyndman, and Gertig (2007), Hyndman and Booth (2008), Hyndman and Ullah (2007) and Yao, Müller, and Wang (2005). Hyndman and Ullah (2007) used non-parametric smoothing on each curve $y_i(t)$ separately to obtain estimates of the smooth functions $x(t_i)$. Then they proposed a functional principal component approach to decompose the time series of functional data into a number of principal components and their scores. Their model can be written as follows:

$$x_t(i) = \mu(i) + \sum_{k=1}^{K} B_{t,k}\phi_k(i) + e_t(i). \tag{21}$$

Where $\mu(i)$ is the weighted functions and we will discuss this term in more detail later. ϕ_k is the k'th principal component and $B_{t,k}$ are the corresponding scores. Furthermore $e_t(i)$ is the error term which is i.i.d. random functions with mean zero. Because the principal components scores are uncorrelated, Hyndman and Ullah (2007) suggested that each univariate time series can be forecasted using a univariate time series model. By multiplying the forecasted principal component scores with the principal components, estimated future curves are obtained. In the paper of Shang, Hyndman, and Shang (2018) they extend the Hyndman–Ullah approach in several directions. They introduce geometrically decreasing weights in the principal component decomposition to allow more recent data to affect the results more than data in the distant past.

$$\mu(x_i) = \sum_{t=1}^{n} w_t \hat{f}_t(i).$$

Here $\hat{f}_t(i)$ are the estimated curves at maturity date i at time t and $w_t = k(1 - k)^{n-t}$ is a geometrically decreasing weight with $0 < k < 1$.

With yield curves recent data has more weight on the future curves hence we want to use the method explained in Shang et al. (2018) to forecast these curves. If we follow the steps from the article Hyndman and Shang (2009) we obtain

$$y_t(x_i) = \hat{\mu}(x_i) + \sum_{k=1}^{K} B_{t,k}\phi_k(x_i) + \hat{\sigma}(x_i)\hat{\epsilon}_{t,i}. \tag{22}$$

Here $\hat{\sigma}(x_i)\hat{\epsilon}_{t,i}$ is the error term where $\hat{\sigma}(x_i)$ depends on the maturity date and $\hat{\epsilon}_{t,i}$ is an overall error. As stated before this is an independent univariate time series model hence we can apply regular multivariate forecasting on this series.

For the sake of this paper we use a slightly different and robuster model for our forecast. Because of time constraints and in the scope of this paper we leave the error term out of our model. The same applies to the weighted average function. This function is hard implement in our model. Hence we use a mean average of the interest at a certain maturity date, f_i. So for example the mean interest over our time period of yield bonds with maturity date of three months. Therefore our forecast model is more robust and can be written as

$$y_t(x_i) = \hat{\mu}(i) + \sum_{k=1}^{K} B_{t,k}\phi_k(x_i)$$

where $\hat{\mu}(i) = \frac{1}{n}\sum_{t=1}^{n} f_t(i)$.

6.2 Forecasting Yield curves using fPCA

We use the packages "forecast", "pracma" and "ftsa" where "ftsa" is developed by Shang et al. (2018) to apply their method. We first make a data frame using the equation 22. This can be transformed to an functional time series and we can than directly forecast this using the packages "ftsa".

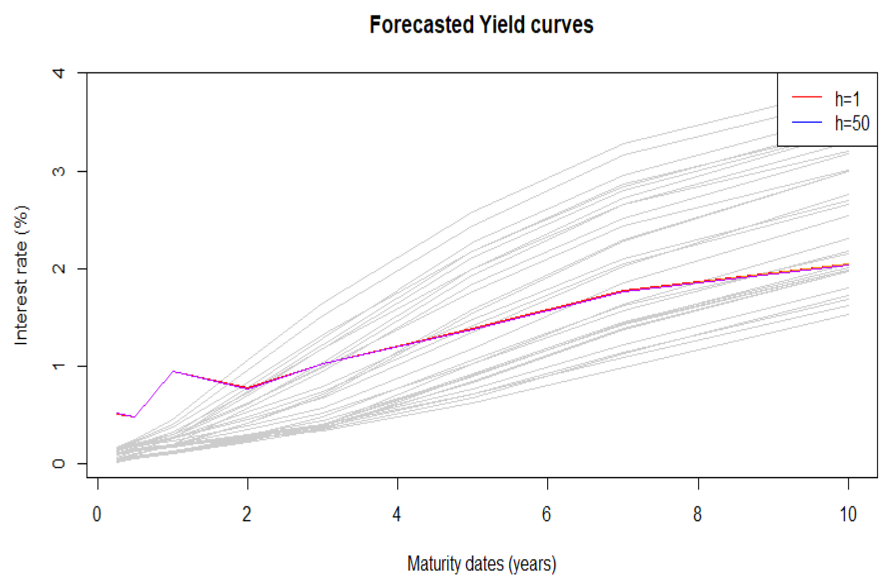

Figure 19: Fifty forecasted yield curves in comparison with thirty past yield curves.

The forecast goes to its mean very fast hence there is little variance in the forecasted curves. This could be explained by the small differences amongst the yield curves in our data. Furthermore our forecast is higher than the last two yield curves of our data. This is shown in figure 20. This shows that our intercept which is our mean function is too high. The rest of the yield curve might be skewed as a result. Here we can see that our forecast is too robust as we do not give these curves a bigger weight in our mean function but we use just the mean of the whole time period. Nevertheless we do see in figure 19 that they have the same trend as our data. They all have the same nod between the six and eight year maturity date where the curves flatten.

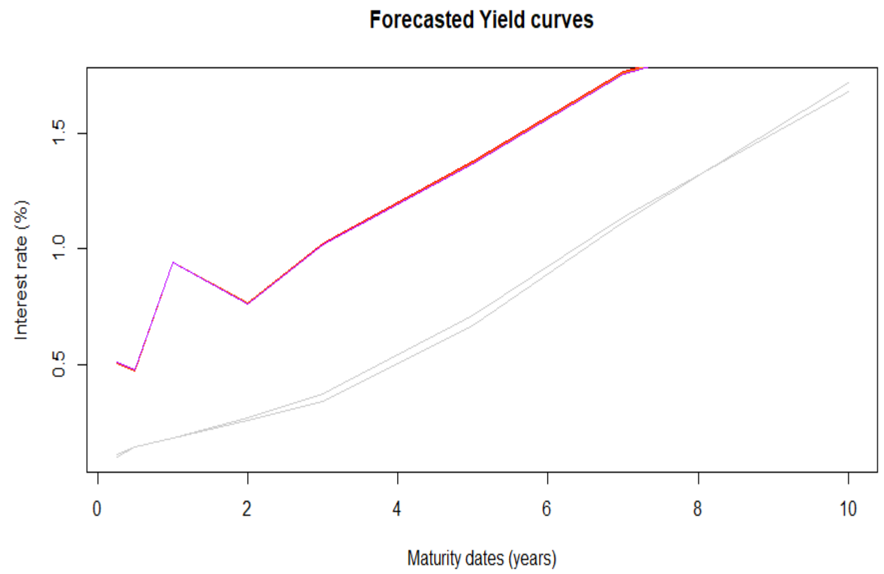

Figure 20: Fifty forecasted yield curves in comparison with the two last yield curves.

7 Conclusion and further work

We introduced functional data and showed a way to represent the data using functional basis systems. We explored PCA and fPCA and focused on a economic application, namely yield curves. To represent our data of yield curves we used B-splines. We discussed PCA on yield curves and using this method certain statements can be done regarding the first three principal components and to what they relate to, namely.

PC1: Parallel shift of yield curves/level/intercept on the y-axes;

PC2: Tilting of the curve/slope;

PC3: The flexing/curvature of the curve.

We have validated these claims using fPCA and we had similar values for the principal components. So most of the variance is explained by the first principal component. The second and third together explain the rest of the variance. Hereby we showed that we can expand our functional analyses as we have similar results as multivariate analyses. Although the statements are verified using fPCA we did discretize the function domain in a grid of points before applying the fPCA. Thus one could question our methods as truly being FDA and the effective use of it. Nevertheless this opens the way for more research as fine tuning a new analyses will eventually improve this method. We use yield curves which are uncorrelated but it could be very interesting to validate that we can apply the same strategy to dependent data.

We also sought the impact of extreme economical scenario's on our principal components. First we will discuss the crisis scenario. Because the principal components where very similar as in the overall scenario we can hold our claim which we did for the overall scenario regarding the interpretation of the principal components. We have seen that in times of crisis that the second principal component explains significant more variance than in the overall situation. As the second principal component relates to the slope which means that the slope together with the intercept of the curve explain almost all the variance in the crisis scenario. The third principal component did not have any significant changes. The explained variance is bigger but still negligible with respect to the first and second principal component.

The fPCA in the growth scenario gave very different results than our overall and crisis scenarios. The first principal component does not relate to the intercept. The growth model implies that the first principal component relates to the curve of the yield curve. As a result with this assumption we can link the second principal to the intercept. For the third principal component we can still hold our claim that it relates to the curvature. Hence the first and second principal components in the rapid growth scenario have interchanged their interpretations but we should be careful with these conclusions. As the statements are not very strong and might be proved wrong. Hence to make any strong and valid statements about the growth scenario further research is needed.

Furthermore the domain of our data of the scenarios is not the same and is significantly smaller than in the overall situation. Hence our results and conclusions do not have the same significance as statements over the overall data and might be skewed. Thus for further research one should use data from other countries and cross check the claims and conclusions made in this paper.

We investigated the forecasting of functional time series with a weighted average function. We saw that our forecast converged to the mean fast but further it was relatively well fitted. We used a robust and simplified model for our forecast. For future work one could explore what kind of model should be fitted to the error term. Furthermore using the weighted average function as in Hyndman and Shang (2009) could also improve the fit even more.

References

Choudhry, M. (2019). *Analysing and interpreting the yield curve.* John Wiley & Sons.

Cochrane, J. H. (2009). *Asset pricing: Revised edition.* Princeton university press.

Cœuré, B. (2016). Assessing the implications of negative interest rates. In *Speech at the yale financial crisis forum, yale school of management, new haven* (Vol. 28, p. 2016).

De Boor, C. (2001). A practical guide to splines 2001. *Appl. Math. Sci.*

Erbas, B., Hyndman, R. J., & Gertig, D. M. (2007). Forecasting age-specific breast cancer mortality using functional data models. *Statistics in Medicine*, *26*(2), 458–470.

European Central Bank. (2021). *Euro area yield curves.*

Forbes. (2020). *Longest economic expansion in united states history.* ://www.forbes.com/sites/davidmarotta/2020/01/21/longest-economic-expansion-in-united-states-history/?sh=7a0dd0cc62a2.

Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications* (Vol. 200). Springer Science & Business Media.

Hyndman, R. J., & Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, *24*(3), 323–342.

Hyndman, R. J., & Shang, H. L. (2009). Forecasting functional time series. *Journal of the Korean Statistical Society*, *38*(3), 199–211.

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

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.

Kneip, A., & Gasser, T. (1992). Statistical tools to analyze data representing a sample of curves. *The Annals of Statistics*, 1266–1305.

Liu, X., & Müller, H.-G. (2004). Functional convex averaging and synchronization for time-warped random curves. *Journal of the American Statistical Association*, *99*(467), 687–699.

Mehl, A. (2009). The yield curve as a predictor and emerging economies. *Open Economies Review*, *20*(5), 683.

Panaretos, V. M., & Zemel, Y. (2016). Amplitude and phase variation of point processes. *The Annals of Statistics*, *44*(2), 771–812.

PWC. (2019). *What us life insurers should do about low and negative interest rates.* `https://oceanexplorer.noaa.gov/okeanos/explorations/ex1907/dailyupdates/nov7/nov7.html`.

Ramin Nakisa. (2021). *Principal component analysis.* :http://nakisa.org/bankr-useful-financial-r-snippets/principal-component-analysis/.

Ramsay, & Silverman. (2005). *Functional data analysis.*

Ramsay, J., Hooker, G., & Graves, S. (2009). Introduction to functional data analysis. In *Functional data analysis with r and matlab* (pp. 1–19). Springer.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, *14*(1), 1–17.

Rao, C. R. (1987). Prediction of future observations in growth curve models. *Statistical Science*, 434–447.

Rognlie, M. (2016). What lower bound? monetary policy with negative interest rates. *Unpublished manuscript*.

Schumacher, J. M., et al. (1981). *Dynamic feedback in finite-and infinite-dimensional linear systems* (Vol. 143). Mathematisch Centrum Amsterdam.

Shang, H. L., Hyndman, R. J., & Shang, M. H. L. (2018). Package 'fds'. *Journal of the Korean Statistical Society*, *38*(3), 199–221.

Tang, R., & Müller, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika*, *95*(4), 875–889.

The Guadrian. (2011). *Global financial crisis: five key stages 2007-2011.* :https://www.theguardian.com/business/2011/aug/07/global-financial-crisis-key-stages.

Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika*, *23*(1), 19–23.

Wright, J. H. (2006). The yield curve and predicting recessions.

Yao, F., Müller, H.-G., & Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American statistical association*, *100*(470), 577–590.