

**MEF UNIVERSITY**

**RISK PARAMETER CALCULATION  
USING PRINCIPAL COMPONENT ANALYSIS OF  
YIELD CURVES: THE CASE OF BORSA ISTANBUL  
FIXED INCOME MARKET**

**Capstone Project**

**Hayrettin Konuk**

**İSTANBUL, 2017**



**MEF UNIVERSITY**

**RISK PARAMETER CALCULATION  
USING PRINCIPAL COMPONENT ANALYSIS OF  
YIELD CURVES: THE CASE OF BORSA ISTANBUL  
FIXED INCOME MARKET**

**Capstone Project**

**Hayrettin Konuk**

**Advisor: Asst. Prof. Dr. Levent Guntay**

**İSTANBUL, 2017**

## MEF UNIVERSITY

Name of the project: Risk Parameter Calculation Using Principle Component  
Analysis of Yield Curves: The Case of Borsa Istanbul Fixed Income Market  
Name/Last Name of the Student: Hayrettin Konuk  
Date of Thesis Defense: 16/10/2017

I hereby state that the graduation project prepared by Hayrettin Konuk has been completed under my supervision. I accept this work as a "Graduation Project".

16/10/2017  
Asst. Prof. Dr. Levent Guntay

I hereby state that I have examined this graduation project by Hayrettin Konuk which is accepted by his supervisor. This work is acceptable as a graduation project and the student is eligible to take the graduation project examination.

16/10/2017  
Prof. Dr. Özgür Özlük

Director  
of  
Big Data Analytics Program

We hereby state that we have held the graduation examination of Hayrettin Konuk and agree that the student has satisfied all requirements.

### THE EXAMINATION COMMITTEE

Committee Member

1. Asst. Prof. Dr. Levent Guntay

2. Prof. Dr. Özgür Özlük

Signature

.....

.....

## **Academic Honesty Pledge**

I promise not to collaborate with anyone, not to seek or accept any outside help, and not to give any help to others.

I understand that all resources in print or on the web must be explicitly cited.

In keeping with MEF University's ideals, I pledge that this work is my own and that I have neither given nor received inappropriate assistance in preparing it.

---

Hayrettin Konuk

16.10.2017

Signature

# EXECUTIVE SUMMARY

## RISK PARAMETER CALCULATION USING PRINCIPAL COMPONENT ANALYSIS OF YIELD CURVES: THE CASE OF BORSA ISTANBUL FIXED INCOME MARKET

Hayrettin Konuk

Advisor: Asst. Prof. Dr. Levent Güntay

OCTOBER , 2017, 23 pages

To enable a trustworthy clearing operation, clearinghouses require conservative margins to avoid the risk of incurring a loss in case one counterparty defaults. When margin requirements for fixed income instruments are calculated, yield curves of each instrument are stressed using their first three principal components. All instruments in an account are then evaluated against each stressed yield curve and the margin requirement is calculated as the difference of the combined value of these instruments calculated with the worst of the stressed yield curves between their combined values calculated with related unstressed curves. The aim of this project is to construct a tool for applying principle component analysis (PCA) on daily zero coupon yield curve of Turkish Treasury Securities. The analysis employs a yield curve panel data set obtained consisting historical zero coupon yield curves. The data set includes interest rates of 60 different maturities varying between overnight and 15 years and 1250 daily observations between December 2010 and December 2015. The result of this analysis provides a method that could be run at the end of each clearing day to determine the major components of the yield curve such as level/height, slope and curvature that describes at least 95% of the variation in interest changes and subject to stress shocks.

**Key Words:** exploratory data analysis, principle component analysis, yield curves.

# ÖZET

## RISK PARAMETER CALCULATION USING PRINCIPAL COMPONENT ANALYSIS OF YIELD CURVES: THE CASE OF BORSA ISTANBUL FIXED INCOME MARKET Hayrettin Konuk

Tez Danışmanı: Yrd. Doç. Dr. Levent Guntay

EKİM, 2017, 23 sayfa

Takasın güvenli bir şekilde tamamlanmasının sağlanabilmesi için, takas kurumunun herhangi bir karşı tarafının temerrüdü halinde oluşacak zararlar nedeniyle maruz kalacağı riskleri karşılayabilmek için gerekli teminat tutarını makul bir muhafazakârlıkta temin ediyor olması gerekmektedir. Sabit getirili enstrümanların teminat gereksinimlerinin hesaplanmasında, her bir enstrümanın kendi fiyatına şok vermek yerine, her bir enstrümanın verim eğrisinin ilk üç temel bileşenine şok verilir. Bir hesapta bulunan tüm enstrümanlar şok verilmiş farklı verim eğrileri ile değerlendirilir ve şoklanmış verim eğrilerinin en kötülerinin ürettiği toplam değer ile şoklanmamış eğriler ile üretilmiş toplam portföy değeri arasındaki fark teminat gereksinimi olarak hesaplanır. Bu projenin amacı, Türk Hazine Tahvilleri'nden oluşan kuponsuz verim eğrileri üzerinde günlük keşifsel veri analizi akabinde temel bileşenler analizi yapacak bir araç üretmektir. Analizde, tarihsel kuponsuz verim eğrilerinden oluşmuş verim eğrisi panel veri seti kullanılmaktadır. Veri seti, Aralık 2010 ve Aralık 2015 arasındaki 1250 günlük gözlem için, gecelik ve 15 yıllık altmış farklı vade dilimine ait faizlerden oluşmaktadır. Analiz sonucu ise her takas günü sonunda çalıştırılarak seviye, eğim ve büküm gibi verim eğrilerindeki getiri değişimlerinin en az %95'ini açıklayan, şok uygulanacak ilk üç temel bileşenin tespit edilmesini sağlayan bir metodoloji üretmektir.

**Anahtar Kelimeler:** keşifsel veri analizi, temel bileşenler analizi, verim eğrileri.

## TABLE OF CONTENTS

Academic Honesty Pledge .....	vi
EXECUTIVE SUMMARY .....	vii
ÖZET .....	viii
TABLE OF CONTENTS.....	ix
1. INTRODUCTION .....	1
2. LITERATURE REVIEW .....	2
2.1. Literature.....	2
2.2. Term Structure, Yield and Yield Curve Models.....	3
2.3 Principle Component Analysis on Yield Curves .....	4
2.3.1. Principal Component Analysis .....	4
2.3.2. Methodology of Principle Component Analysis .....	5
2.3.3. Steps of a PCA Process.....	7
2.3.4. Interpretation of the First Three Components.....	7
3. EXPLORATORY DATA ANALYSIS OF TURKISH TREASURY YIELD CURVE ...	8
3.1. Data.....	8
3.2. Preprocessing and Exploratory Data Analysis.....	9
4. PRINCIPLE COMPONENT ANALYSIS.....	11
4.1. Results of Principle Component Analysis on Yield Curves .....	12
4.2. Interpretation of the Results.....	13
5. CONCLUSION.....	14
REFERENCES .....	16
APPENDIX.....	18
TABLES AND GRAPHS .....	23



# 1. INTRODUCTION

Takas Istanbul provides central counterparty clearinghouse (CCP) services for listed derivatives and cash equity markets of Borsa Istanbul. It will also be the CCP for Borsa Istanbul Fixed Income Market. Takas Istanbul is also planned to undertake the counterparty credit risk of market participants by requiring margin collateral from them. Each clearing member of a clearinghouse that has cleared its trades needs to pledge collateral, the size of which relates to the risk in its positions. The main risk factors involved in the fixed income markets stem from the yield curve changes. Yield curves are sometimes modelled by the first three principle components that are closely associated with the level, slope and curvature of the yield curve.

While margining fixed income instruments, instead of stressing each instrument's individual price, yield curves are stressed using their first three principal components. All instruments in an account are then evaluated against each stressed yield curve and the margin requirement is given as the combined value of these instruments calculated with the worst outcomes of the stressed yield curves.

A counterparty could be monitoring the explanatory power of these 3 principal components on a daily basis. On a monthly basis it could evaluate and, if needed, update each yield curve's first three principal components together with a risk parameter that decides how much of a principal component will be used to simulate the stressed curves.

As a result, the objective of this project is to construct a methodology that determines and updates the principle components at the end of each day. We also plan to verify whether the explanatory power of the first 3 principal components explain the majority of the variation in the yield curve.

This study is organized as follows. Section 2 reviews the related literature regarding principal component analysis of yield curves, second subsection includes a very short brief on the term structure of interest rates and yield curves, and third subsection reviews the principle component analysis. Section 3 describes the data source, structure, preparation and exploratory data analysis. And Section 4 provides the application of principle component analysis (PCA, henceforth). First sub-section of Section 4 presents the empirical results of the principal component analysis of the historical yield curve data and the second sub-section of section 4 is the interpretation of the results. Section 5 concludes.

## 2. LITERATURE REVIEW

### 2.1. Literature

The main center of the works that have been processed until these days mainly focus on the structure of interest rates (Litterman and Scheinkman, 1991; Dai and Singleton, 2000). Cochrane and Piazzesi (2005) support the literature by arguing that the structure of interest rates demonstrate that a linear combination of the forward rates is the base factor for figuring out the excess returns on bonds in all maturities. To develop an outline that can be used as a method to develop a base scenario of an extensive historical term structure variation, many scholars attempted to gather movements in principal components. Stress scenarios put forward by Lorethan (1997) and Rodrigues (1997) are examples of these efforts.

Jamshidian and Zhu (1997) pioneered the Principal Component Analysis in analyzing fixed income scenarios. They aim to get a discrete approximation of the portfolio value distributions. On the other hand, Loretan (1997) and Frye (1997) concentrate on the Principal Component Analysis in the context of a VaR methodology. These two scholars mainly figure out the VaR of a fixed income portfolio. Explanation of principal component based scenarios is the main step in the calculation. In this step they define separate “stress” in each of the directions constituted by the principal components  $s$  and “combined” stress as linear combinations of the PCs. A small set of large prefixed stress factors are used in these studies. One of them is using the 2.33 times the PCA standard deviation which corresponds to the ninety-ninth percentile VaR.

Abdymomunov and Gerlach (2014) emphasize some limitations of the PCA method. These limitations mainly appear in stress testing interest rate exposures. The main argument is that PCA creates economically unreasonable scenarios.

Before PCA is applied to financial series, it is a reasonable process to understand whether PCA is in fact a meaningful procedure given the distributional properties of the data. Controlling whether the data is stationarity is required to be aware of the presence of stochastic trends. Different studies point out that interest rates could be non-stationary whereas first differencing usually achieves stationarity (Niffiker et. al, 2000. Lardic et al., 2003). Lardic et al. (2003) also highlight that the original variables should be centered and variance-reduced and they amount to using the correlation matrix of the changes. Moreover, more precise results could be obtained if daily data could be employed for the analysis.

## 2.2. Term Structure, Yield and Yield Curve Models

The term structure of interest rates is one of the most commonly used method of pricing bonds. The term structure of interest rates explains the relationship between the yields and maturities of a group of bonds with the same credit rating. Mainly, the term structure refers to risk-free Treasury securities but it can also be referred to riskier securities. The term structure of interest rates is also called the “yield curve”. Normally plotted on a graph as a curve showing representations of various bond yields for progressively longer maturities is the main property of the term structure that gives this synonym to it. Generally it is presumed that the bond is a zero-coupon bond security with a no-call provision. An example of the term structure of interest rates and the corresponding yield curve plot - based on our calculations- is provided in Table 1 and Figure 1:

MATURITY	YIELD
OverNight (O/N)	12.11%
2 year	11.76%
4 years	11.59%
6 years	11.46%
8 years	11.35%
10 years	11.21%
12 years	11.14%
14 years	11.06%

Table 1: Interest Rates per Maturities

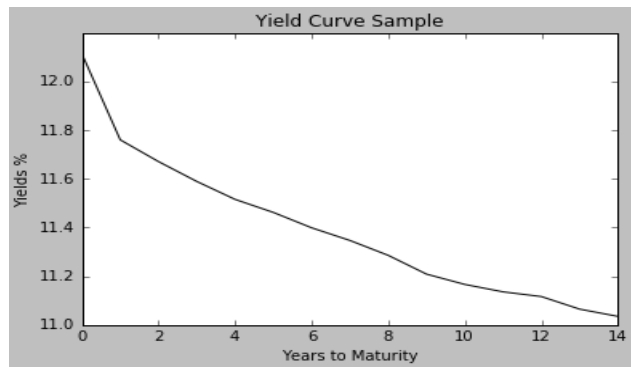


Figure 1: Treasury Zero Coupon Yield Curve on 17.12.2015

We can observe changes in the shape and level of the yield curve over time. When economic activity is expected to accelerate in the future, the yield curve tends to become steeper. If the economy is expected to slow down in the future, the yield curve tends to become flatter, since future rates are expected to be lower than they are now. The yield to maturity (YTM) means the equality of the rate of interest at which the market price of a bond between the present values of its expected future cash flows.

Upward and downward movements in the yield curve in different periods of time are the precursor of profits and losses for most financial institutions. The traditional “full models” of yield curves, treat every point of the curve as a random variable in its own right. On the other hand, in order to evaluate their resilience to yield curve movements, most firms

are on the side of “component models”. Polynomials, variance matching and principle components are the result of decomposing a whole model into components and it can be found various ways to apply decomposition process.

### **2.3 Principle Component Analysis on Yield Curves**

The yield curve on a particular date describes variations in interest rates as a function of the term to which a deposit is committed. Due to this, movements in the yield curve from one period to the next are important drivers of profits and losses of the financial institutions. Principal component analysis is the simplest one among the many orthogonalization techniques. These techniques turn a set of correlated variables into a set of uncorrelated variables. For financial risk management applications, principal component analysis is in demand at the highest rate. Independent from the high or low correlation, PCA can be used for each and every set of stationary time series. However, the best result can be taken from the highly correlated systems. A set of zero coupon returns of different maturities can be an example of this.

#### **2.3.1. Principal Component Analysis**

Principal component analysis (PCA) is a well-known statistical technique and also called a useful machine learning technique that has found application in fields such as face recognition and image compression, and is a common technique for finding patterns in data of high dimension (Smith, 2002). Other popular applications of PCA include exploratory data analyses and de-noising of signals in stock market trading, and the analysis genome data and gene expression levels in the field of bioinformatics (Raschka, 2015).

PCA is an unsupervised method. So, class labels are not taken into consideration while identifying underlying dimensions that shows the correlations between a set of factors such as “Years to Maturities”. This information ensures a new smaller set of uncorrelated factors to replace the original set of correlated factors as the extracted features in subsequent multivariate analysis.

Each principal component is the eigenvector of the covariance/correlation matrix of the input factors with the largest eigenvalue that maximizes the variance. Here the aim is maximizing the variance with the minimum number of components. Accordingly, PCA brings feature extraction and dimension reduction together in just one step.

In machine learning, source of the complexity and time consumption with the data size constitute is number of the input dimensions. For the sake of memory usage and computation, reducing dimension can be applied as an efficient way (Alpaydm, 2010).

In the light of the facts referred above, it's an extremely powerful machine learning tool for reducing the dimensions of large, highly correlated systems. For instance, in order to get a close range approximation to the changes in each rate in a system of sixty different interest rates, only three principal components are sufficient. Moreover, due to the diagonal covariance matrix, the principal components are not correlated. Hence, determining only three variances instead of thousands of different variances and covariances is the main task. As a result, it is understood evidently that PCA is a computationally efficient method.

Studies often summarize term structures (yield curves) by a small set of linear combinations of yields, so our analysis doesn't require the information about the class labels. Due to this reasons, researchers don't prefer a supervised dimension reduction technique such as Linear Discriminant Analysis or a nonlinear dimension reduction technique such as Kernel Principal Component Analysis for yield curve/term structure analysis.

### **2.3.2. Methodology of Principle Component Analysis**

Principal Component Analysis (PCA) is a well-known statistical method from multivariate analysis used in managing and explaining interest rate risk. Principle components are defined by explaining the largest proportion of the overall variance of a dataset without any overlap. Another definition for them can be (uncorrelated) axes creating a new coordinate system. The new axes are called as loadings. If an axis has more explanatory power for the dataset, the value of the corresponding weight will be larger.

While component matrix demonstrates the linear correlations between the original fields in the rows, it shows the derived components in the columns. The correlations among the components and the original inputs are called loadings; they are typically used for the interpretation and labeling of the derived components (Tsiptsis and Chorianopoulos, 2009).

To eliminate the complexity of a dataset, selecting the highest weights PCs can be a solution. By this means the original dataset can be transformed into new axes. These new coordinates are called as scores. The original dataset creates three outputs: the loadings, their associated weights and their scores. The structural changes of dataset and also its potential dislocations are emphasized here.

The eigenvalue is a measure of the variance that each component accounts for. To decide on which components are kept the eigenvalue criterion is the most widely used criterion. The main argument is that a component should be considered insignificant if it does worse than a single field. There is one unit of standardized variance in each single field. So the components with eigenvalues below 1 are not extracted. Extraction of the components are realized according to descending order of importance. In other words, the largest part of the variance is on the first one of the original fields. Extraction stops at component which has an eigenvalue below the threshold of 1 (Credit Suisse, 2012).

Briefly, PCA analyzes the covariance matrix of the dataset in order to capture as much variance as possible, with the fewest PCs possible. Axes created after this procedure namely the loadings will each have a certain length and an orientation. Respectively the eigenvectors and eigenvalues of the diagonalized correlation matrix of the database have found out this. The standard deviation of the scores is called as the eigenvalue. Eigenvectors on the other hand stand for the loadings. If the vector is longer meaning that the eigenvalue is larger, it will have more information and the associated PC will play a more important role.

Spectral decomposition of a covariance matrix or a correlation matrix is the main base of PCA. If PCA is carried out on a correlation matrix, correlations of returns are the only factors that affect the results. However if the input to PCA is a covariance matrix, the volatility of the returns as well as the correlations of returns will shape the results. No general relationship is discovered between the principal components of a covariance matrix and those of its correlation matrix (Alexander, 2008).

Especially the “Finance” literature uses the correlation matrix rather than the covariance matrix. However, the eigen decomposition of the covariance matrix (if the input data was standardized) yields the same results as an eigen decomposition on the correlation matrix, since the correlation matrix can be understood as the normalized covariance matrix (Raschka, 2015).

Also the principal components are orthogonal, and therefore additive and statistically independent. For the risk management purposes, additivity is important. Because it allows evaluation of the impact of even one unit of added parallel shift risk to an existing position. Since it allows the factors to be managed separately, say to hedge a parallel shift

without having to think about its effect on the other factors as discussed in Niffikeer et al. (2000), statistical independence is also important.

### **2.3.3. Steps of a PCA Process**

Step 1: Centre and standardize data (normalizing)

Step 2: Calculation of the covariance matrix  $\Sigma$

Step 3: Finding the Matrix P such that  $P\Sigma P'$  is a diagonal matrix

Step 4: Selection of the eigenvectors (=meaningful components)

Step 5: Creation of the reconstructed variables (multiplication of each PC coefficient vector by the corresponding score + add back in the mean of the data)

Step 6: Comparison of the old and the new datasets

### **2.3.4. Interpretation of the First Three Components**

As discussed in Alexander (2009) in a perfectly correlated system of returns on financial assets or changes in interest rates the elements of the first eigenvector are equal. If it is needed to say in more clear way, the more highly correlated the system the more similar the values of the elements of the first eigenvector. Hence, the first principal component captures a common trend in assets or interest rates. This means that when the first principal component changes but the other components are fixed in a specific time, then the returns (or changes in interest rates) all move by roughly the same amount. That is why the first component is generally called as the trend component.

If the system has no natural ordering then it is not possible to talk about intuitive interpretation of the second and higher order principal components. But if the system is ordered, such as a set of interest rate changes of different maturities or a set of returns on futures of different maturities, a change in slope of the term structure is captivated by the second principal component. Then the elements of the second eigenvector are decreasing (or increasing) in magnitude, so that when the second principal component alters while the other components fixed then the returns (or changes in interest rates) move up at one end of the term structure and down at the other end. That is why the second component is also called as the tilt component.

Likewise, the elements of the third eigenvector are usually decreasing (or increasing) and then increasing (or decreasing) in magnitude. Thus if the third principal component changes when the other components are fixed, then the returns (or changes in interest rates)

move up (or down) at both ends of the term structure and down (or up) in the middle. That is why the third component is called as the curvature or convexity component. Higher order principal components have similar movement analysis described by cubic polynomials (fourth component), quartic polynomials (fifth component) and so on.

Also estimates of principal components based on historical data can be quite unstable. Some simple rules of thumb – such as taking moving averages of our principal components – is a method to reduce instability. Using too many principal components also have some serious consequences in applying them. Sometimes the aim is just to add principal components representing stable relationships to get a more accurate forecast. However, this addition can take the process to a point where additional principal components merely lead to the model tracking noise and weaken the forecasting ability of our model.

### **3. EXPLORATORY DATA ANALYSIS OF TURKISH TREASURY YIELD CURVE**

Before applying the PCA analysis, in this section we first define and inspect the yield curve over a period time and make qualitative observations.

#### **3.1. Data**

The source of the data is Bloomberg. Our analysis is built on the data set consisting historical zero coupon yield curves including interest rates belongs to 60 different maturities between overnight and 15 years for at least a 5 year look back period between December 2010 and December 2015 of 1250 daily observations.

Before applying the PCA on the interest rate data, initially the missing values filled by linear interpolation and we provide an exploratory data analysis. Afterwards, daily differences (in basis points) of the historical yields on each maturity ladder are calculated and these returns are scaled for normalizing.



### 3.2. Preprocessing and Exploratory Data Analysis

Before exploring the data, missing values filled by linear interpolation. Afterwards, in Figure 2 we plot the line graph of each column that represents the yield to maturities of each days on maturity ladders for presenting the historical volatility of 60 different ladders.

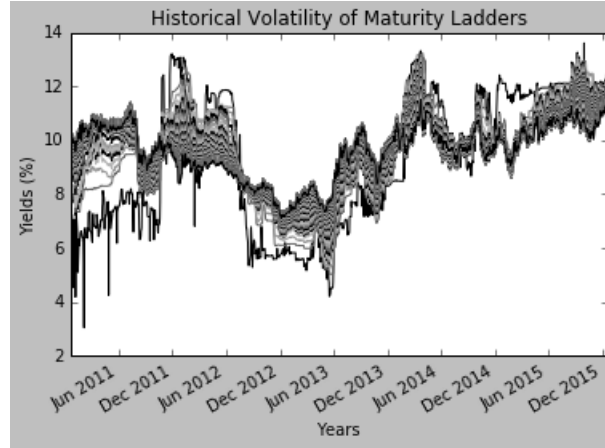


Figure 2 : Historical interest rate levels of each maturity ladder

In Figure 2 each line represents interest rate levels of each different maturity ladder -starting from O/N to 15 years- time series data in our observation period. Dark bold black line represents the level of O/N rates looks more volatile than the others. Following this presentation, we transpose the data matrix to draw the line graph of each historical yield curves.

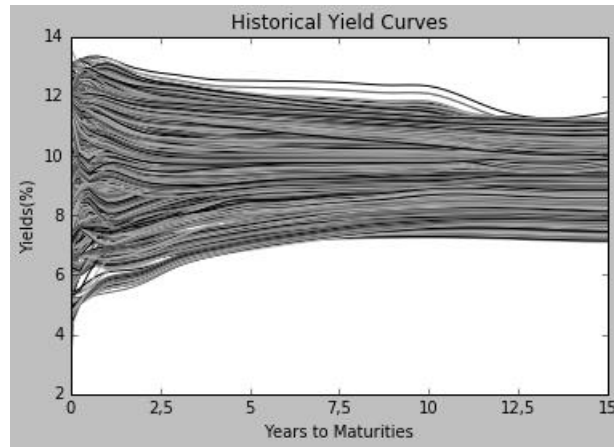


Figure 3: Yield Curves

The yield curve paths in Figure 3 show that first, Yield Curves are generally moving together in successive days. Second, risk free rates are very volatile for short maturities that we'll also see from the following graph. Third, although the short rates are differentiated in

a wide range, yield ranges are shrinking at longer maturities; such as overnight (O/N) yields are fluctuating between 3% and 13.6% range while the yields are fluctuating between 7% and 11.5% range at the 15-year maturity, in our period of observation.

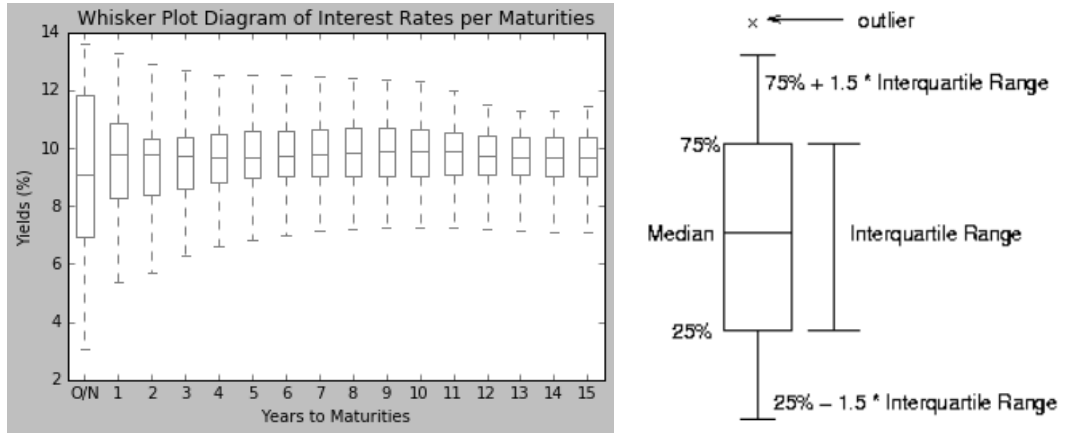


Figure 4: Box-Whisker Plot Graph of interest rate levels of different maturities

Box-and-whisker plot graphs in Figure 4 are created for monitoring both rate level and dispersion during the period of observation. In a box and whiskers plot, the center line in the box is the median, the edges of the box are the lower and upper quartiles (25th and 75th percentile), whilst the whiskers highlight the last data point within a distance of 1.5 x (upper – lower quartile) from the lower and upper quartiles. Values outside the whiskers are plotted separately as dots and suspected to be outliers.

Also we can see the descriptive statistics of the different maturities form the below Table 2 that confirms our interpretations above such as the volatility of O/N rates with its standard deviation value of 2.585 % which is higher than the other maturities.

Descriptive Statistics of Interest Rates Per Maturities																
	ON	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
count	1250	1250	1250	1250	1250	1250	1250	1250	1250	1250	1250	1250	1250	1250	1250	1250
mean	9.26	9.60	9.48	9.53	9.58	9.64	9.68	9.72	9.74	9.76	9.78	9.72	9.63	9.56	9.54	9.57
std	2.59	1.78	1.55	1.41	1.30	1.23	1.18	1.15	1.11	1.10	1.08	1.03	0.97	0.95	0.96	0.99
min	3.04	5.39	5.71	6.29	6.64	6.82	7.01	7.17	7.22	7.24	7.24	7.23	7.20	7.16	7.12	7.10
25%	6.94	8.29	8.38	8.62	8.82	8.96	9.03	9.04	9.04	9.06	9.06	9.09	9.09	9.06	9.02	9.01
50%	9.08	9.80	9.78	9.71	9.69	9.66	9.73	9.80	9.84	9.88	9.92	9.88	9.75	9.66	9.66	9.70
75%	11.85	10.88	10.33	10.40	10.49	10.57	10.62	10.66	10.67	10.68	10.66	10.56	10.45	10.37	10.35	10.39
max	13.60	13.28	12.88	12.70	12.55	12.52	12.50	12.48	12.42	12.36	12.34	11.97	11.50	11.27	11.27	11.45

Table 2: Descriptive Statistics of Interest Rate Levels at Different Maturities

In addition to the descriptive statistics above, we can investigate the correlations of interest rate levels of different maturity ladders with each other.

Correlation Matrix of Maturity Based Interest Rates																
	ON	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ON	1.00	0.89	0.82	0.77	0.72	0.68	0.64	0.61	0.58	0.56	0.54	0.52	0.50	0.49	0.50	0.52
1	0.89	1.00	0.98	0.95	0.92	0.88	0.85	0.81	0.78	0.76	0.74	0.73	0.71	0.70	0.71	0.73
2	0.82	0.98	1.00	0.99	0.97	0.95	0.92	0.90	0.88	0.86	0.84	0.83	0.81	0.80	0.81	0.82
3	0.77	0.95	0.99	1.00	0.99	0.98	0.96	0.94	0.92	0.90	0.89	0.88	0.86	0.85	0.86	0.87
4	0.72	0.92	0.97	0.99	1.00	0.99	0.98	0.97	0.96	0.94	0.93	0.92	0.91	0.90	0.90	0.91
5	0.68	0.88	0.95	0.98	0.99	1.00	1.00	0.99	0.98	0.97	0.96	0.95	0.94	0.93	0.93	0.94
6	0.64	0.85	0.92	0.96	0.98	1.00	1.00	1.00	0.99	0.99	0.98	0.97	0.96	0.95	0.95	0.96
7	0.61	0.81	0.90	0.94	0.97	0.99	1.00	1.00	1.00	1.00	0.99	0.98	0.97	0.96	0.96	0.97
8	0.58	0.78	0.88	0.92	0.96	0.98	0.99	1.00	1.00	1.00	1.00	0.99	0.98	0.96	0.97	0.98
9	0.56	0.76	0.86	0.90	0.94	0.97	0.99	1.00	1.00	1.00	1.00	0.99	0.98	0.97	0.97	0.98
10	0.54	0.74	0.84	0.89	0.93	0.96	0.98	0.99	1.00	1.00	1.00	1.00	0.98	0.97	0.97	0.98
11	0.52	0.73	0.83	0.88	0.92	0.95	0.97	0.98	0.99	0.99	1.00	1.00	0.99	0.99	0.99	0.99
12	0.50	0.71	0.81	0.86	0.91	0.94	0.96	0.97	0.98	0.98	0.98	0.99	1.00	1.00	1.00	1.00
13	0.49	0.70	0.80	0.85	0.90	0.93	0.95	0.96	0.96	0.97	0.97	0.99	1.00	1.00	1.00	1.00
14	0.50	0.71	0.81	0.86	0.90	0.93	0.95	0.96	0.97	0.97	0.97	0.99	1.00	1.00	1.00	1.00
15	0.52	0.73	0.82	0.87	0.91	0.94	0.96	0.97	0.98	0.98	0.98	0.99	1.00	1.00	1.00	1.00

Table 3: Correlation Matrix of Interest Rate Levels at Different Maturities

#### 4. PRINCIPLE COMPONENT ANALYSIS

From a risk management perspective, when we are dealing with interest rates, we should take into consideration the interest rate level changes.

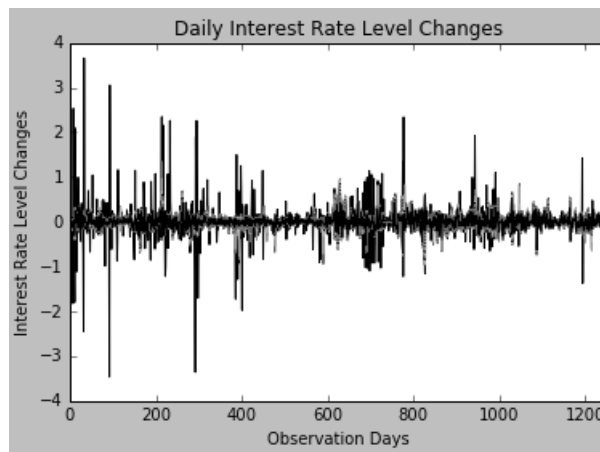


Figure 5: Daily interest rate level changes per maturity ladders

In this study, mean of the daily interest rate level differences -in basis points- is subtracted from the daily interest level differences for each maturity ladder and afterwards each observation is divided by the standard deviation of each maturity ladder for normalizing the observation data. Following this short preprocessing, principle component analysis (PCA) is applied<sup>1</sup> and below the empirical results are obtained.

#### 4.1. Results of Principle Component Analysis on Yield Curves

We applied the principal component analysis initially for the first ten components and receive the below results. According to the results of the analysis, the first three components explain 95.88% (equals to the sum of variance explained by three components; that is 86.01%, 7.02% and 2.85% respectively) of the total variance in daily yield curve changes.



Table 4: Importance of Principle Components

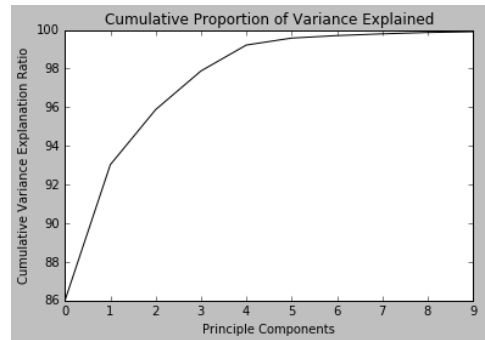


Figure 6: Explained cumulative variance ratio

Pursuant to the Kaiser’s criterion, we should only retain principal components for which the variance is above 1 when principal component analysis was applied to standardized data and due to the variances that we obtain, we should take into consideration the first four principal components. However, our approach in this study is similar to the treatment of confidence interval in value at risk calculations which we generally assume 95% is sufficient. Hence, the first three components are sufficient to explain the 95.6% of the total variance.

---

<sup>1</sup> We used the scikit-learn library of Python for carrying on the PCA analysis.

## 4.2. Interpretation of the Results

As it can be seen from the results, we achieve more than 95% explanation level with the first three principal components of the yield curve.

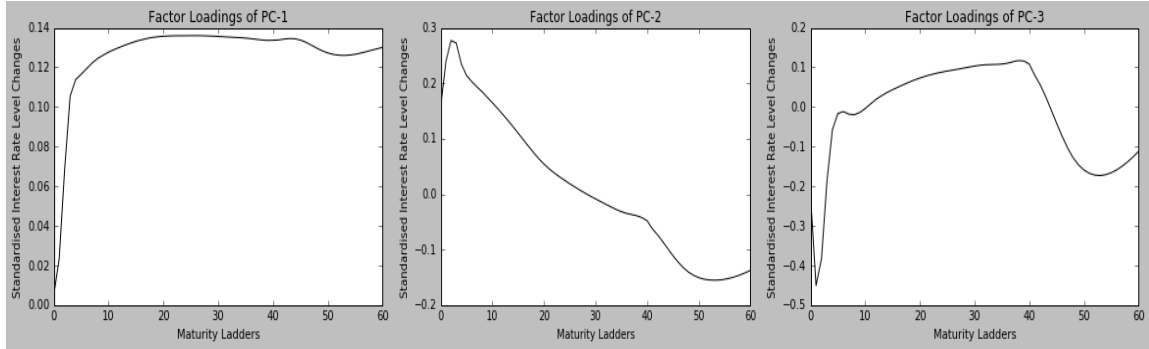


Figure 7: Factor Loading Graphs of First Three Components

Factor loadings of first principle component, which explains the 86% of the total variance, always represents positive coefficients for all of the maturity ladders, and also very volatile until to 2 year maturity and flattens at the following years. It causes rising interest rates at each maturity ladder. The factor loadings of second principle component represents positive coefficients on the short end while it represents negative coefficients on the long end following a decreasing coefficients. These two loadings represent the level and the slope of the yield curve with the third factor loading which represents the curvature effect. And these representations also show the characteristics of the yield curve that can be seen on the sample yield curve at the Figure 1.

First Three Factor Loadings per Maturities																
	O/N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
PC - 1	0.005	0.114	0.124	0.130	0.134	0.136	0.136	0.136	0.135	0.135	0.134	0.134	0.130	0.126	0.127	0.130
PC - 2	0.158	0.235	0.185	0.144	0.098	0.054	0.025	0.002	-0.018	-0.035	-0.048	-0.098	-0.141	-0.155	-0.152	-0.138
PC - 3	-0.224	-0.057	-0.019	0.019	0.050	0.073	0.088	0.098	0.107	0.111	0.109	-0.007	-0.128	-0.172	-0.159	-0.112

Table 5: Factor Loadings of First Three PCs

Since each principal component is a linear combination of the variables, total sum of the multiplication of each PC's loadings (for each maturity point) with their own standard deviation provides us the yield curve changes for each PC and each historical day.

According to figures that we obtain from our analysis, factor loading of first principle component for 2 years of maturity is 0.124 and this implies that a unit change in PC 1 causes

a 0.124 basis point increase in 2-years rate. If the 2-year rate is 10%, then it will become 10.01% due to a unit change. For the three factors, it causes a 0.29 ( $0.124+0.185-0.019=0.290$ ) basis point change for a unit change in 2-year rate and 10% will become 10.03%.

However, according to characteristics of the historical yield curves, we obtain the variances for each principal components. So instead of the unit changes we can apply the standard deviations of each principal component for determining the change in the interest rates. Thus the parallel shift factor (PC-1) changes the 2-year rate by 1 basis points ( $7.24 \times 0.124 = 0.9$ ); the slope factor (PC-2) changes the 2-year rate by 0.4 basis points ( $2.07 \times 0.185 = 0.38$ ) and the curvature factor(PC-3) changes the 2-year rate by -0.03 basis points ( $1.32 \times -0.019 = -0.03$ ). And also if we multiply the sum of these level changes (1.26) with  $\pm 2.33$  standard deviation which corresponds to the 1<sup>st</sup> and 99<sup>th</sup> percentiles of the normal distribution, we can obtain the Value at Risk as the  $\pm 2.93$  basis points ( $(0.9+0.38-0.03) \times (\pm 2.33)$ ) within the 99% significance level and we can assume that 10% 2-year rate will increase to %10.29 as the interest rate risk.

## 5. CONCLUSION

In this study we apply the Principal Components Analysis (PCA) to determine the factors that are main reasons for changes in the risk free term structure. PCA quantifies the movements of the yield curve in terms of three main factors: level, slope, and curvature. This study shows that the first three principal components explain 95.8% of the total variance of the changes in Turkish Treasury zero coupon interest rates. This result is in line with the results of studies from the fixed income markets of other countries. Also this study shows that the historical volatility of the term structure represents a  $\pm 2.93$  basis point change with a 99% significance level in 2-year rate which is indicative for Turkish fixed income market.

Since the first three principle components are sufficient for explaining the variance of interest rate changes; risk management become a matter of managing exposure to these factors. So, for Takas Istanbul as the CCP of fixed income market in Turkey, it will be sufficient to give stress shocks on to the first three principle components determined in this study for calculating the risk per maturity as the base for margin requirements for its members.



## REFERENCES

- Abdymomunov, A., & Gerlach, J. (2014). Stress testing interest rate risk exposure. *Journal of Banking & Finance*, 49, 287-301.
- Alexander, C. (2008). *Market Risk Analysis, Quantitative Methods in Finance* (Vol. 1). John Wiley & Sons.
- Alexander, C. (2009). *Market Risk Analysis, Value at Risk Models* (Vol. 4). John Wiley & Sons.
- Alpaydin, E. (2010). *Introduction to Machine Learning, Second Edition*. MIT Press.
- Cochrane, J. H., & Piazzesi, M. (2009). Decomposing the yield curve. University of Chicago working paper.
- Dai, Q., & Singleton, K. J. (2000). Specification analysis of affine term structure models. *The Journal of Finance*, 55(5), 1943-1978.
- Frye, J. (1997). Principals of risk: Finding value-at-risk through factor-based interest rate scenarios. *NationsBanc-CRT*, April.
- Jamshidian, F., & Zhu, Y. (1996). Scenario simulation: Theory and methodology. *Finance and Stochastics*, 1(1), 43-67.
- Lardic, S., Priaulet, P., & Priaulet, S. (2003). PCA of the yield curve dynamics: questions of methodologies. *Journal of Bond Trading and Management*, 1(4), 327-349.
- Litterman, R. B., & Scheinkman, J. (1991). Common factors affecting bond returns. *The Journal of Fixed Income*, 1(1), 54-61.
- Loretan, M. (1997). Generating market risk scenarios using principal components analysis: methodological and practical considerations. *Federal Reserve Board Working Paper*, 23-53.
- Niffikeer, C. I., Hewins, R. D., & Flavell, R. B. (2000). A synthetic factor approach to the estimation of value-at-risk of a portfolio of interest rate swaps. *Journal of Banking & Finance*, 24(12), 1903-1932.
- Raschka, S. (2015a). PCA in 3 steps  
[http://sebastianraschka.com/Articles/2015\\_pca\\_in\\_3\\_steps.html](http://sebastianraschka.com/Articles/2015_pca_in_3_steps.html), Retrieved on 9/3/2017.
- Raschka, S. (2015b). Python machine learning. Packt Publishing Ltd.
- Rodrigues, A. P. (1997). Term structure and volatility shocks. *Manuscript, Federal Reserve Board*, 500.



Smith, L. I. (2002). A tutorial on principal components analysis. *Cornell University, USA*, 51(52), 65.

Tsiptsis, K. K., & Chorianopoulos, A. (2011). *Data mining techniques in CRM: inside customer segmentation*. John Wiley & Sons.

## APPENDIX

### PYTHON SCRIPT

Created on Tue Sep 05 21:45:30 2017

@author: hkonuk

```
"""
```

```
import numpy as np
```

```
from sklearn.decomposition import PCA
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
from sklearn.preprocessing import scale
```

```
#import matplotlib.ticker as ticker
```

```
import numpy as np
```

```
#from matplotlib.ticker import MultipleLocator, FormatStrFormatter
```

```
import matplotlib as mpl
```

```
#matplotlib inline
```

```
#Importing the historical yield curve dataset
```

```
df=pd.read_excel('C:/Users/hkonuk/Documents/termpdata.xlsx','verim')
```

```
# Checking how many observation do we have for each row
```

```
print df.count()
```

```
# Filling NAs by linear interpolation
```

```
yields=df.interpolate()
```

```
# Checking how many observation do we have after changing NAs by interpolation
```

```
print yields.count()
```

```
# Plotting the line graph of each column that represents
```

```
# the yield to maturity of related days to maturity ladder
```

```
# for presenting the historical volatility of 60 different ladder.
```

```
yields.plot(legend=None, title='Historical Volatility of Maturity Ladders')
```

```
plt.xlabel('Years')
```

```

plt.ylabel('Yields (%)')
#plt.style.use('grayscale')
mpl.rc("figure", facecolor="white")

# Transposing the matrix to draw the line graph of each yield curve
yield_curves=yields.transpose()
yield_curves.plot(legend=False, title='Historical Yield Curves')
plt.xlabel('Years to Maturities')
plt.ylabel('Yields(%)')
plt.style.use('grayscale')
#mpl.rc("figure", facecolor="white")

# Creating box-and-whisker plots graph for monitoring
# both rate level and dispersion during the period of observation.
sf=pd.read_excel('C:/Users/hkonuk/Documents/termpdata_short.xlsx','verim')
syields=sf.interpolate()
color = dict(boxes='DarkGreen', whiskers='DarkOrange', medians='DarkBlue',
caps='Gray')
ax = syields.plot(kind='box', color=(0.5,0.5,0.5), sym='r+',title='Whisker Plot
Diagram of Interest Rates per Maturities')
plt.xlabel('Years to Maturities')
plt.ylabel('Yields (%)')
plt.style.use('grayscale')
#mpl.rc("figure", facecolor="white")

#Sample Yield Curve from 17.12.2015
mf=pd.read_excel('C:/Users/hkonuk/Documents/getiri_tablo.xlsx','mercan')
mf.plot(legend=None, title='Yield Curve Sample')
plt.xlabel('Years to Maturity')
plt.ylabel('Yields %')
mpl.rc("figure", facecolor="white")

```

```

#Summarize the descriptive statistics of interest rates belong to each maturity
ladder(column)
yields_stat=yields.describe()
print yields_stat
#Calculate correlation of interest rates between each maturity ladder(column) with
eachother
yields_cor=yields.corr(method='pearson')
#Plotting a heatmap reflecting the correlations.
plt.matshow(yields.corr())

# APPLYING PCA TO DAILY INTEREST RATE LEVEL DIFFERENCES
# Taking differences of each row to get the
#historical daily interest rate differences of each maturity ladder(column)
yc_fark = np.diff(yields, axis=0)
plt.plot(yc_fark)
plt.title('Daily Interest Rate Level Changes') # per Maturity Ladders')
plt.style.use('grayscale')
plt.xlim((0,1250))
plt.xlabel('Observation Days')
plt.ylabel('Interest Rate Level Changes')
#Scaling the values (normalizing: subtracting mean of each series and divide them
by their own std dev.)
y_c_fark = scale(yc_fark)

#Fitting processed data onto PCA model and deriving the first 10 PCs by
transforming the processed data.
#Create a PCA model with two principal components
pca = PCA(n_components=10)
pca.fit(y_c_fark)

```

```

#Get the components from transforming the original data.
scores = pca.fit_transform(y_c_fark)
#pc_y_c = pca.transform(y_c_fark)
# Reconstruct from the 10 dimensional scores
reconstruct = pca.inverse_transform(scores)
#The residual is the amount not explained by the first two components
residual=y_c_fark-reconstruct

#Extracting the factor loading vectors for the first three components
F_PC_1 = pca.components_[0]
plt.plot(F_PC_1)
plt.title('Factor Loadings of PC-1') # per Maturity Ladders')
plt.style.use('grayscale')
plt.xlabel('Maturity Ladders')
plt.ylabel('Standardised Interest Rate Level Changes')
#print first_pc_loadings
S_PC_1 = pca.components_[1]
plt.plot(S_PC_1)
plt.title('Factor Loadings of PC-2') # per Maturity Ladders')
plt.style.use('grayscale')
plt.xlabel('Maturity Ladders')
plt.ylabel('Standardised Interest Rate Level Changes')
#print second_pc_loadings
T_PC_1 = pca.components_[2]
plt.plot(T_PC_1)
plt.title('Factor Loadings of PC-3') # per Maturity Ladders')
plt.style.use('grayscale')
plt.xlabel('Maturity Ladders')
plt.ylabel('Standardised Interest Rate Level Changes')
#print third_pc_loadings

```

```

#The amount of variance that each PC explains
var_ycf= pca.explained_variance_ratio_
print('explained variance ratio (daily differences- first ten components): %s'
      % str(var_ycf))
#Cumulative Variance explained
cumvar_ycf=np.cumsum(np.round(pca.explained_variance_ratio_,
decimals=4)*100)
print('explained cumulative variance ratio (daily differences- first ten components):
%s'
      % str(cumvar_ycf))

#Plotting the explained cumulative variance ratio
plt.plot(cumvar_ycf)
plt.title('Cumulative Proportion of Variance Explained') # per Maturity Ladders')
plt.style.use('grayscale')
plt.xlabel('Principle Components')
plt.ylabel('Cumulative Variance Explanation Ratio')
#Creating a function that calculates the main results of PCA on the same table.
def pca_summary(pca, y_c_fark, out=True):
    names = ["PC"+str(i) for i in range(1, len(pca.explained_variance_ratio_)+1)]
    a = list(np.std(pca.transform(y_c_fark), axis=0))
    b = list(pca.explained_variance_ratio_)
    c = [np.sum(pca.explained_variance_ratio_[0:i]) for i in range(1,
len(pca.explained_variance_ratio_)+1)]
    columns = pd.MultiIndex.from_tuples([("sdev", "Standard deviation"),
("varprop", "Proportion of Variance"), ("cumprop", "Cumulative Proportion")])
    summary = pd.DataFrame(zip(a, b, c), index=names, columns=columns)
    if out:
        print("Importance of components:")
        print(summary)
    return summary
summary = pca_summary(pca, y_c_fark)

```

## TABLES AND GRAPHS

### Tables List

Table 1: Interest Rates per Maturities .....	3
Table 2: Descriptive Statistics of Interest Rate Levels at Different Maturities .....	10
Table 3: Correlation Matrix of Interest Rate Levels at Different Maturities .....	11
Table 4: Importance of Principle Components .....	12
Table 5: Factor Loadings of First Three PCs .....	13

### Graph List

Figure 1: Treasury Zero Coupon Yield Curve on 17.12.2015 .....	3
Figure 2 : Historical interest rate levels of each maturity ladder .....	9
Figure 3: Yield Curves .....	9
Figure 4: Box-Whisker Plot Graph of interest rate levels of different maturities....	10
Figure 5: Daily interest rate level changes per maturity ladders.....	11
Figure 6: Explained cumulative variance ratio .....	12
Figure 7: Factor Loading Graphs of First Three Components.....	13