Principal Component Analysis applied to European Yield Curve

Fabrizio Basso*

*Student ID: 10512476
*Emails: 10512476@mydbs.ie; fabrbasso@gmail.com

Applied Financial Analysis - CA1 Course Code: B9FT106

Abstract

This paper aims to explore the time series' proprieties of the features extracted by using the Principal Component Analysis (PCA) technique on the European AAA-rated Government Bond Yield curve. The PCA can greatly simplify the problem of modelling the yield curve by massively reducing its dimensionality to a small set of uncorrelated features. It finds several applications in finance and in the fixed income particularly from risk management to trade recommendation. After selecting a subset of Principal Components (PCs), this paper first analyzes their nature in comparison to the original rates and the implications in terms of information retained and lost. Then the time-series characteristics of each PC are studied and, when possible, Auto-Regressive Moving-Average (ARMA) models will be fitted on the data. One hundred observations of the original dataset are set aside as a test set to evaluate the predictive power of these models. Eventually, further analyses are performed on the PCs to evaluate the presence of heteroscedasticity and GARCH-ARCH models are fitted when possible. Tests are performed on the fitted coefficient to investigate the real nature of the conditional variance process.

Contents

1 Introduction

The modelling of the yield curve is an extremely daunting task. The major challenge is its high dimensionality. For instance, the yield curve used in this analysis, the European AAA-rated Government curve, is made up of 30 different maturities or 3'700 observations. The dataset consists of 15 years of daily readings. To further complicate the task, interest rate series are highly correlated among themselves, causing a severe problem of multi-collinearity. To overcome these two problems, it is now a well-established practice to employ the Principal Component Analysis (PCA).
PCA is a technique for dimensional reduction. Its core idea is to decrease the dimensionality of a dataset, consisting of a large number of interrelated variables, by extracting a subset of uncorrelated features, while retaining as much information as possible of the original dataset. The bibliography around the PCA and its applications to the Yield Curve analysis is vast, and the debate around how many PCs to retain and what they represent is mostly resolved in favour of three risk factors, frequently defined to as level, slope and curvature (See Hull [Hul12], Page 491). PCA analysis of the yield curve has ample applications in the day-by-day operations in the financial sector as it can be used to monitor the richness/cheapness of the curve (See Credit Suisse [PG12] (link)) or for risk management purpose (See Hull [Hul12], Page 493 for a very simple risk management application). Aim of this paper is to explore the consequences of applying the PCA to model the Yield Curve in terms of the meaning of the PCs and information retained/lost. Moreover, the time series proprieties of the selected PCs are studied in depth. In the first place, the statistical feature of the PCs' time series are investigated using their autocorrelogram (ACF) and partial-autocorrelogram (PACF) and, if the data exhibit evidence of it, an ARMA model will be fitted. One-hundred observations of the original dataset are set aside as a test set to evaluate the predictive power of these models. Further analyses will be then conducted on the PCs to assess if they are affected by heteroscedasticity and if possible GARCH-ARCH models will be fitted and test performed on their coefficients to explore the nature of the conditional variance processes. Given the widespread application of the PCA to finance, understanding the time series characteristics of the PCs extracted has several important implications. For instance, if a particular PC shows persistence in its returns, then positive shocks are likely to follow positive shocks, while if it exhibits heteroscedasticity, the situation of high volatility might be persistent and show up in clusters. These conditions, if present, should be taken in due consideration if an asset allocator is evaluating a trading opportunity or a risk manager is assessing the risk of a specific exposure relying on the PCA.

2 The Dataset

The dataset used in the analysis is made up of 3'700 daily observations (about 15 years) from the Euro Denominated, AAA-rated Government Bond Yield curve (the Yield Curve) - from 2005-05-03 to 2019-10-17 - for all the maturities ranging from 1 Year to 30 Years. The data are collected and made available from the ECB website. The ECB provides these data split into two huge csv file, collecting all the key interest rates from the Euro Area. The first file covers all the readings from 2004 to the end of 2018 (link), while the second captures all the data regarding the current year (link). Dealing with these dataset takes a toll on the computer performances due to the sheer size of the files (2.5 GB in total) and requires several lines of code to clean, select and sort the data. To not encumber the analysis notebook with unnecessary data from the original dataset, the uploading, the filtering and the ordering of the data into a well tided-up Pandas' DataFrame is performed in a different Notebook ("PCA - Yield Curve - Dataset Uploading"). At the end of the process, the dataset size shrinks from 4.2 million of rows by 40 columns to 3'700 rows by 30 columns. This DataFrame is then stored and uploaded in the analysis Notebook ("INSERT NAME"). Fig. 1 below reports a set of descriptive statistics of a selection of yields from the final dataset.

	count	mean	std	min	25%	50%	75%	max
Yield 1Y	3700.0	0.830	1.610	-0.908	-0.483	0.109	1.950	4.540
Yield 3Y	3700.0	1.127	1.623	-0.958	-0.378	0.511	2.299	4.738
Yield 5Y	3700.0	1.480	1.602	-0.919	-0.075	1.061	2.802	4.730
Yield 7Y	3700.0	1.812	1.577	-0.824	0.203	1.582	3.243	4.736
Yield 10Y	3700.0	2.200	1.545	-0.680	0.563	2.184	3.665	4.776
Yield 15Y	3700.0	2.570	1.506	-0.481	0.931	2.717	3.971	4.872
Yield 20Y	3700.0	2.727	1.469	-0.347	1.142	2.886	4.056	4.985
Yield 25Y	3700.0	2.780	1.434	-0.266	1.270	2.891	4.086	5.098
Yield 30Y	3700.0	2.787	1.404	-0.211	1.355	2.832	4.052	5.175

Figure 1: Descriptive Statistics for a subset of Yields from the dataset

3 Exploratory Data Analysis

The dataset used in the analysis is made up of the daily observation from the AAA-Rated European Government Yield Curve. The data cover 3'700 days, about 15 years, from 2005-05-03 to 2019-10-17. The dataset includes all the maturities ranging from 1 year up to 30. Over this period, the Yield Curve changed radically by shifting massively downward by flattening in its term structures, as shown by Fig. 2.

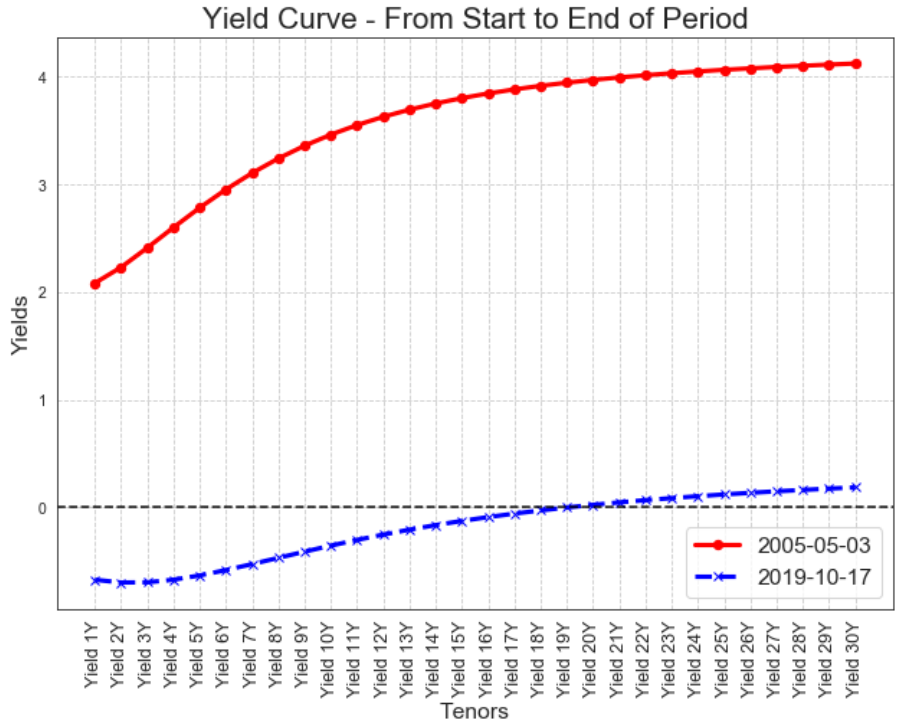

Figure 2: Yield Curve at the end and Beginning of Dataset

The time series of the rates across all the maturities, even from simple graphical exploration, seem to be non-stationary processes. This characteristic is confirmed by performing the Augmented Dickey-Fuller test on each of the tenor of the Yield Curve. The test results are reported in Fig. 4. As a result, the time series will be considered in their first difference to remove the unit-root.

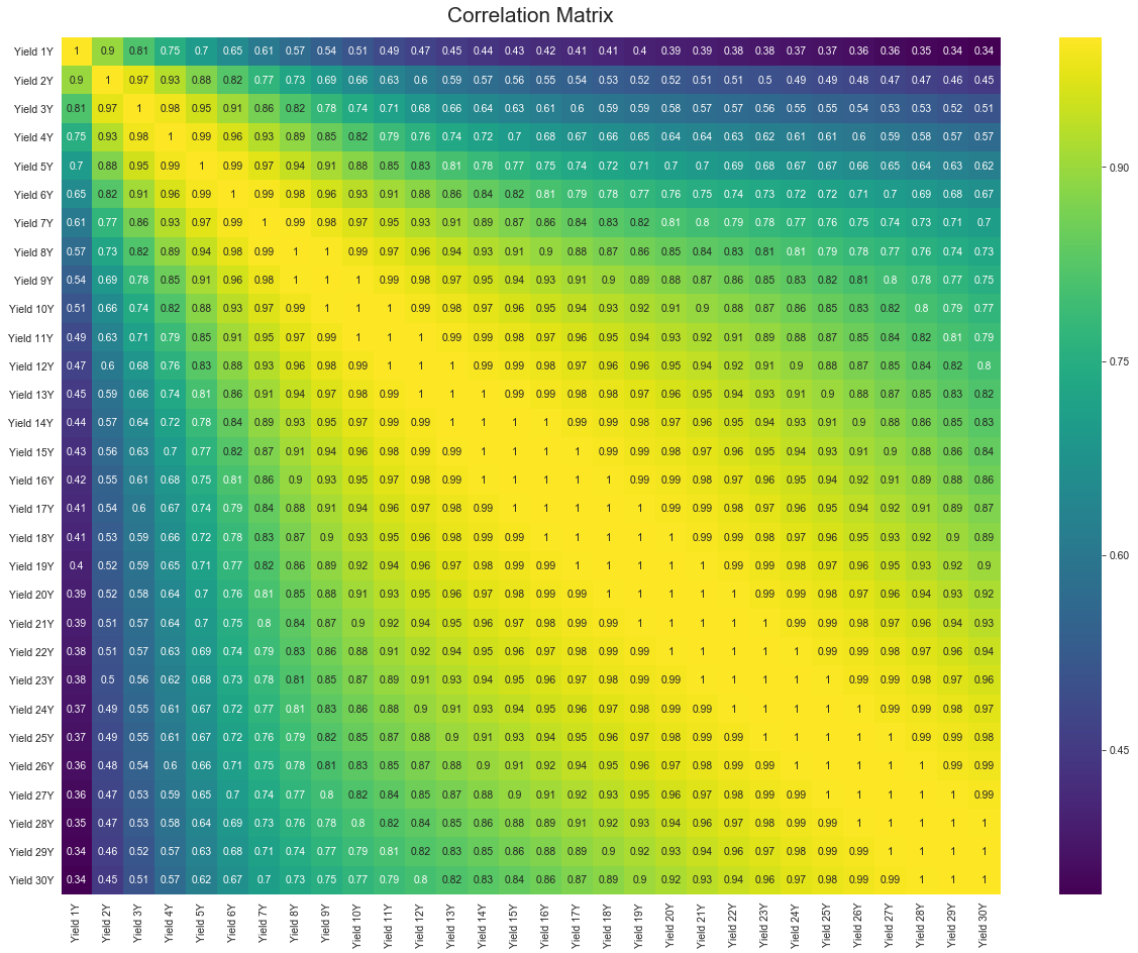

Figure 3: Yields Correlation matrix.

Since later on, in the analysis, it will be necessary to evaluate the performances of the fitted models, the dataset is split between training and test set:

- **Training**: 3'599 observations (one observation is lost when the first-difference operator is applied), from 2005-10-30 to 2019-05-31; and
- **Test**: 100 observations, from 2019-05-31 to 2019-10-17.

The left side of Fig. 6 provides a graphical representation on the evolution of the rates over the 3'599 days of the dataset. The right part shows the slope of each bucket of the curve, calculated as the difference between the tenor with the shortest maturity against the longest. The main take-away from these graphs is the increasing synchronisation of yields movements as the maturity of the tenors increases. In the last batch of tenors, from 20-years to 30-years maturities, for instance, all the yields trade in a very tight pack, as also shown by the slope that has the smaller range among the four pools of rates. All segments of the yield curve moved downward in the covered period.

The correlation matrix calculated on the daily variations confirms the insights gathered by the graphical exploration of the yields (See Fig. 3). It can be noticed that all the maturities above 19-years are highly correlated among themselves, with the coefficient ranging from 0.9 to 1. This characteristic of the rates is what makes their modelling particularly hard: it is a high dimensional system of highly correlated time series. Within this framework, the use of dimensional reduction techniques, such as PCA, is of great help.

The PCA model is sensitive to the scale of the features analysed. In the case of this study, all the variables are on the same scale and, as a result, the differences as well. All the means are

	Is Stationary?	P-Value		Is Stationary?	P-Value
Yield 1Y	False	0.797	Yield 16Y	False	0.961
Yield 2Y	False	0.859	Yield 17Y	False	0.961
Yield 3Y	False	0.871	Yield 18Y	False	0.961
Yield 4Y	False	0.894	Yield 19Y	False	0.967
Yield 5Y	False	0.911	Yield 20Y	False	0.967
Yield 6Y	False	0.924	Yield 21Y	False	0.965
Yield 7Y	False	0.935	Yield 22Y	False	0.964
Yield 8Y	False	0.944	Yield 23Y	False	0.963
Yield 9Y	False	0.953	Yield 24Y	False	0.961
Yield 10Y	False	0.957	Yield 25Y	False	0.959
Yield 11Y	False	0.959	Yield 26Y	False	0.956
Yield 12Y	False	0.962	Yield 27Y	False	0.953
Yield 13Y	False	0.959	Yield 28Y	False	0.950
Yield 14Y	False	0.960	Yield 29Y	False	0.946
Yield 15Y	False	0.961	Yield 30Y	False	0.942

Figure 4: Augmented Dickey-Fuller test on all YC tenors

already very close to zero but, as a precaution, they are subtracted from each series by applying the 'StandardScaler' provided by SciKit-Learn (deactivating the standard deviation scaling). As expected, the time series are practically identical after this transformation, as fig. 5 shows. However, another relevant feature emerges from this graph: the elevated frequency of outliers across all maturities.

Note: the high number of outliers across all the maturities raises the issue about how to deal with them and what is the proper way to scale the data. PCA can be affected by the presence of outliers, and this can constitute a **limitation** of the analysis conducted. However, in defence of the approach used in this paper, outliers seem to be equally numerous on both sides of the distributions and then less likely to bias the analysis in a specific direction. Testing different scaling procedure robust to outliers such as quartile-transformation to verify their impact on the PCA can be an exciting field for further researches.

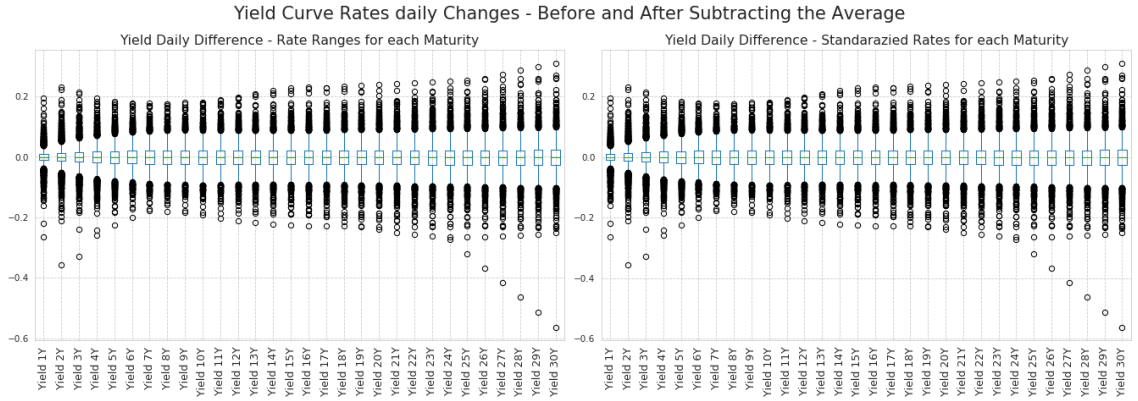

Figure 5: Rate Daily Difference before and after standardization

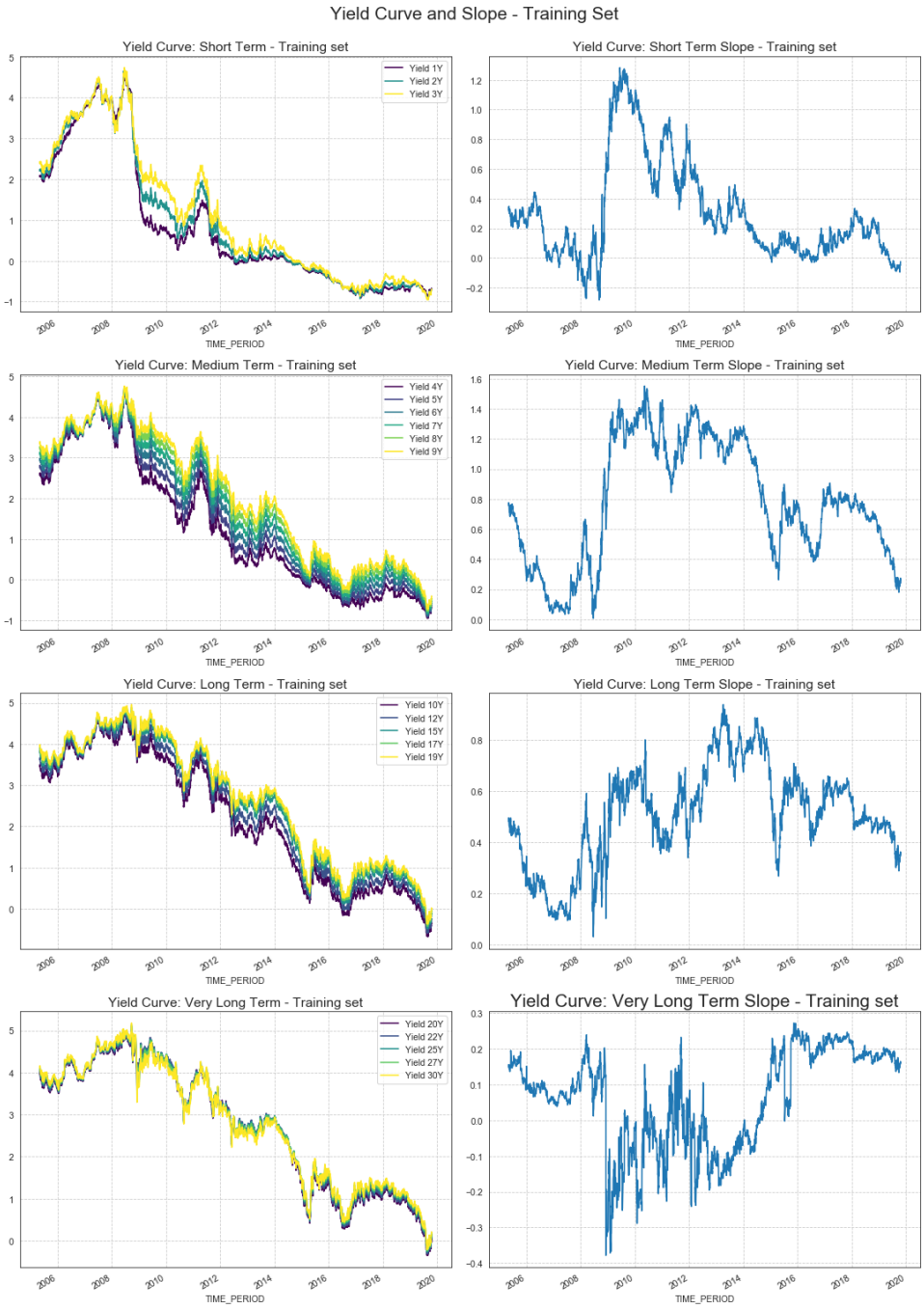

Figure 6: Yields by Maturity Buckets

4 Extracting the Principal Components

The PCA is applied to the dataset of the daily difference in the Yield Curve. To extract the PCs, this paper used the function provided by Scikit-Learn library. The PCA extracts new uncorrelated features from the original dataset, where each new PC, tries to capture as much information as possible from the original data. In other words, each new PC tries to explain as much volatility as possible of the data. Then, once the new features are extracted, it is worth exploring how much volatility of the original dataset each PC is able to explain. These results are represented in Fig. 7.

	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7	PC_8	PC_9
variance_explained	86.64%	9.12%	2.93%	0.96%	0.28%	0.06%	0.01%	0.00%	0.00%
variance_explained_cum	86.64%	95.76%	98.69%	99.65%	99.93%	99.99%	100.00%	100.00%	100.00%

Figure 7: PCs explained Variance

From the table above it can be observed that the first PC - PC-1 - alone is able to explain 86.64% of the total variance of the dataset. As we move to the second PC - PC-2 -, this captures 9.12% of the total variance of the dataset and the third a little less than 3%. In line with what is a well-established approach in the literature, only the first three principal components will be considered to reconstruct the Yield Curve movements (see for instance Crump and Gospodinov[Cru19] (link) for a recent application). The first three PCs considered together are then able to capture 98.69% of the original dataset's information. That's quite an achievement. If this level of information loss (about 1.3%) can be accepted, then the total model complexity is reduced from 30 features to 3. The results achieved in terms of explained variance are in line with those recorded by Crump and Gospodinov[Cru19] on the US yield curve: by using three PCs, they can explain 96.8% of the variance on the forward curve and, using different assumptions, between 93.4% and 99.8% of the US yield curve. Before starting to analyse the time series proprieties of the PCs, the general relationship between the PCs and the original Yields is explored (4.1) along with a more in-depth analysis of the information loss due to the use of the PCA (4.2).

4.1 Principal Components and Yields

To study the relationship between the PCs and the Yields requires to focus on the eigenvectors used to transform the original data. Since only the first 3 PCs are retained, only the three eigenvectors associated with these features are considered. The three eigenvectors are plotted in fig. 8.

The eigenvectors confirm the traditional interpretation of the first three PCs provided by the literature on this theme.

- **PC-1 - Levels**: The first principal component seems to be associated with parallel shifts in the Yield Curve. A change in the value of the first PC propagates positively across all the tenors. Moreover, the graph tells that the longer the maturity the greater the effect;
- **PC-2 - Slope**: The second PC relates to the slope of the Yield Curve. The eigenvector associated with this principal component shows that when the PC-2 experience a positive change, its variation has a positive impact on the maturities up to approximately 14 years and then it turns negative;
- **PC-3 - Curvature**: The third PC is connected with the curvature of the Yield Curve. The eigenvector associated with this principal component shows that when the PC-3 experience a positive change, its variation has a positive effect on the "belly" of the Yield Curve (maturities between 5 and 21 years) while it is negative on the two "wings" (maturities up to 5 years and above 21);

Further confirmation of this interpretation of the first three PCs comes from studying their correlation with the yields change. The correlations are shown in Fig. 9, to facilitate the reading of the matrix, a subset of tenors is selected. The red box at the right end of the matrix highlights the correlations of the PCs. It can be observed that the PC-1 has a positive correlation with all the tenors, the PC-2 has positive correlations with all the maturities up to 15 years and then turns negative. The PC-3 has a positive correlation with the intermediate maturities and negative on

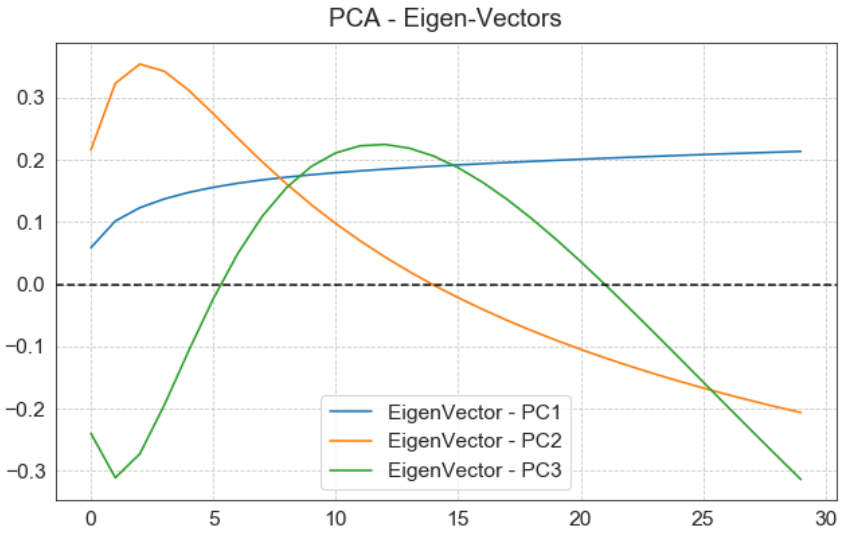

Figure 8: Eigenvectors associated with the first three PCs.

the extremities. The correlation matrix confirms another critical feature of the PCs. The purple box at the bottom-right of the matrix shows how the correlation of the PCs among themselves is always zero. This feature is essential, as it allows us to study and model each PC independently from the others.

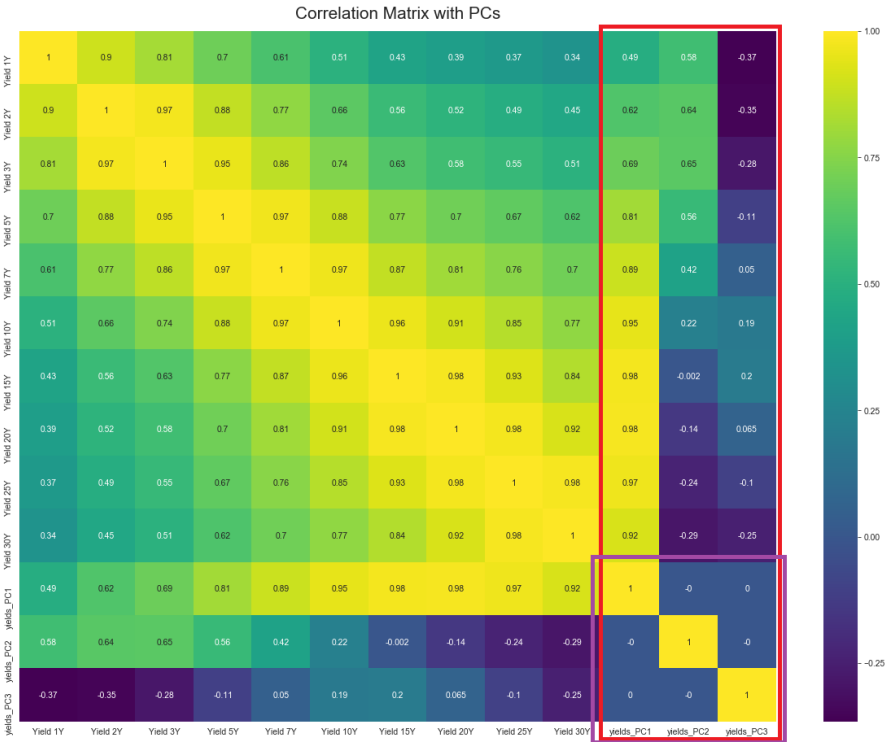

Figure 9: Yield Changes and PCs Correlation Matrix.

4.2 Principal Components and Information Loss

Paragraph 4 highlighted how the first three PCs were able to explain 98.69% of the variance of the dataset. As a result, "only" 1.3% of the information is lost. This paragraph explores in greater detail the nature of this loss and its implications. An ideal situation would be where the loss of information distributes uniformly across all the tenors of the Yield Curve. On the contrary, if the loss concentrates in a specific bucket of maturities or, even worse, in a single tenor, that would be a serious shortcoming to the applicability of the PCA to model the Yield Curve. To carry out this study, the yield changes are reconstructed starting from the three PCs considered and their eigenvectors by applying the following relationship:

$$I_t^{rec} = PC_{t,[1,\dots,i]} * E^{-1}_{[1,\dots,i]} \tag{1}$$

The equation 1 states that the reconstructed Yield changes (I_t^{rec}) in each period t using the first i PCs ($PC_{t,[1,\dots,i]}$) are equal to the product of the value of the PCs in t by the first i eigenvectors ($E^{-1}_{[1,\dots,i]}$). In this specific case i=3. Once the reconstructed Yields changes are calculated, the errors against the actual changes are computed as follow [Eq. 2]:

$$error_t = I_t^{rec} - Y_t \tag{2}$$

These errors represent the **"loss of information"** resulting by using a subset of PCs for each maturity. If all the available PCs (30) were used the errors would be 0 for all the tenors in any t. Using the errors as calculated in Eq. 2, the following metrics are used to measure the accuracy of the reconstructed Yield changes:

- **MAE**: Mean Absolute Error;
- **MSE**: Mean Squared Error; and
- **RMSE**: Root Mean Squared Errors.

Fig. 10 reports the scores achieved on these metrics:

	MSE	RMSE	MAE		MSE	RMSE	MAE
Yield 1Y	0.0002	0.0133	0.0081	Yield 16Y	0.0	0.0034	0.0019
Yield 2Y	0.0001	0.0088	0.0049	Yield 17Y	0.0	0.0038	0.0021
Yield 3Y	0.0000	0.0049	0.0026	Yield 18Y	0.0	0.0040	0.0021
Yield 4Y	0.0000	0.0039	0.0022	Yield 19Y	0.0	0.0040	0.0021
Yield 5Y	0.0000	0.0048	0.0028	Yield 20Y	0.0	0.0038	0.0019
Yield 6Y	0.0000	0.0056	0.0033	Yield 21Y	0.0	0.0034	0.0017
Yield 7Y	0.0000	0.0058	0.0033	Yield 22Y	0.0	0.0029	0.0014
Yield 8Y	0.0000	0.0054	0.0030	Yield 23Y	0.0	0.0022	0.0010
Yield 9Y	0.0000	0.0047	0.0024	Yield 24Y	0.0	0.0014	0.0006
Yield 10Y	0.0000	0.0036	0.0018	Yield 25Y	0.0	0.0007	0.0003
Yield 11Y	0.0000	0.0025	0.0012	Yield 26Y	0.0	0.0009	0.0005
Yield 12Y	0.0000	0.0016	0.0008	Yield 27Y	0.0	0.0019	0.0010
Yield 13Y	0.0000	0.0015	0.0009	Yield 28Y	0.0	0.0030	0.0016
Yield 14Y	0.0000	0.0021	0.0012	Yield 29Y	0.0	0.0043	0.0022
Yield 15Y	0.0000	0.0028	0.0016	Yield 30Y	0.0	0.0056	0.0028

Figure 10: Errors associated to reconstructed Yield Changes.

Fig. 10 shows that the distribution of the errors is far away from being uniform. The loss of information seems to affect the short end part of the Yield Curve in particular. To better assess

the concentration of the information loss, the errors are recalculated in percentage terms of the cumulative error observed for each metric across all tenors. The results are reported in Fig. 11

	MSE	RMSE	MAE		MSE	RMSE	MAE
Yield 1Y	28.07	11.44	12.74	Yield 16Y	1.79	2.89	2.98
Yield 2Y	12.39	7.60	7.68	Yield 17Y	2.25	3.24	3.25
Yield 3Y	3.75	4.18	4.16	Yield 18Y	2.50	3.41	3.34
Yield 4Y	2.43	3.37	3.45	Yield 19Y	2.50	3.41	3.27
Yield 5Y	3.61	4.10	4.46	Yield 20Y	2.27	3.25	3.04
Yield 6Y	4.91	4.78	5.17	Yield 21Y	1.84	2.93	2.68
Yield 7Y	5.31	4.98	5.19	Yield 22Y	1.31	2.47	2.20
Yield 8Y	4.70	4.68	4.69	Yield 23Y	0.77	1.89	1.62
Yield 9Y	3.44	4.01	3.84	Yield 24Y	0.32	1.22	0.98
Yield 10Y	2.07	3.11	2.84	Yield 25Y	0.07	0.57	0.46
Yield 11Y	1.00	2.16	1.90	Yield 26Y	0.12	0.74	0.79
Yield 12Y	0.42	1.40	1.32	Yield 27Y	0.56	1.61	1.63
Yield 13Y	0.35	1.28	1.39	Yield 28Y	1.47	2.62	2.54
Yield 14Y	0.68	1.78	1.95	Yield 29Y	2.92	3.69	3.48
Yield 15Y	1.22	2.38	2.53	Yield 30Y	4.97	4.81	4.46

Figure 11: Errors as percentage of the total error according each metric.

From Fig. 11 it is possible to see that, for instance, the 1Y tenor by itself accounts for the highest share of the errors according to all the metrics, followed by the 2Y and 7Y maturities. By converse, the loss of information is almost non-existent in other maturities such as the 25Y. Fig. 12 displays the loss of information for buckets of maturities. Again, the results confirm that the loss is concentrated in the 1Y-5Y bucket (from 30.6% to 50.2%) followed by the 6Y-10Y (from 20.4% to 21.7%).

	MSE	RMSE	MAE
1-5 Years	50.2489	30.6859	32.4928
6-10 Years	20.4354	21.554	21.7141
11-15 Years	3.66798	9.00146	9.08705
16-20 Years	11.3023	16.2015	15.8761
21-25 Years	4.315	9.08722	7.93029
26-30 Years	10.0305	13.4699	12.8998

Figure 12: Errors as percentage of the total error per buckets of maturities.

To further study the loss of information in the 1Y tenor, the actual and the reconstructed rate changes are compared and then regressed on each other. Fig. 13 below shows that the two series share the same mean. However, the reconstructed series shows a lower standard deviation in comparison to the original data.

By analysing the regression results (Fig. 14) of the actual yield changes against the reconstructed ones, it follows that the constant and the coefficient are consistent with the absence of bias in the reconstructed series. However, the R^2 of the regression is only 0.708, meaning that the regressor is only partially able to explain the original data. As a comparison, the same regression using the actual and reconstructed 25Y rates achieves a R^2 of 1.

	Yield 1Y	1Y Yield Reconstructed
count	3599.000000	3599.000000
mean	-0.000752	-0.000752
std	0.024586	0.020691
min	-0.263676	-0.172638
25%	-0.009647	-0.008971
50%	-0.000269	-0.000511
75%	0.009483	0.007969
max	0.194397	0.142050

Figure 13: Stats comparison between 1Y actual rates changed and reconstructed.

OLS Regression Results

Dep. Variable:	Yield 1Y	R-squared:	0.708
Model:	OLS	Adj. R-squared:	0.708
Method:	Least Squares	F-statistic:	8731.
Date:	Sun, 17 Nov 2019	Prob (F-statistic):	0.00
Time:	20:39:17	Log-Likelihood:	10447.
No. Observations:	3599	AIC:	-2.089e+04
Df Residuals:	3597	BIC:	-2.088e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.246e-18	0.000	5.62e-15	1.000	-0.000	0.000
1Y Yield Reconstructed	1.0000	0.011	93.441	0.000	0.979	1.021

Omnibus:	2056.251	Durbin-Watson:	1.935
Prob(Omnibus):	0.000	Jarque-Bera (JB):	105381.275
Skew:	-1.999	Prob(JB):	0.00
Kurtosis:	29.206	Cond. No.	48.3

Figure 14: OLS regression results of daily changes of 1Y rate actual and reconstructed.

Note: The lack of uniformity in the distribution of the information loss among the tenors and its concentration in specific maturities or specific segments of the curve can constitute a severe **limitation** on the application of the PCA to model the Yield Curve to obtain trading indications or to monitor risks. In the already mentioned paper from Credit Suisse [PG12] aiming at creating trading recommendations, the authors claim that by using the first three PCs they manage to retain 95% of the information from the original dataset. However, no study is performed with regards to where the information is lost. If, as in this paper, the information loss is concentrated in specific tenors of the curve, the trading recommendations relative to those specific maturities should be disregarded or considered with a lower level of confidence.

5 Analysing PCs Time Series

In this section, the focus moves on to study the statistical proprieties of the selected PCs as time series. To carry out the analysis, the following steps will be implemented on each PC:

1. **Step 1**: Autocorrelation (ACF) and Partial-Autocorrelation (PACF) of each PC are analysed to identify any ARMA process present in the series;
2. **Step 2**: Relying on the results from step 1, an ARMA model is fitted on each PC; and
3. **Step 3**: The residuals of the regressions are studied to identify the presence of any heteroskedasticity effect. The White test for heteroskedasticity and an analysis of the ACF and PACF of the squared residual are used to implement this step;

4. **Step 4**: if heteroskedasticity is observed, a Garch-Arch Model is fitted on the residual.

5.1 Step I - Principal Components: Exploratory Data Analysis

In this section, the time series of the three PCs are studied through their Autocorrelation (ACF) and Partial-Autocorrelation (PACF) functions. Moreover, the historical distributions of each PC are compared with those of Gaussian function using a Q-Q and a P-P plot to obtain some insights on their distributions.

1. **PC - 1** (Fig. 15): the ACF and PACF graphs seem to suggest that the first lag $(t-1)$ has some explanatory power with regards to the reading in t. The two graphs can be consistent with an AR(1), MA(1) or even an ARMA(1,1) process. At this stage, it is hard to be more precise as the signal gets quickly lost in the noise. In any case, the coefficients of the process are likely to be very small. The Q-Q and Probability Plots show the data points fall along the Normal line in the middle of the graph, but curve off in the extremities. This behaviour usually means your data have more extreme values than would be expected if they truly came from a Normal distribution. From a visual inspection of the graph on the top, these extreme values seem to be clustered in specific periods, possibly suggesting that the time series is affected by heteroskedasticity;

2. **PC - 2** (Fig. 16): the ACF and PACF graphs seem to suggest that the first lag $(t-1)$ has some explanatory power with regards to the reading in t. The two graphs can be consistent with an AR(1), MA(1) or even an ARMA(1,1) process. At this stage, it is hard to be more precise as the signal gets quickly lost in the noise. In any case, the coefficients of the process are likely to be very small. The Q-Q and Probability Plots show the data points fall along the Normal line in the middle of the graph, but curve off in the extremities. This feature seems more marked here in comparison to the one observed for the PC-1. This behaviour usually means your data have more extreme values than would be expected if they truly came from a Normal distribution. From a visual inspection of the graph on the top, these extreme values seem to be clustered in specific periods, possibly suggesting that the time series is affected by heteroskedasticity;

3. **PC - 3** (Fig. 17): the ACF and PACF graphs does not suggest any particular ARMA(p,q) process to describe the data. The Q-Q and Probability Plots are similar to those observed for the two other PCs. They show that the data points fall along the Normal line in the middle of the graph but curve off in the extremities. This feature seems to be more marked here in comparison to the one observed for the PC-1 and PC-2. This behaviour usually means data have more extreme values than would be expected if they truly came from a Normal distribution. From a visual inspection of the graph on the top, these extreme values seem to be clustered in specific periods of time, possibly suggesting that the time series is affected by heteroskedasticity.

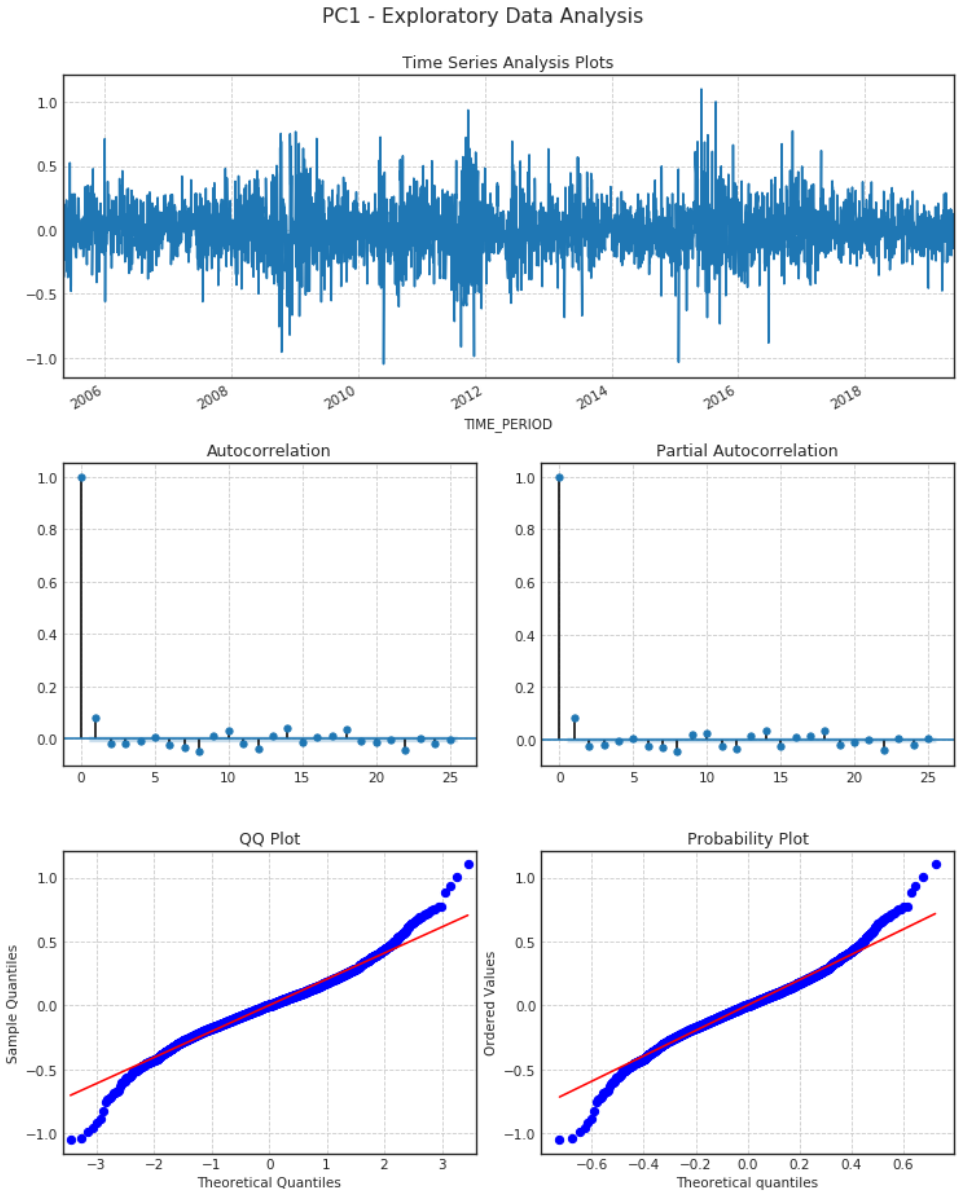

Figure 15: PC-1 Exploratory Data Analysis.

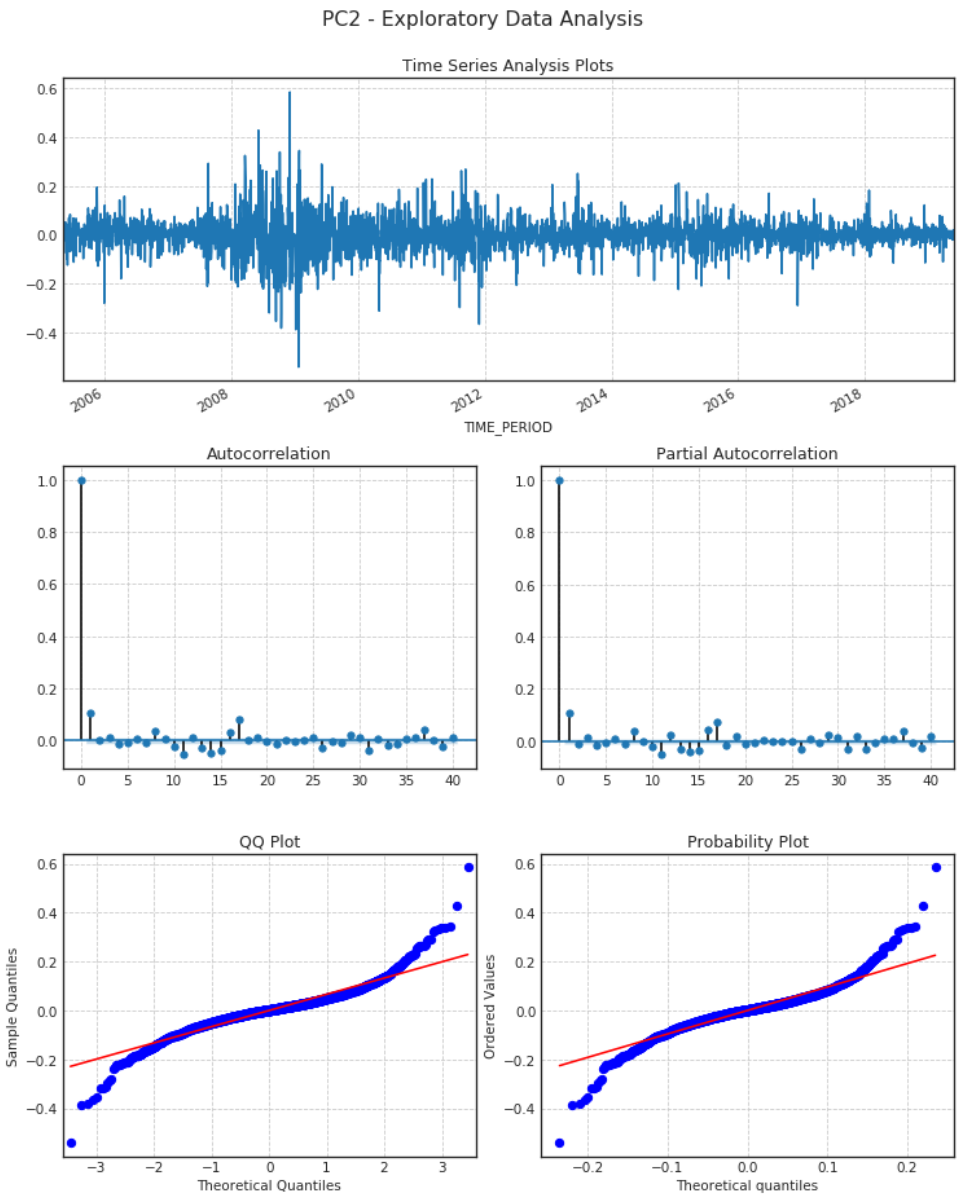

Figure 16: PC-2 Exploratory Data Analysis.

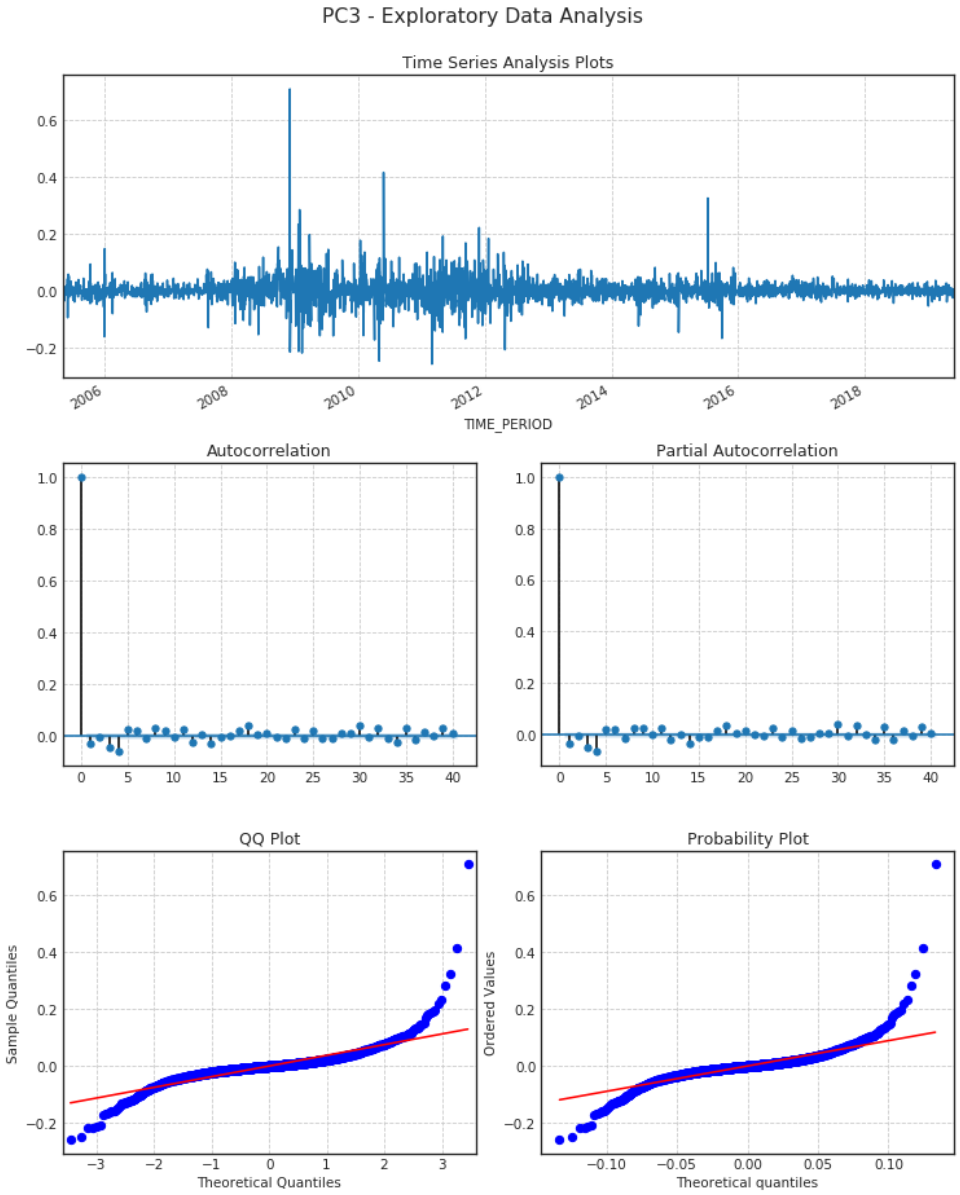

Figure 17: PC-3 Exploratory Data Analysis.

In general, all three PCs does not show high correlations among their historical values. As a result, past data have low predictive power for future observations. Some positive autocorrelation has been spotted in PC-1 and PC-2, and this is going to be investigated in the next subsection.

5.2 Step II - Principal Components: ARMA Models Analysis

To estimate the proper order -p and q- of the ARMA(p, q) models to fit on the extracted PCs, all the lags up to 6 periods have been considered for the AR(q) and MA(q) components. The best model to fit the data is selected according to the **Bayesian information criterion** (BIC). This procedure is applied to the three chosen PCs to verify the conclusions that emerged in the preceding section.

For each PC, the resulting best ARMA(p, q) process is reported along with a preliminary analysis of the residual.

1. **PC - 1** (Fig. 15): The implementation of the procedure described above on the PC-1 highlights the significance of the MA(1) component in describing future observations. However, the coefficient associated with this feature is quite low, 0.0853, although significantly different from zero, as shown by the t-test and associated p-value. This model performs very poorly in explaining the dependable variable as confirmed by the R^2 (Fig. 18): 0.007. The analysis of the regression's residuals (Fig. 15) shows that the significance of the first lag in the autocorrelation function has disappeared. The QQ plot confirms the heavy tail distribution already stressed in section 5.1;

2. **PC - 2** (Fig. 16): The analysis on the PC-2 highlights again the significance of the MA(1) component in describing future observations. The coefficient associated with this feature is still quite low, 0.1092, albeit higher than the one observed for the PC-1, and significantly different from zero, as shown by the t-test and associated p-value. The R^2 (Fig. 18) of the model is still very low: 0.012. The analysis of the regression residuals (Fig. 15) shows that the significance of the first lag in the autocorrelation function has disappeared. The QQ plot confirms the heavy tail distribution already stressed in section 5.1;

3. **PC - 3** (Fig. 17): As for the PC-3, the lowest score for the BIC metric is achieved for a model made up of the constant only, confirming that there is no significant ARMA(p, q) process in this time series.

	PC_1	PC_2	PC_3
R2	0.007	0.012	-0.000
MSE	0.041	0.004	0.001
MAE	0.152	0.045	0.022

Figure 18: ARMA models Statistics.

ARMA Model Results

Dep. Variable:	y	No. Observations:	3598
Model:	ARMA(0, 1)	Log Likelihood	622.858
Method:	css-mle	S.D. of innovations	0.204
Date:	Tue, 19 Nov 2019	AIC	-1239.717
Time:	16:43:00	BIC	-1221.152
Sample:	0	HQIC	-1233.101

	coef	std err	z	P>|z|	[0.025	0.975]
const	1.889e-05	0.004	0.005	0.996	-0.007	0.007
ma.L1.y	0.0853	0.017	5.077	0.000	0.052	0.118

Roots

	Real	Imaginary	Modulus	Frequency
MA.1	-11.7171	+0.0000j	11.7171	0.5000

ARIMA Order: I: 0, ar: 0, ma: 1

Figure 19: PC-1 ARMA Process.

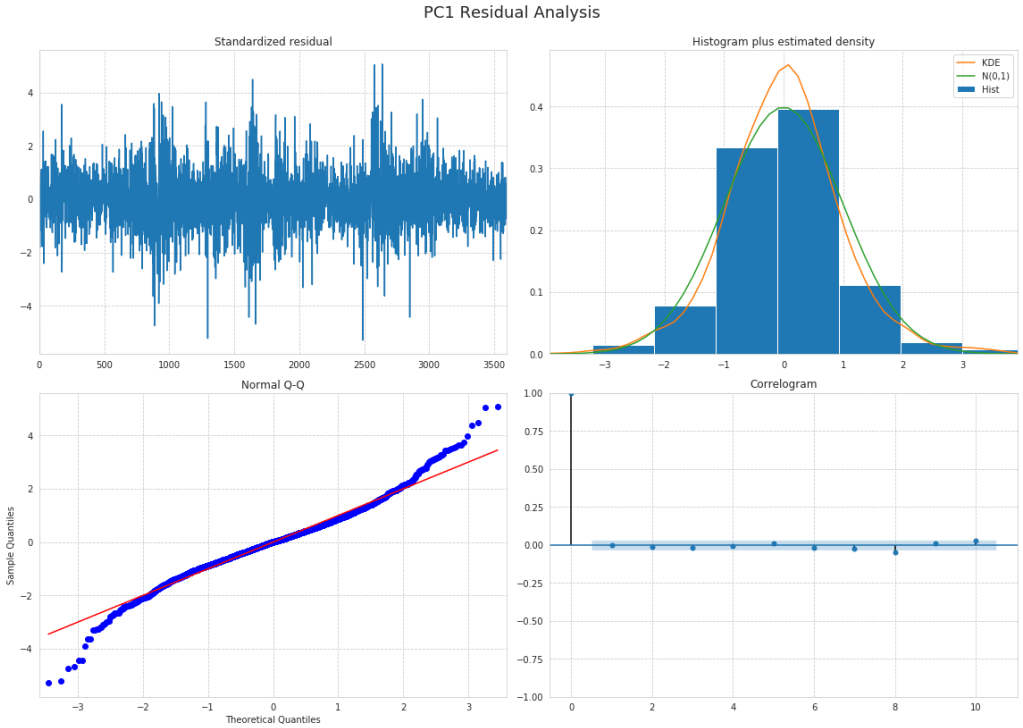

Figure 20: PC-1 ARMA Residuals.

ARMA Model Results

Dep. Variable:	y	No. Observations:	3599
Model:	ARMA(0, 1)	Log Likelihood	4682.802
Method:	css-mle	S.D. of innovations	0.066
Date:	Tue, 19 Nov 2019	AIC	-9359.605
Time:	16:42:42	BIC	-9341.040
Sample:	0	HQIC	-9352.989

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.647e-06	0.001	-0.001	0.999	-0.002	0.002
ma.L1.y	0.1092	0.017	6.571	0.000	0.077	0.142

Roots

	Real	Imaginary	Modulus	Frequency
MA.1	-9.1595	+0.0000j	9.1595	0.5000

ARIMA Order: I: 0, ar: 0, ma: 1

Figure 21: PC-2 ARMA Process.

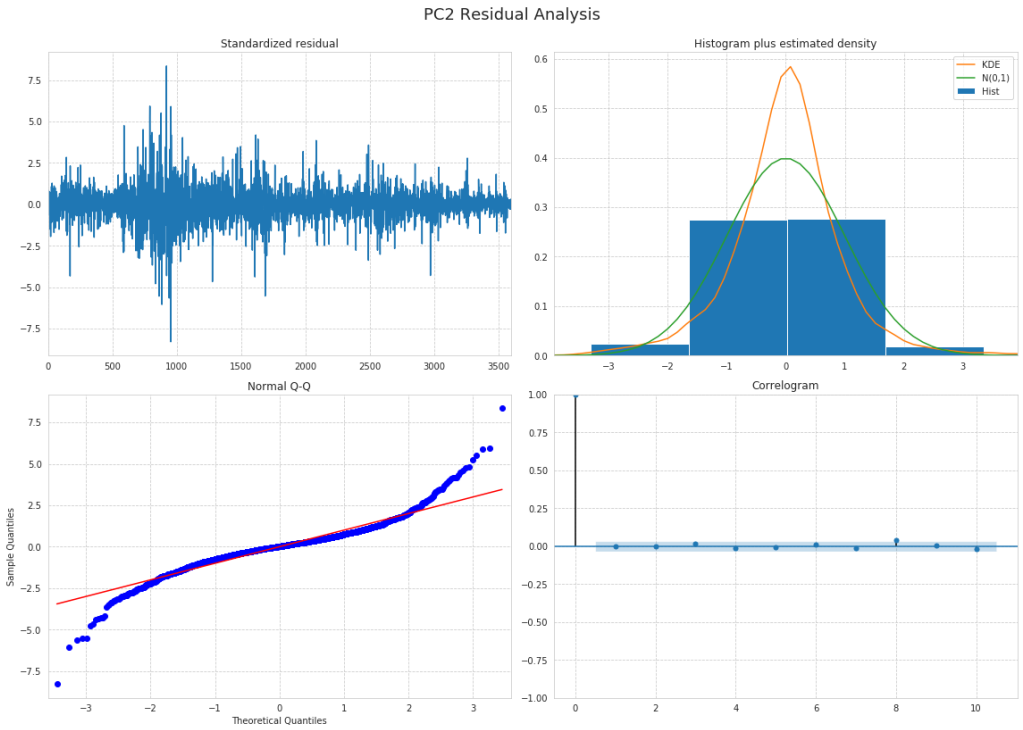

Figure 22: PC-2 ARMA Residuals.

ARMA Model Results

Dep. Variable:	y	No. Observations:	3599
Model:	ARMA(0, 0)	Log Likelihood	6706.933
Method:	css	S.D. of innovations	0.038
Date:	Tue, 19 Nov 2019	AIC	-13409.866
Time:	16:42:04	BIC	-13397.489
Sample:	0	HQIC	-13405.455

	coef	std err	z	P>|z|	[0.025	0.975]
const	-7.246e-10	0.001	-1.16e-06	1.000	-0.001	0.001

ARIMA Order: I: 0, ar: 0, ma: 0

Figure 23: PC-3 ARMA Process.

Despite having a significant coefficient, the two ARMA models fitted on PC-1 and PC-2 deliver a very dismal performance in explaining the dependent variable as shown by the R^2 metrics. Further confirmations of their poor performances come from their scores in forecasting the 100 observations set aside as a test set at the beginning of the analysis, Fig. 25 and Fig. 26. It covers the period from the end of May-19 to mid-October-19. The results, reported in Fig. 24, show values of the RMSE in line with the Standard Deviation of the time series while the R^2 is negative, meaning that the simple average of the series provides a better forecast.

	PC_1	PC_2
R2_f	-0.025115	-0.047832
RMSE_f	0.204961	0.053815
MAE_f	0.157500	0.040005

Figure 24: PCs out-of-sample Scores

Note: Even though a MA(1) component is significant for the PC-1 and PC-2, its explanatory and predictive power is minimal. Both the in-sample and out-of-sample analysis confirm this conclusion. As a result, future movements of the PCs cannot be explained by their own past. However, this does not impinge the general benefits of using the PCA, since it does not rule out that the three PCs could be modelled using other exogenous features, rather than modelling the original 30 features.

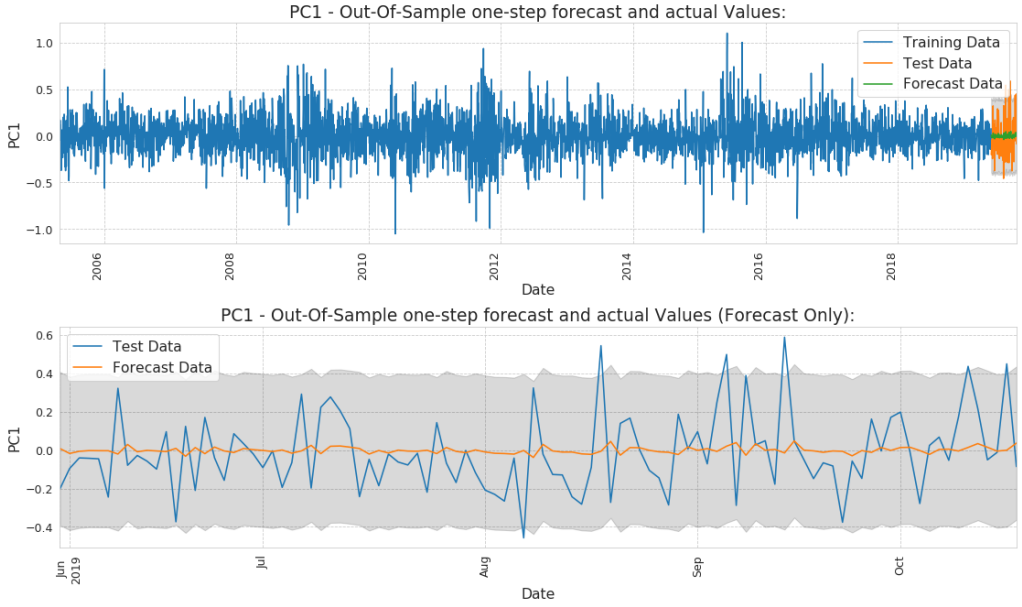

Figure 25: PC-1 In-sample and Forecast

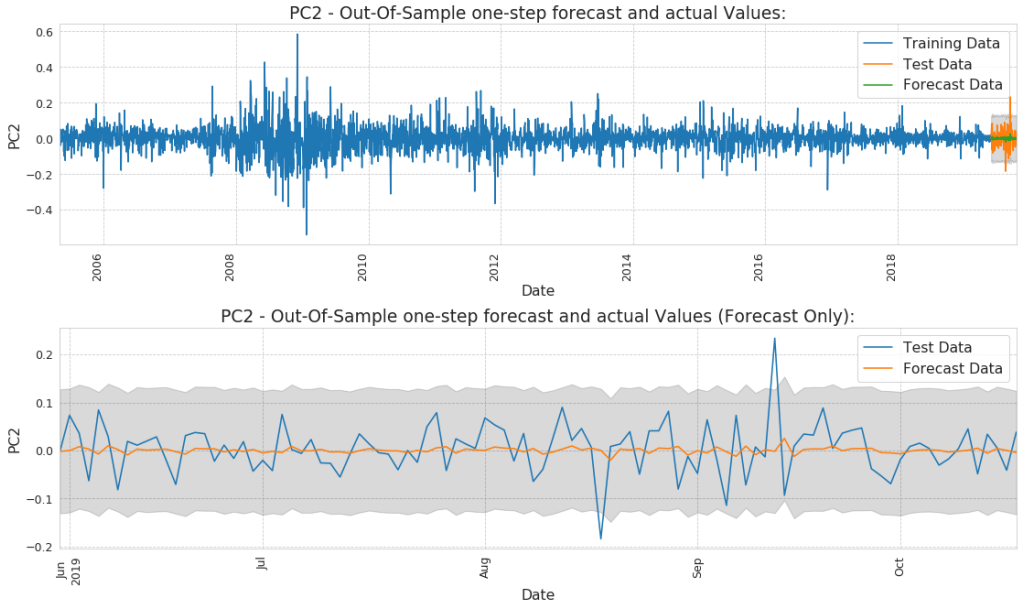

Figure 26: PC-2 In-sample and Forecast

5.3 Step III - Principal Components: Testing for Heteroskedasticity

In this section, the presence of heteroskedasticity is explored in greater details. The analysis is carried out in two steps. First, the White-Test for heteroskedasticity is performed on the residuals of the regressions fitted in the preceding section. Then, the ACF and the PACF of the squared residuals are studied to verify if there are any hints of a Garch-Arch process of a specific order.

- White-Test for Heteroscedasticity: The test is performed on the residual of the ARMA processes estimated in Section 5.2. It is based on an auxiliary regression using the squared residual as the independent variable. The test's H_0 is that the time series is homoscedastic. The test can be performed either on the residual of the auxiliary regression or its coefficients. As a result, two statistics are produced: the LM-Statistic (test on residual) and the F-Statistic (test on coefficients). The results are reported in the table below (Fig. 27):

	Stats - PC1	Stats - PC2	Stats - PC3
LM Statistic	86.96	438.13	37.56
LM-Test p-value	0.00	0.00	0.00
F-Statistic	44.52	249.23	18.96
F-Test p-value	0.00	0.00	0.00

Figure 27: White-Test for Heteroscedasticity Results

The P-value associated with all the statistics is 0, and therefore the H_0 is rejected for each PCs. As a result, the White-Test confirms the presence of heteroscedastic in the PCs series.

- Squared Residual Analysis: The squared residuals are analysed through their ACF and PACF to identify the order of the Garch-Arch process.

1. **PC-1 Residuals** (Fig. 28): The ACF shows a strong persistence of the significance of the past observation. A similar effect can be spotted for the PACF, although less marked. The two graphs together suggest the presence of a GARCH(p, q) process where both the component q related to σ^2 and p related to ϵ^2 are present;
2. **PC-2 Residuals** (Fig. 29): Again the ACF shows a significance persistence of the significance of the past observation. The PACF, on the other hand, shows a faster decay in its values. The two graphs together suggest the presence of a GARCH(p, q) where the component q related to σ^2 is somewhat dominant in comparison to p connected to ϵ^2;
3. **PC-3 Residuals** (Fig. 30): ACF and PACF exhibit some significance for the short term lags, however, the signal quickly decays into noise. Overall, they suggest the presence of a GARCH(p, q) process, albeit less complex than those identified for the PC-1 and PC-2.

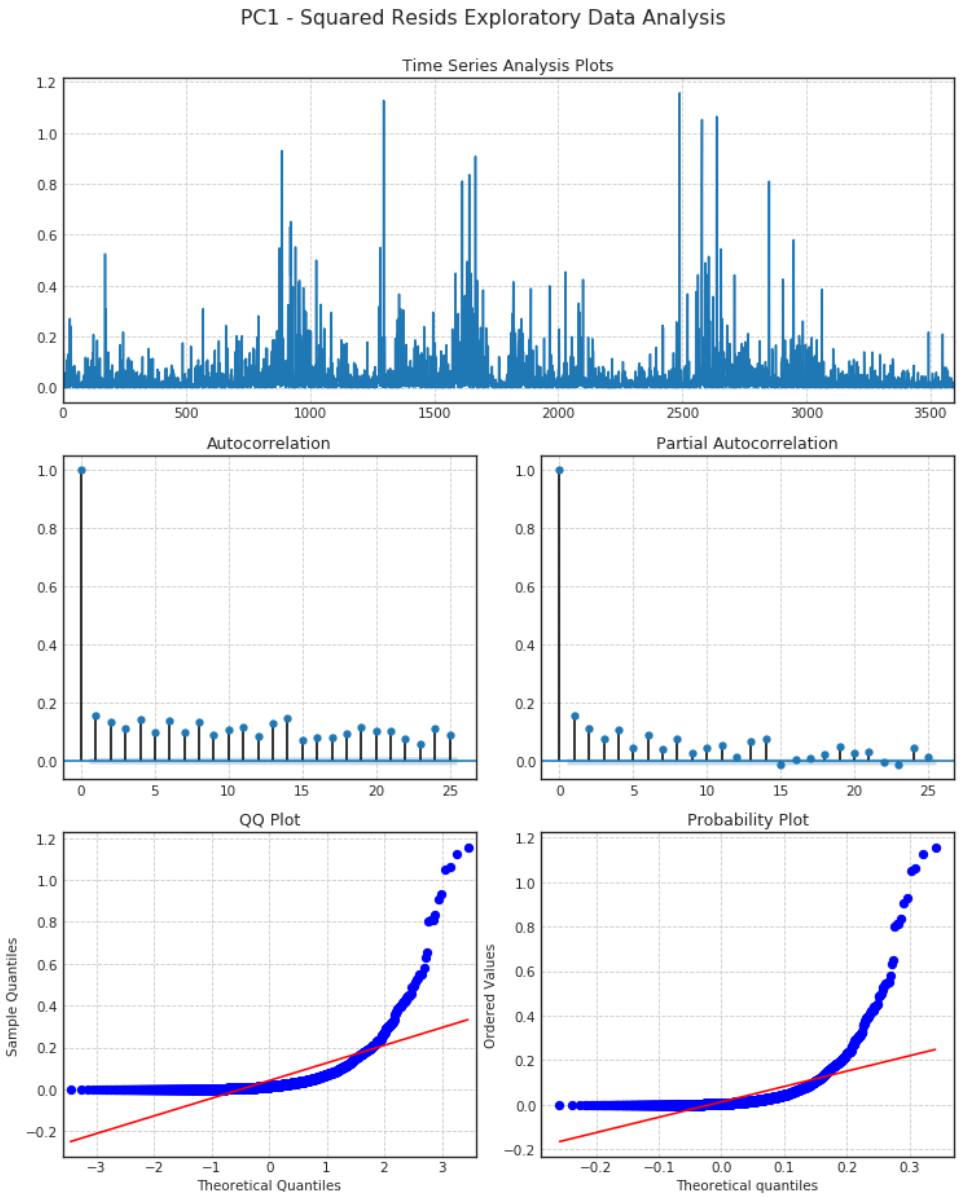

Figure 28: PC-1 Squared Residual Exploratory Data Analysis.

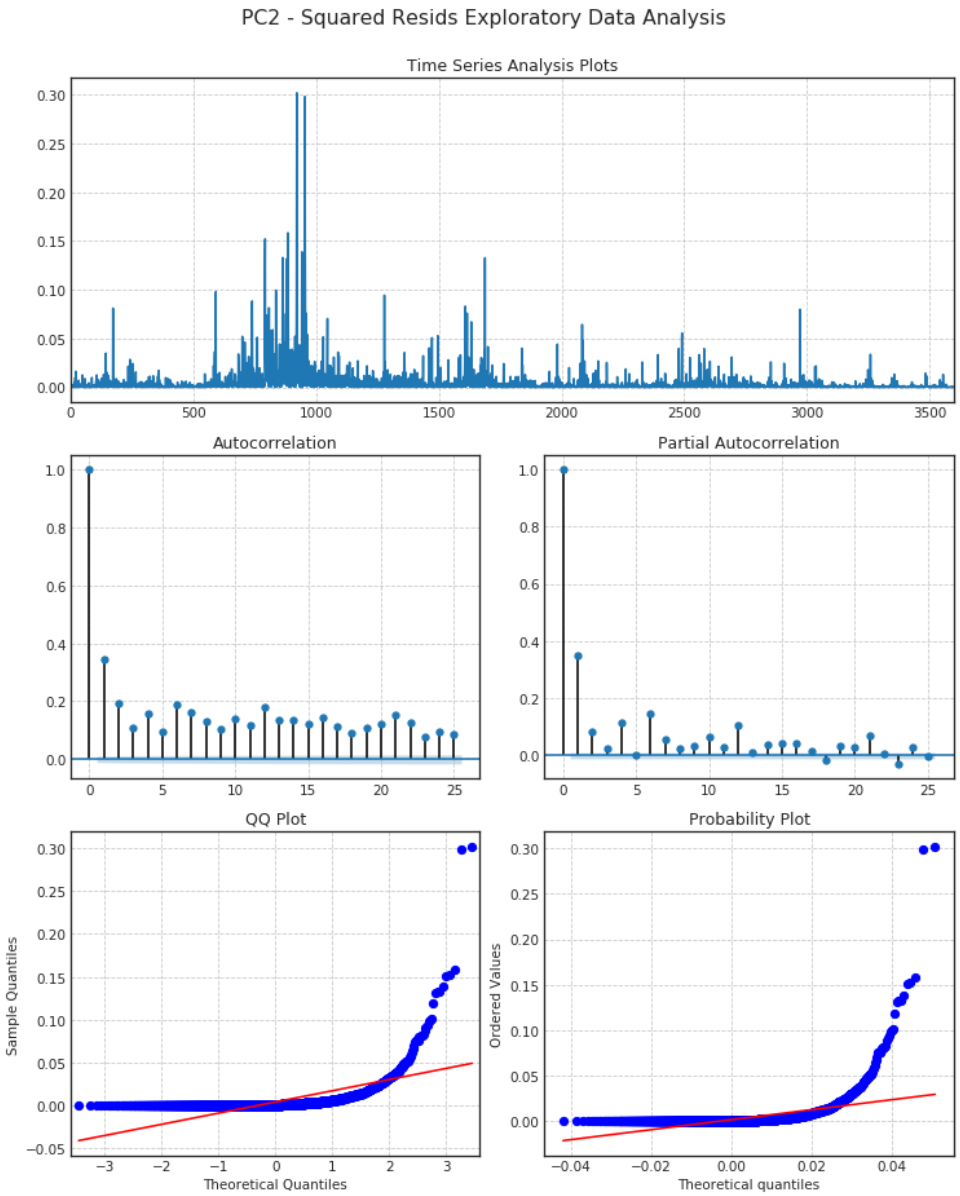

Figure 29: PC-2 Squared Residual Exploratory Data Analysis.

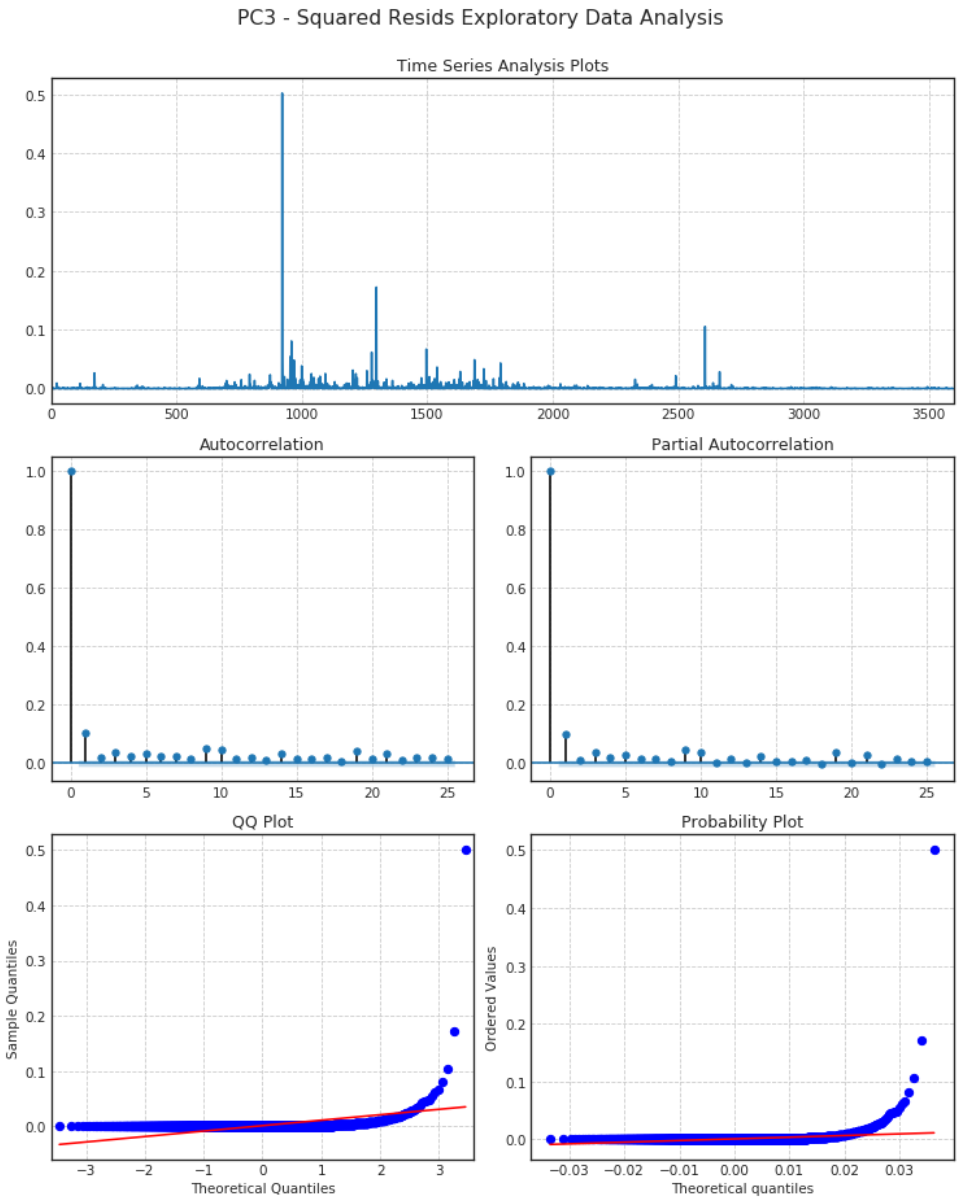

Figure 30: PC-3 Squared Residual Exploratory Data Analysis.

5.4 Step IV - Principal Components: Estimating GARCH models

In this section, a GARCH(p, q) process is estimated on the time series of the residuals derived from the ARMA models used on the PCs. The order p and q of the GARCH process is defined relying on the significance of the coefficients and the insights gathered from the preceding section. After fitting the GARCH process, the estimated conditional variance is used to standardise the residual. A specific sub-paragraph is dedicated to the analysis of the GARCHs' parameters and their ramifications.

1. **PC-1 Residuals** (Fig. 31 and Fig. 32): a GARCH(1,2) appears to be the best model to fit the squared residuals from the MA(1) process fitted before. Once divided by the conditional variance, the squared residuals show no sign of correlation with their past values in the ACF and PACF. The comparison between the QQ-plots of the original residual and the standardised exhibit that the data standardisation reduces the number of extreme values in the two tails, albeit they remain too heavy to be consistent with a normal distribution. An asymmetric component has been tested but resulted not significant;

2. **PC-2 Residuals** (Fig. 33 and Fig. 34): again a GARCH(1,2) appears to be the best model to fit the squared residuals from the MA(1). Once divided by the conditional variance, the squared residuals show no sign of correlation with their past values in the ACF and PACF. The comparison between the QQ-plots of the original residual and the standardised exhibit that the data standardisation reduces the number of extreme values in the two tails, albeit they remain too heavy to be consistent with a normal distribution. An asymmetric component has been tested but resulted not significant;

3. **PC-3 Residuals** (Fig. 35 and Fig. 36): The same process seems to apply for the PC-3, regardless of the suggestions coming from the ACF and PACF. A GARCH(1,2) is still the best model for representing the squared values from the PC-3 regression. Once divided by the conditional variance, the squared residuals show no more sign of correlation with their past values in the ACF and PACF. The comparison between the QQ-plots of the original and the standardised residual exhibit that the data standardisation reduces the number of extreme values in the two tails, albeit they remain too heavy to be consistent with a normal distribution. An asymmetric component has been tested but resulted not significant;

```
Iteration:      5,   Func. Count:     42,   Neg. LLF: 15703.38002334353
Iteration:     10,   Func. Count:     80,   Neg. LLF: 15702.062103683817
Optimization terminated successfully.    (Exit mode 0)
            Current function value: 15702.061851240323
            Iterations: 12
            Function evaluations: 94
            Gradient evaluations: 12
                      Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                  resid   R-squared:                      -0.000
Mean Model:             Constant Mean   Adj. R-squared:                 -0.000
Vol Model:                      GARCH   Log-Likelihood:               -15702.1
Distribution:                  Normal   AIC:                           31414.1
Method:            Maximum Likelihood   BIC:                           31445.1
                                        No. Observations:                3597
Date:                Fri, Nov 22 2019   Df Residuals:                    3592
Time:                        16:02:48   Df Model:                           5
                            Mean Model
============================================================================
                 coef    std err          t      P>|t|  95.0% Conf. Int.
----------------------------------------------------------------------------
mu             0.0575      0.297      0.194      0.846 [ -0.524,  0.639]
                              Volatility Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
omega          5.3980      2.244      2.406  1.615e-02     [  1.000,  9.796]
alpha[1]       0.0587  1.321e-02      4.444  8.819e-06 [3.281e-02,8.459e-02]
beta[1]        0.6183      0.153      4.030  5.582e-05     [  0.318,  0.919]
beta[2]        0.3097      0.145      2.133  3.296e-02   [2.507e-02,  0.594]
============================================================================
```

Figure 31: PC-1 residuals Estimated Garch(1,2).

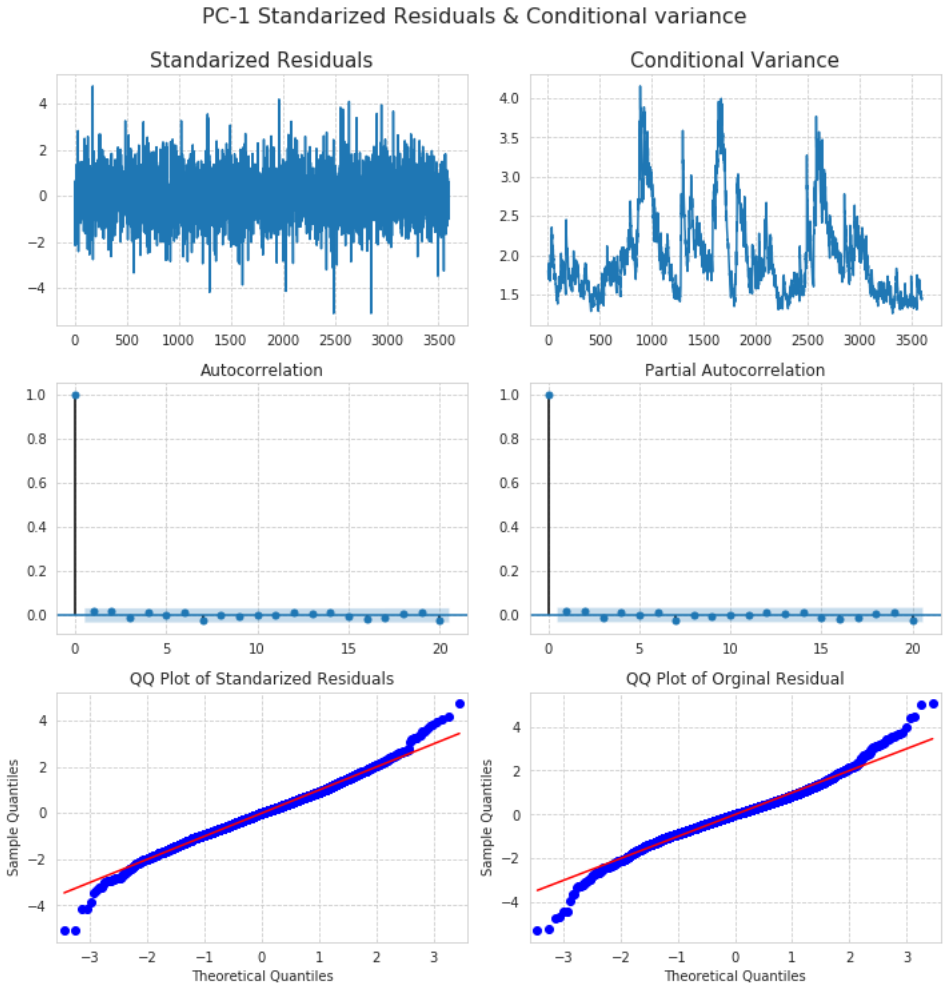

Figure 32: PC-1 Standardized Residual and Conditional Variance.

```
Iteration:      5,   Func. Count:     42,   Neg. LLF: 11227.121488838075
Iteration:     10,   Func. Count:     81,   Neg. LLF: 11222.493176468557
Optimization terminated successfully.    (Exit mode 0)
            Current function value: 11222.477800909866
            Iterations: 13
            Function evaluations: 103
            Gradient evaluations: 13
                      Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                  resid   R-squared:                      -0.000
Mean Model:             Constant Mean   Adj. R-squared:                 -0.000
Vol Model:                      GARCH   Log-Likelihood:               -11222.5
Distribution:                  Normal   AIC:                            22455.0
Method:            Maximum Likelihood   BIC:                            22485.9
                                        No. Observations:                 3598
Date:                Fri, Nov 22 2019   Df Residuals:                     3593
Time:                        16:00:34   Df Model:                            5
                            Mean Model
==============================================================================
                 coef    std err          t      P>|t|     95.0% Conf. Int.
------------------------------------------------------------------------------
mu             0.1227  7.541e-02      1.628      0.104 [-2.506e-02,  0.271]
                          Volatility Model
==============================================================================
                 coef    std err          t      P>|t|    95.0% Conf. Int.
------------------------------------------------------------------------------
omega          0.4627      0.198      2.338  1.937e-02 [7.488e-02,  0.851]
alpha[1]       0.1223  2.571e-02      4.759  1.946e-06 [7.195e-02,  0.173]
beta[1]        0.5030      0.110      4.580  4.651e-06   [  0.288,  0.718]
beta[2]        0.3691      0.110      3.364  7.684e-04   [  0.154,  0.584]
==============================================================================
```

Figure 33: PC-2 residuals Estimated Garch(1,2).

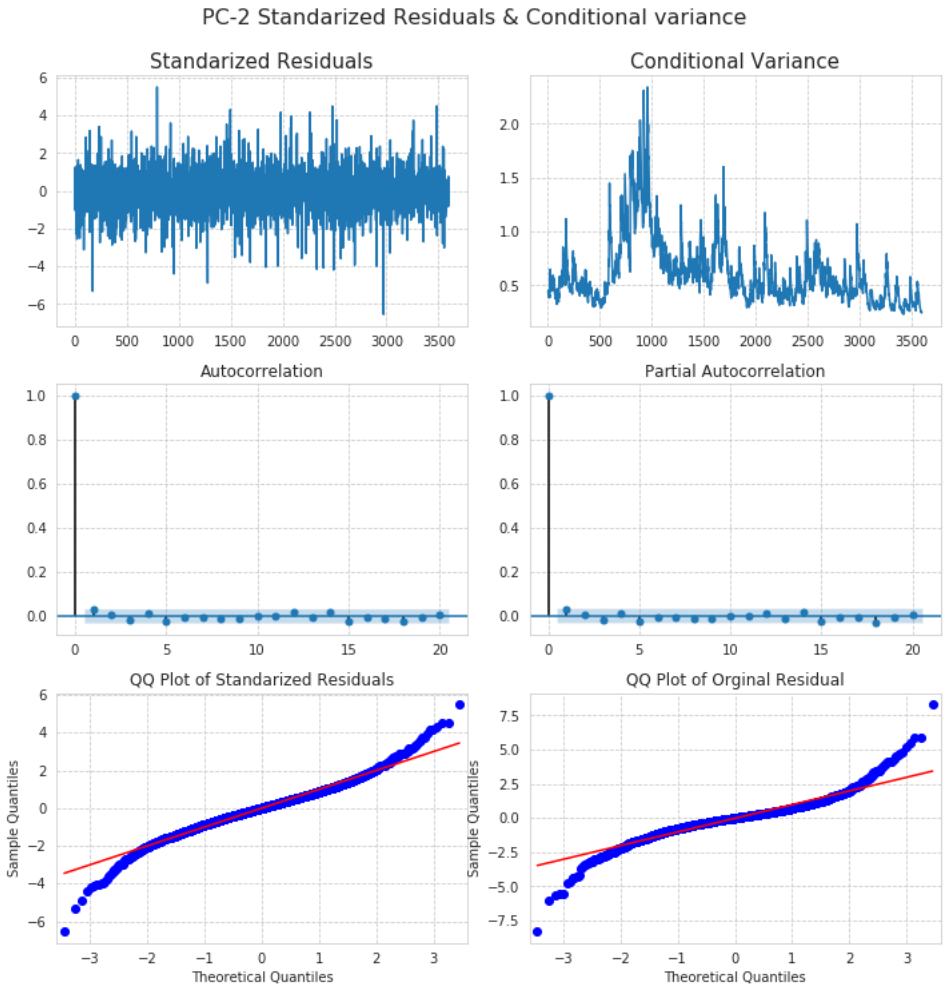

Figure 34: PC-2 Standardized Residual and Conditional Variance.

```
Iteration:      5,   Func. Count:     42,   Neg. LLF: 8915.473086328144
Iteration:     10,   Func. Count:     82,   Neg. LLF: 8892.956504524112
Optimization terminated successfully.    (Exit mode 0)
            Current function value: 8892.974020033122
            Iterations: 14
            Function evaluations: 114
            Gradient evaluations: 13
                      Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                  resid   R-squared:                      -0.001
Mean Model:             Constant Mean   Adj. R-squared:                 -0.001
Vol Model:                      GARCH   Log-Likelihood:               -8892.97
Distribution:                  Normal   AIC:                           17795.9
Method:            Maximum Likelihood   BIC:                           17826.9
                                        No. Observations:                3598
Date:                Fri, Nov 22 2019   Df Residuals:                    3593
Time:                        16:00:21   Df Model:                           5
                            Mean Model
============================================================================
                 coef    std err          t      P>|t|     95.0% Conf. Int.
----------------------------------------------------------------------------
mu            -0.0970  3.771e-02     -2.573  1.008e-02 [ -0.171,-2.312e-02]
                              Volatility Model
============================================================================
                 coef    std err          t      P>|t|     95.0% Conf. Int.
----------------------------------------------------------------------------
omega          0.0453  5.760e-02      0.786      0.432 [-6.761e-02,  0.158]
alpha[1]       0.0891  2.082e-02      4.278  1.886e-05  [4.826e-02,  0.130]
beta[1]        0.1866  4.804e-02      3.883  1.032e-04  [9.239e-02,  0.281]
beta[2]        0.7244  4.059e-02     17.846  3.089e-71    [  0.645,  0.804]
============================================================================
```

Figure 35: PC-3 residuals Estimated Garch(1,2).

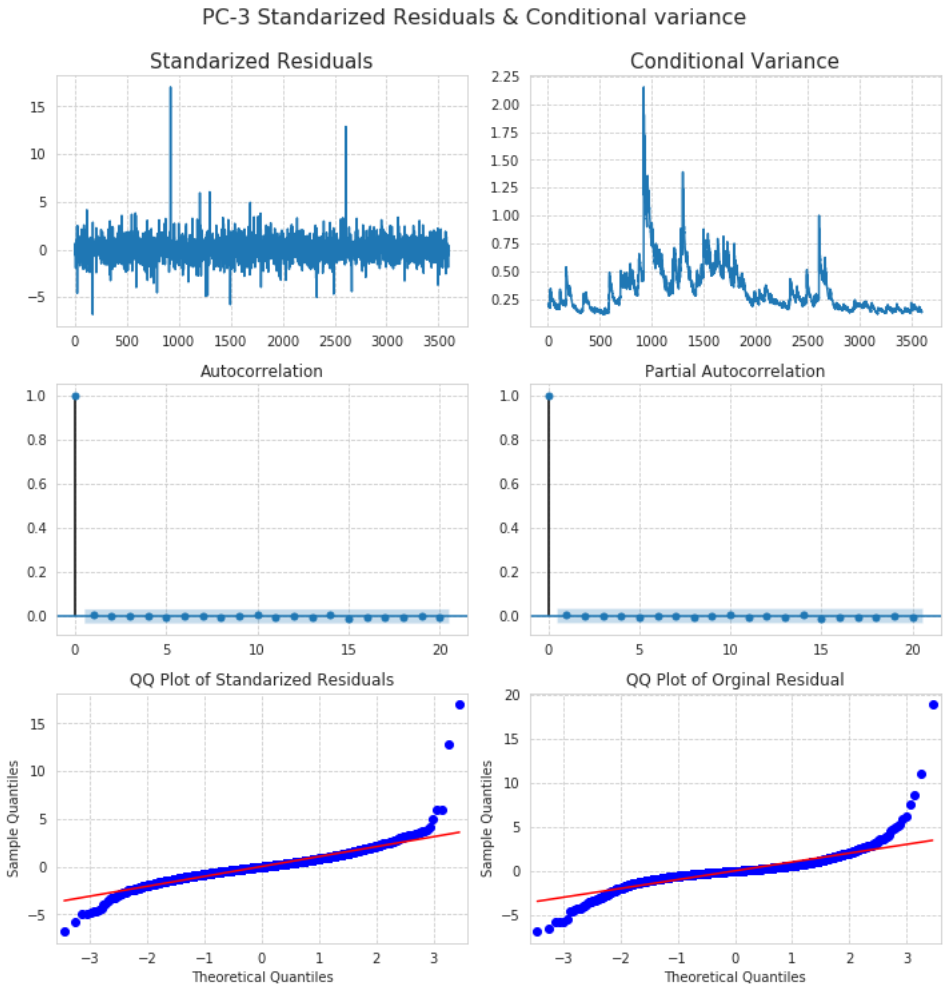

Figure 36: PC-3 Standardized Residual and Conditional Variance.

5.4.1 GARCH Coefficients analysis: Integrated-GARCH?

A GARCH(p, q) is said to be **integrated** when the sum of its coefficients is equal to 1 [Eq.3]:

$$\sum_{n=1}^{p} \alpha_n + \sum_{m=1}^{q} \beta_m = 1 \tag{3}$$

Under the general condition $\sum \alpha + \sum \beta < 1$, the volatility itself is mean reverting, and it fluctuates around σ, the square root of the unconditional variance [Eq.4]:

$$\sigma^2 = \frac{\omega}{1 - \sum_{n=1}^{p} \alpha_n - \sum_{n=1}^{q} \beta_m} \tag{4}$$

Where ω is the constant of the GARCH process. However, when the condition outline in Eq.3 is satisfied, the conditional variance is un-determined and not mean-reverting anymore (i.e. shocks to the variance are persistent). The three GARCH(1,2) models used on the PCs have levels in their coefficient very close to unity.

	Coefficients			
	Alpha-1	Beta-1	Beta-2	Tot
PC-1	0.0587	0.6183	0.3097	**0.9867**
PC-2	0.1223	0.503	0.3691	**0.9944**
PC-3	0.0891	0.1866	0.7244	**1.0001**

Figure 37: GARCH estimated Coefficients.

In order to test if the sum of the coefficients is equal to 1, a likelihood ratio (LR) test is performed, according to the following steps (see Greene[Gre99] and Hamilton[Ham94]):

1. Estimate the unrestricted GARCH(1,2). Obtain the model log-likelihood L_{unres};
2. Estimate the GARCH(1,2) subject to the restriction $\alpha_1 + \beta_1 + \beta_2$=1. Obtain the model log-likelihood L_{res}; and
3. Calculate the likelihood ratio statistic:

$$LR = -2 \ln \left(\frac{L_{res}}{L_{unres}} \right) \tag{5}$$

Where:

$$H_0 : \alpha_1 + \beta_1 + \beta_2 = 1 \text{ and } H_1 : \alpha_1 + \beta_1 + \beta_2 < 1 \tag{6}$$

and:

$$L = -\frac{T}{2} \left(1 + \log(2\pi) + \log \left(\frac{\hat{\epsilon}' \hat{\epsilon}}{T} \right) \right) \tag{7}$$

Since there is one linear restriction that distinguishes the restricted from the unrestricted model, under the null hypothesis that the restriction holds in the data, the LR statistic will follow a $\chi^2(1)$ distribution.

The values of the log-likelihood function for the restricted and the unrestricted model are calculated using the EViews Statistical Software. The statistics and the relative P-values are reported in the Fig. 38

	LR unres	LR res	LR_stat	P_Value
PC_1	862.110	848.080	0.033	0.856
PC_2	5344.383	5316.142	0.011	0.918
PC_3	7683.505	7648.391	0.009	0.924

Figure 38: LR test for I-GARCH.

The results from the test produce strong evidence that the variance of the PCs can be modelled as an Integrated-GARCH. All the P-values confirm with wide margins that the H_0 cannot be rejected.

Note: The GARCH models describing the PCs conditional variance display clear signs of being **integrated**, as the tests above show. This feature of the PCs' residuals variance has serious implications on the PCA use to model the interest rate Yield Curve. Since the variance of the PCs is not mean reverting, both positive and negative shocks are supposed to be long-lasting. This means, for instance, that when a sharp increase in the conditional variance is observed, rather than converging back to an average, it is more likely to signal a permanent shift in the series' behaviour.

6 Conclusions

The paper studied some of the consequences of applying the PCA to model to the daily difference observed in the European AAA-rated Government bond Yield Curve and the time series characteristics of the first three principal components extracted.

As a result of the analysis, two main conclusions can be drawn:

- **The information loss is not uniformly distributed across all the tenors:** the research highlighted the importance of assessing where the information is lost when selecting a subset of Principal Components to models the Yield Curve. Even when the overall loss of information is marginal, in the range of 5% for instance, if it concentrates on specific tenors or segments of the Yield Curve, it implies a massive loss of information as well as the ability to reproduce the characteristics of those specific rates. As a result, any risk measurement or trade suggestion coming for these tenors relaying on PCA extracted features should not be considered or treated with an appropriated degree of confidence.
- **The Shocks on the Principal Component Variance are extremely persistent:** This study highlighted that the PCs extracted from European AAA-rated Government Bond Yield Curve exhibits strong heteroscedastic proprieties over time. After estimating a GARCH process on the squared residuals of the PCs, a test on the coefficients pointed out that the conditional variance models are integrated. This finding has important implication for the use of PCA on the Yield Curve. For instance, in estimating the total VAR of a fixed income portfolio using the PCs, the use of long term variances of the PCs (as in Hull [Hul12], Page 493) can be misleading as the conditional variance processes have no average variance they converge to. On the other hand, shorter-term estimation of the variance can provide a better gauge of the actual risk of the portfolio.

References

[Cru19] Gospodinov N. Crump, R.K. Deconstructing the Yield Curve. *SSRN Journal*, 2019.

[Gre99] William H. Greene. *Econometric Analysis - 4th Edition.* Pearson, 1999.

[Ham94] James Douglas Hamilton. *Time Series Analysis.* Princeton University Press, 1994.

[Hul12] John C. Hull. *Options, Futures, and other Derivatives - 8th Edition.* Prentice Hall, 2012.

[PG12] Marion Pelata and Panos Giannopoulos. PCA Unleashed. *Credit Suisse Securities Research and Analytic*, 2012.