

SOFIA UNIVERSITY “St. KLIMENT OHRIDSKI”
FACULTY OF ECONOMICS AND BUSINESS
ADMINISTRATION

Statistical Methods for Risk Analysis
2015 - 2016



Project

Theme:

Using Extreme Value Theory and Copulas to Evaluate Market Risk

Student:

Maria Todorova Zlateva
Fac. № 78986
Statistics, Econometrics and Actuarial Science

Lecturer:

Assoc. Prof. Dr. Boyan Lomev

Sofia
15th February 2016

Table of Contents

Data Acquisition and Data Preparation	3
Calculation of Returns	6
Approaches to Estimation and Forecasting of Time Series. Box-Jenkins Approach. VAR Models	8
Box-Jenkins Approach	9
Identification of the Process.....	9
Specification of the Model.....	13
Estimation of the Model and Hypothesis Testing.....	17
Extreme Value Theory and Semi-Parametric CDF	25
Estimation of the t Copula.....	28
Simulation of the Portfolio Returns Using t Copula	31
Risk Measures. Value-at-Risk. Average Value-at-Risk.....	33
VAR Model	37
Appendix	39
Bibliography	56

Data Acquisition and Data Preparation

- This first step includes gathering and processing the data. Since the aim is to evaluate market risk of stocks and shares, a useful source is Yahoo Finance. For the purpose of evaluating market risk, the common stocks of six globally significant financial institutions are used – **HSBC, Banco Santander, Deutsche Bank, Unicredit Group, BNP Paribas, Bank of America**¹.
- Another essential part of data gathering is the stock exchange. Due to the fact that the different stock exchanges operate in different currencies and in different days, New York Stock Exchange (NYSE) is singled out to be the one analyzed.
- The period for which we will analyze the time series data is 03-Jan-2005 till 09-Nov-2015 which contains **567 weekly observations** for each index. For each week the average of open and close prices is taken and thus the index is calculated. In MatLab the indices are organized in matrixes.
- Dates should be transformed so as to be used in MatLab format. In Excel they are transformed in series format and then they are re-coded in MatLab with the help of the command:

```
Dates = x2mdate(dates)
```

- So as to visualize the dates in the usual **Date format**, the following command is executed:

```
datestr(Dates)
```

The result is as follows and it is given by the default variable ans

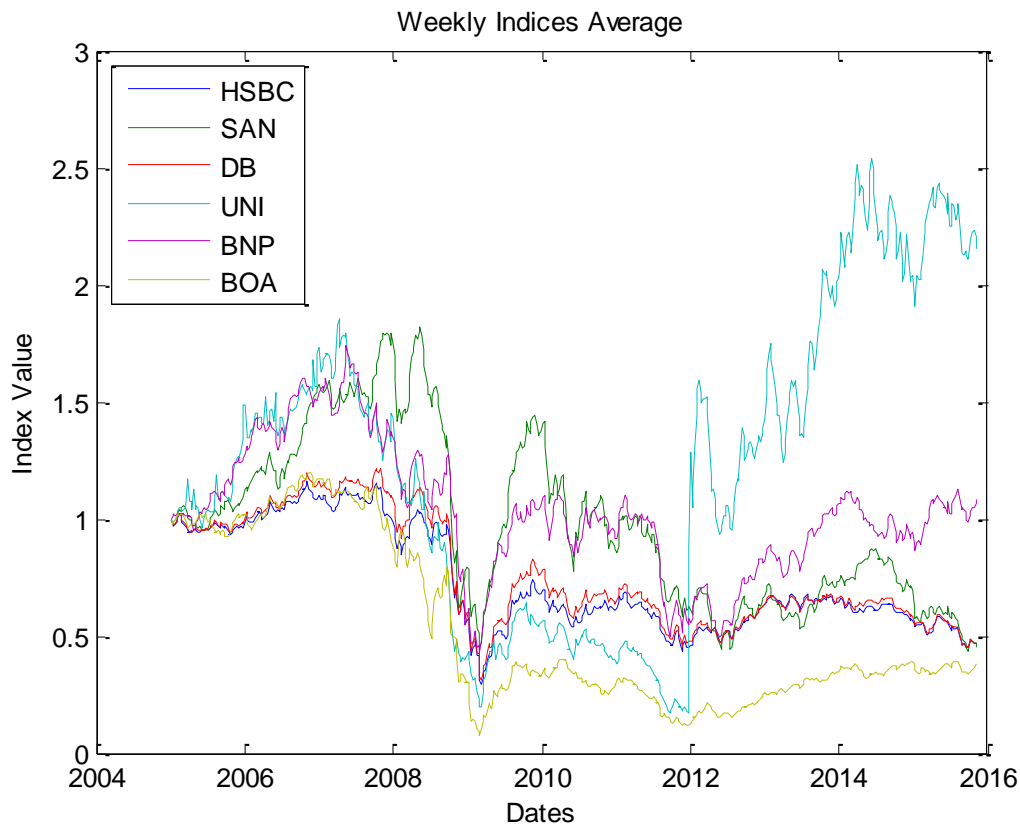
```
ans =
```

```
03-Jan-2005  
10-Jan-2005  
18-Jan-2005  
24-Jan-2005  
31-Jan-2005  
07-Feb-2005  
14-Feb-2005  
.  
.  
.  
.  
.  
.  
.  
.  
.  
.  
09-Nov-2015
```

- All observations are organized in the file `myportfolio_78986.mat`

In order to have a look at the data, it should be visualized by a graph tracing it through the period 03-Jan-2005 till 09-Nov-2015:

¹ In the research afterwards they are mentioned as their abbreviations respectively: HSBC, SAN, DB, UNI, BNP, BOA.



It is obvious that the three indices (UNI, BNP and BOA) have been following upward trend since 2012. The other two indices are downward sloping. Another interesting fact is that the discrepancy between UNI and the rest is getting larger. Therefore, it is natural to assume that it is an outlier in this sample. This fact is going to be checked by further analysis.

The fact that all six indices have trends (downward or upward in the different period of time) implies **non-stationarity**. Moreover, these indices resemble **random walk processes**, which is the classic example of non-stationary process. In order to be sure in this assumption, a formal test is needed. Such test is the Augmented Dickey-Fuller Test. The null hypothesis is that the process which generates the time series possesses unit root:

$$H_0: \rho = 1$$

This hypothesis is equivalent to random walk or non-stationarity. In other words, by applying this test, we will check whether the each index is non-stationary. This is done by the **loop function**:

```
K = ret2price(price2ret(Data));

for i=1:6
    [H,p] = adftest(K(:,i))
end
```

K is the matrix of the indices with dimensions 567x6. The resulting p-values of the test are summarized in the table below:

<u>HSBC</u>	<u>SAN</u>	<u>DB</u>	<u>UNI</u>	<u>BNP</u>	<u>BOA</u>
0.1470	0.3380	0.1713	0.7737	0.5530	0.0773

All p-values are greater than 0.05 which is the default level of significance α . This means that in all six cases we cannot reject the null hypothesis that each of the indices **has a unit root**. Therefore, we stick to the null hypothesis which states that all indices **are not stationary**.

Calculation of Returns

Return in this case is actually the capital gain. Strictly speaking the return has two parts: dividend yield and capital gain. The former is the percentage change in the dividends received from the stock while the latter is the speculative revenue which comes from the changes of market price. Since this case concentrates on the market value, it is natural to focus on the capital gain. It is also define as the percentage change:

$$R\% = \frac{P_1 - P_0}{P_0} \times 100 \quad (1)$$

Another approach is by using logarithms since they are good approximation of percentage change. Furthermore, log-log model are usually used in regression analysis in estimating growth rates and elasticities. As it is shown in (1) since we divide by P_0 the returns matrix will be 566×6 . We lose one observation due to the calculation². This is not desirable in small sample. The command in MatLab is:

```
returns = price2ret(Data);
```

In addition to that we execute another command:

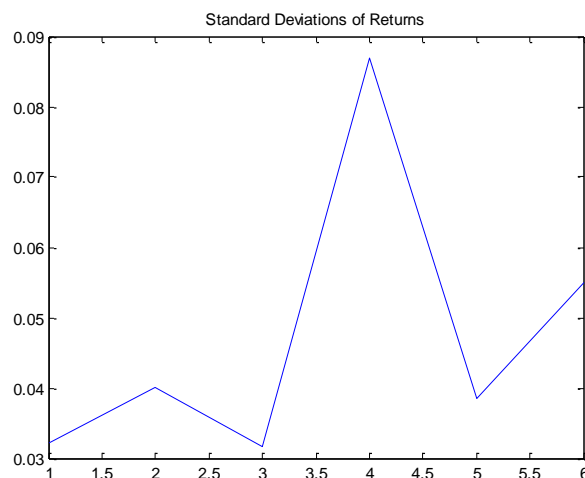
```
T = size(returns,1);
```

It computes an auxiliary variable T which corresponds to the number of rows of the returns matrix – T = 566.

In order to check for outlier in volatility, standard deviations should be estimated as a row vector 1×6 . Since there are six indices a **loop function** can be used:

```
for i=1:6  
stdev(:,i) = std(returns(:,i));  
end
```

And the following graph emerges:



² The matrix previously obtained which is called K, reserves the same of rows as the original dataset 567. Its function is `ret2price(price2ret(Data))` which is the same as $y^{-1}(y(x))$ and it is called inverse function.

Like the previous graph, the index UNI stands out as an **outlier index**. There are two reasons for this fact. The first one is that the volatility is greater. This is obvious from the previous illustration. The second is the different scale of the prices. They are smaller nominally than the others and this is transformed to the returns, too.

The standard deviation follows the same scale as the original data. Besides, this is one of its good features. This measure has, nonetheless, a setback. It is a measure of volatility and therefore it gives the information about the movements of the random variable in both directions (downward and upward). So this is not a reliable measure in terms of risk since risk is associated only with the downward movement. This is why Fr. Fabozzi, St. Stoyanov and Sv. Rachev consider the standard deviation as a good measure of “*uncertainty*” rather than “*risk*”³.

This transformation from prices into returns is remedy of non-stationarity. We are to investigate this issue in the next section.

³ Fabozzi Fr., Stoyanov St., Rachev, Sv. (2008)*Advanced Stochastic Model, Risk Assessment and Portfolio Optimization*

Approaches to Estimation and Forecasting of Time Series. Box-Jenkins Approach and VAR Models.

Since we have obtained the returns of each index and formed a returns matrix, we should find a method or an approach of how to estimate and reliably forecast these 6 indices. This is needed in the process of evaluation of market risk.

Time series econometrics offers many solutions to this problem but there are two main branches of methods that are formed in the literature. They are Box-Jenkins approach and VAR models.

Box and Jenkins proposed that the data should “speak for itself”. This implies that by using historical realizations (lags) of the random variable, we can catch some patterns of its behavior and thus model its future realizations. This is the main idea behind AR, MA, ARMA and ARIMA models. The algorithm behind this approach is **iterative**. The steps which should be followed are:

1). **Identification** of the process – by implementation of tests for stationarity, plotting the data, using sample ACF and PACF the process can be singled out as AR, MA, ARMA or ARIMA. Furthermore, we can determine the lags which we should want to include.

2). **Estimation** of the model – Having specified the model, we should estimate the respective parameters in it. We can choose an estimation technique (OLS, ML etc.) and obtain the values of the parameters as well as their standard errors and t-statistics.

3). **Diagnosis checking** – one of the most essential part of the Box-Jenkins approach is testing for statistical significance. No matter what is the value of a parameter, if it is not statistically significant, it loses its role in the model. In addition, if the residuals are not white noise, then the model is not accurate and adequate enough. The reason is that there should be discrimination between deterministic and purely random part. If this is not done, the model is inappropriate and should be specified again from the very beginning – at step 1. That’s why the Box-Jenkins methodology is iterative procedure.

4). **Forecasting** – this step is achieved only if the model is correctly specified and tested. Forecasting includes simulation procedures and should be divided into short-term and long-term projection. This division will enhance our approach because short-term and long-term forecast have different methodologies.

The other category of models developed by Christopher Sims is VAR models. These models propose different idea of catching the behavior of the random variable. VAR is the abbreviation of Vector Autoregressive model. In this project we will use the abbreviation **VAR** corresponding to **Vector Autoregressive model** and the other abbreviation **VaR**, corresponding to **Value-at-Risk**. VAR uses the regression analysis as its main essence and incorporates part of Box-Jenkins approach. It tries to explain the behavior of indices not only by using their past realizations (at different lags) but also including the past realizations of the other indices (their lags). This approach does not formulate a single regression equation that should be estimated but a system of equations. VAR models are useful in prediction albeit **atheoretical**. Unlike the Box-Jenkins approach where we have some guides which model is adequate and which is not, here we do not judge this so easily. The main specification problem which is met in these models is the number of lags to be included.

VAR method is to be used in this case but only as a supportive approach. The main accent is Box-Jenkins approach.

Box – Jenkins Approach

Identification of the process

This approach (as well as VAR) is reliable only if the time series which is to be estimated is stationary. Stationarity requires that the first and the second moment of a times series should be time irrelevant and the covariance between two time periods depends only on the lags but not on time. If Y_t is a stationary process then:

$$\text{Mean:} \quad E(Y_t) = \mu$$

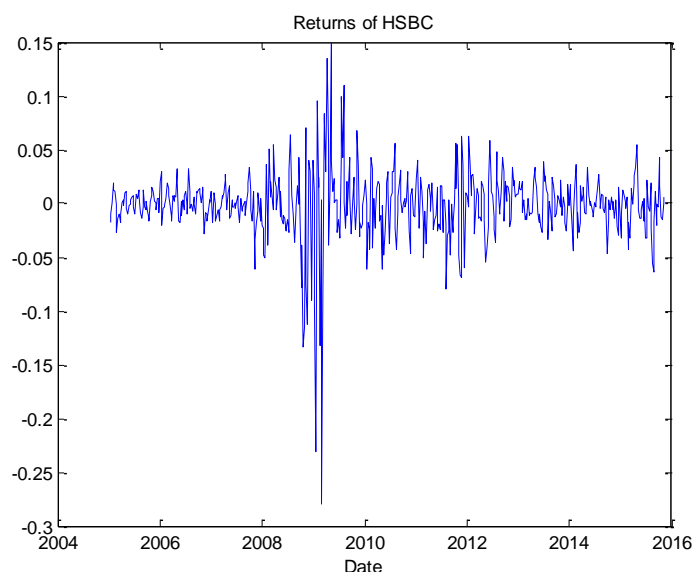
$$\text{Variance:} \quad \text{var}(Y_t) = \sigma^2$$

$$\text{Covariance:} \quad \gamma_k = E [(Y_t - \mu) (Y_{t+k} - \mu)] \quad ^4$$

Stationary processes as defined above have extremely valuable features. One of them is that they tend to return to their mean μ which makes the estimation and forecasting possible. The second of them is that if the variance is stable (time irrelevant), therefore the estimated residuals (error terms, innovations) tend to be homoscedastic. Otherwise, we fall into the case of heteroscedasticity and the residuals will not be white noise (purely random process). They will incorporate some of the patterns of behavior which should belong to the deterministic part of the model and not to the random part.

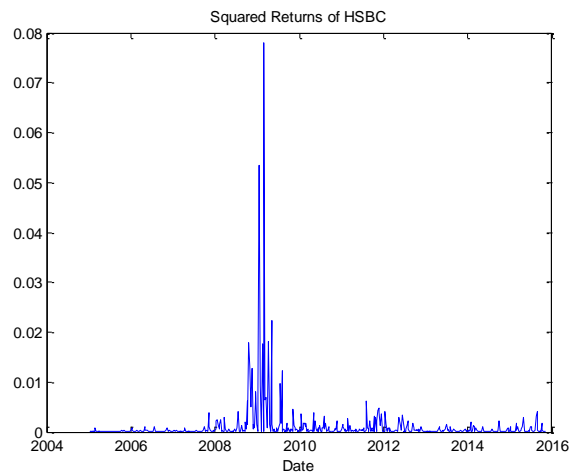
By following Box-Jenkins approach, we can examine the type of process that generates each of the six indices. This is done in three steps: 1). plotting the time series and check its stationarity 2). plotting its sample autocorrelation function and examine its significance 3). plotting its sample partial correlation function and again examine its significance. In addition to these steps, some formal test can be implemented.

Let's start with the first index corresponding to the first column vector in the returns matrix. This is the returns of HSBC. Figure below shows the process:



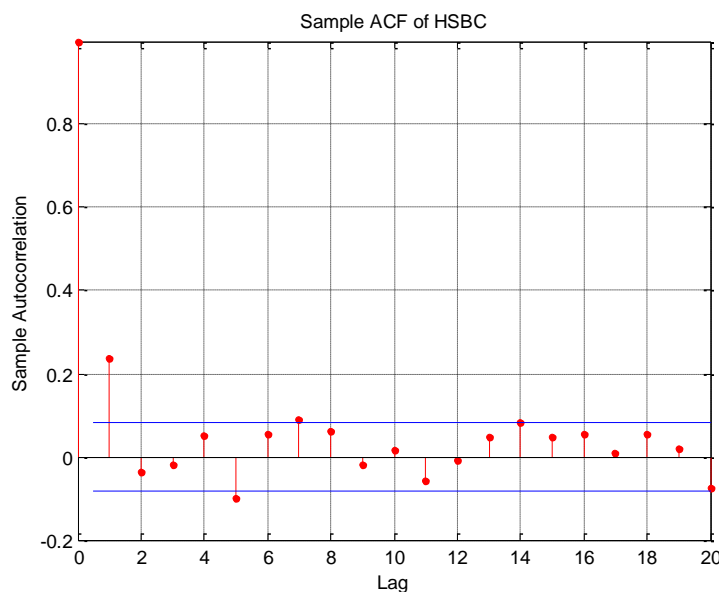
⁴ Gujarati D. (2004) *Basic Econometrics*, McGraw-Hill Company; ch. 21, p.797

The graph shows that the random variable does not vary equally in time. This implies **volatility clustering**. There are moments of low and high variables. We can use this phenomenon if we plot the squared returns of HSBC as well as the other banks:



The process seems, however, to be stationary. The same holds for the other indices. Their plots are presented in the Appendix.

If we are still not sure whether this process is stationary or not, a test is need - by applying correlograms. Let's examine the sample autocorrelation function and respectively sample partial autocorrelation function. They will give us a clue which type of model to use AR(p) or MA (q) and at which value of p and q . The graph below examines again HSBC's returns:

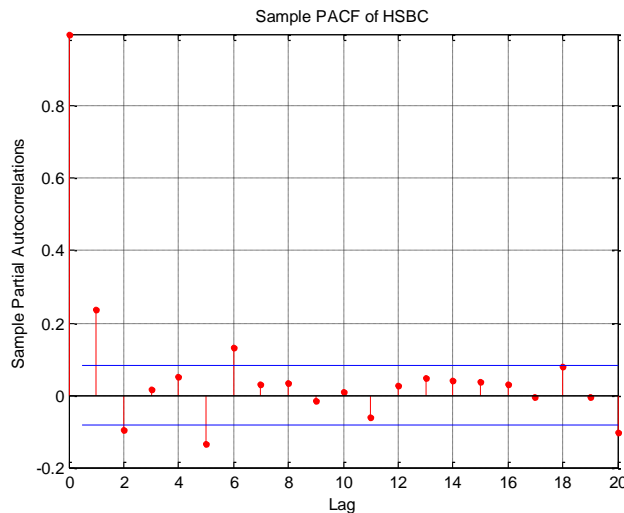


According to the sample ACF, there are two statistically significant correlation indices – one at lag 1 and one at lag 5. The others fall into the confidential interval or are on the upper/lower bound of it. Still there are considered 0. At lag 0, the index has its highest possible value of 1 since it is correlated with itself. The ACF formula is:

$$\widehat{\rho}_k = \frac{\widehat{\gamma}_k}{\widehat{\gamma}_0}, k = lags$$

Therefore, if $k=0$ which happens at lag 0, the coefficient $\widehat{\rho}_k$ is 1. In this case we have variance divided by itself. Since ρ is a correlation measure is obtain values in the interval $[-1, 1]$ and it is a measure of **linear association**. If $\rho = 0$, this does not mean lack of relationship. It means lack of linear relationship. It could be non-linear (quadratic, cubic, exponential etc.).

Another correlogram which should be taken into account is the sample PACF. It represents the correlations between any two lags disregarding the lags in between. In other words, it takes only the ends of the interval. There is a link between these partial autocorrelations and partial regression coefficients. The sample PACF of HSBC is represented in figure below:



Here again we have statistically significant coefficients. There are at lags 1, 2, 5 and 6. Both correlograms look quite alike. Obviously, this time series is stationary and can be modeled without being transformed into stationary.

In order to be completely sure whether there are significant coefficients, we can use a formal test for it. This is the **Augmented Dickey-Fuller Test** mentioned above. Unlike the previous situation which includes non-stationary processes, here we expect all p -values $< \alpha$. In addition, this will mean that we reject the null hypothesis behind the test and all the series in the return matrix are stationary. The results (p-values) are again represented in a table:

<u>HSBC</u>	<u>SAN</u>	<u>DB</u>	<u>UNI</u>	<u>BNP</u>	<u>BOA</u>
1.0000e-03	1.0000e-03	1.0000e-03	1.0000e-03	1.0000e-03	1.0000e-03

The number $1.0000e-03 \approx 0$, so we can easily reject the null hypothesis behind the test and assure ourselves that the processes behind the returns does not possess a unit root. The same results hold if we apply the test for the squared returns (p-values are the same, too).

For the first index, it is appropriate to apply AR (1), judging by both correlograms. The same holds for the other indices. Theoretically, the AR (1) model has exponentially decaying sample ACF and only one significant correlation at lag 1. This, unfortunately, is not the case here. Still, we can approximate that the case is such. To be precise, we should work with ARMA (6,5) model.

In order to make the link between the different time series corresponding to the returns, we can check whether they originate from the same distribution. This test can help us in our statistical analysis whilst making assumptions about the p.d.f of the returns. We can group the returns into

couples for this test. As a result, we will have $\frac{n(n-1)}{2}$ couples. In our case this is $\frac{6(6-1)}{2} = 15$. Let's only check whether the returns of HSBC originate from the same distribution as the returns of the other five banks. The statistical test which checks whether two processes originate from the same probability distribution is called **Kolmogorov-Smirnov Test (K-S test)**. This test has two versions: one-sample test and two-sample test respectively. The former checks whether a process originates from certain empirical distribution. In this occasion, we have only one sample which is to be tested i.e. one-sample K-S test. In the other occasion, we test whether two processes originate from one and the same distribution, without specifying which the distribution in question is.

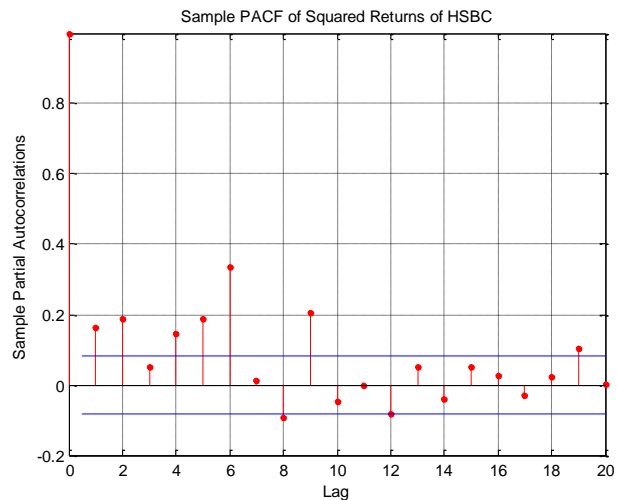
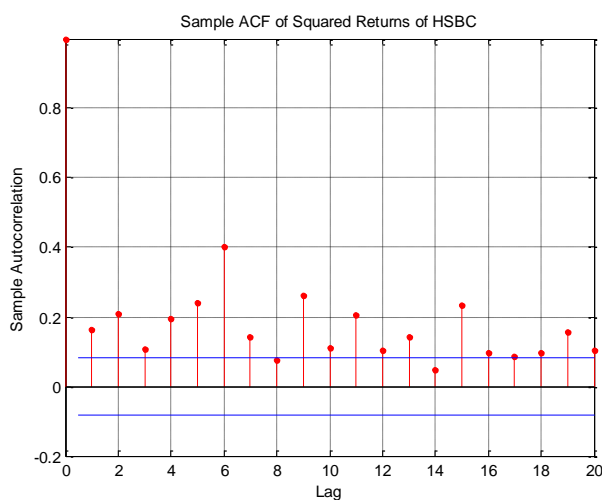
The resulting p-values of the two-sample K-S tests are summarized in the table below:

<i>Variables</i>	SAN	DB	UNI	BNP	BOA
HSBC	0.0048	0.8657	5.1108e-07	0.0039	0.0188

The null hypothesis behind the two-sample K-S test is that both processes come from the same distribution. The respective p-values show that we should reject the null hypothesis for all returns except for Deutsche Banks. Therefore, HSBC and Deutsche Bank have returns which originate from one and the same process. The full table containing the 15 tests is situated in the Appendix.

The table there is structured like a correlation table because the K-S test is a mirror test (despite the direction). The main diagonal divides the 7x7 matrix (table) into two exactly the same triangles. The results show that 8 out of 15 results support the null hypothesis and 7 out of 15 reject it. **Only the returns of Unicredit do not originate from any of the processes as the other five vectors.** We can consider them as a part of other population which in some econometric literature is the definition of an outlier.

It can be useful to check the behavior of the sample ACF and PACF of the squared returns:



Both illustrations imply some volatility of the variance i.e. heteroscedasticity. It is important to catch such type of behavior of the returns since Extreme Value Theory (EVT) which is to be used at a later stage involves Generalized Pareto Distribution (GPD) to be implemented which requires that returns should be normally distributed. No heteroscedasticity is allowed.

Specification of the Model

ARMA approach

Let's conduct an experiment with the specification of the model by ignoring the volatility clustering and the heteroscedasticity problem. Therefore, we stick to the traditional ARMA model. Due to the sample ACF and PACF of the indices, we can assume ARMA (6,5) model. After 30 interactions, MatLab finds the following solution:

ARIMA(6,0,5) Model:

```

-----
Conditional Probability Distribution: Gaussian

```

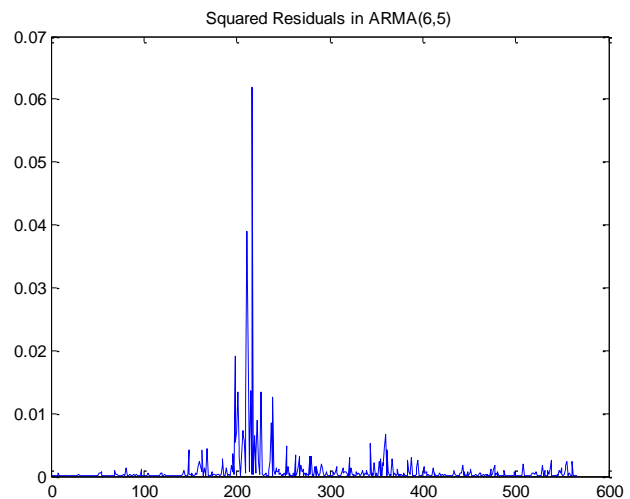
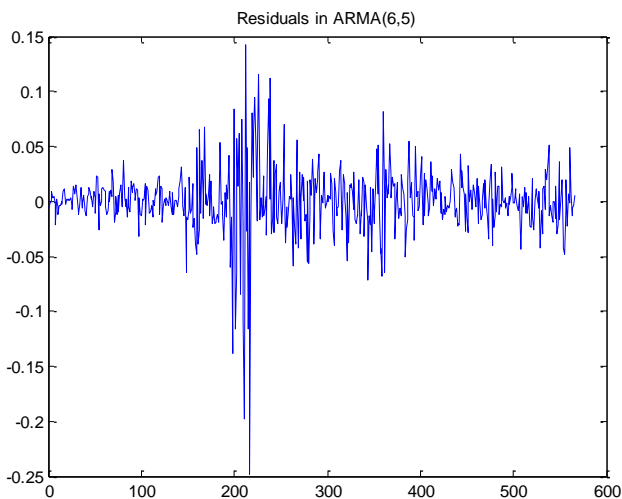
Parameter	Value	Standard Error	t Statistic
Constant	-0.00180342	0.00292799	-0.615926
AR{1}	0.123657	0.113566	1.08886
AR{2}	0.132711	0.134273	0.988368
AR{3}	-0.0126387	0.0973446	-0.129835
AR{4}	-0.229478	0.0708277	-3.23995
AR{5}	-0.600366	0.0930242	-6.45387
AR{6}	0.205925	0.0305107	6.74928
MA{1}	0.153472	0.110706	1.3863
MA{2}	-0.189136	0.12344	-1.53221
MA{3}	-0.0783519	0.0947832	-0.826644
MA{4}	0.337989	0.0866047	3.90266
MA{5}	0.600674	0.108883	5.51668
Variance	0.00092473	3.05239e-05	30.2952

According to the estimation output, there are 7 out of 12 statistically insignificant coefficients. We use the t-test in order to infer significance. The t-statistic goes hand in hand with p-values. Due to the fact that there is no p-values information in the estimation output, we can take advantage of a rule of thumb originating from the p-values:

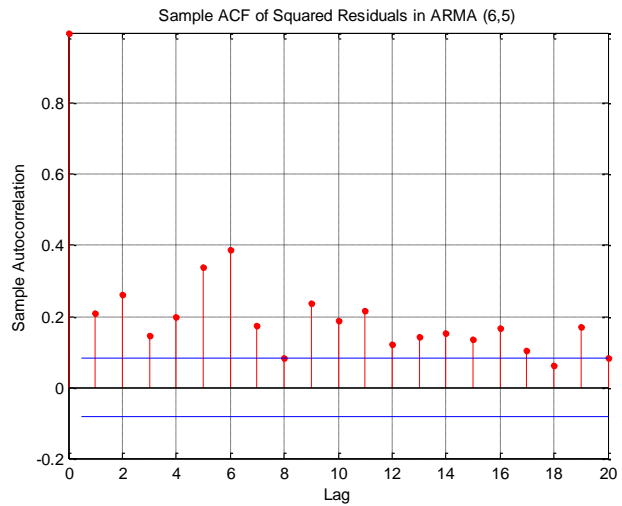
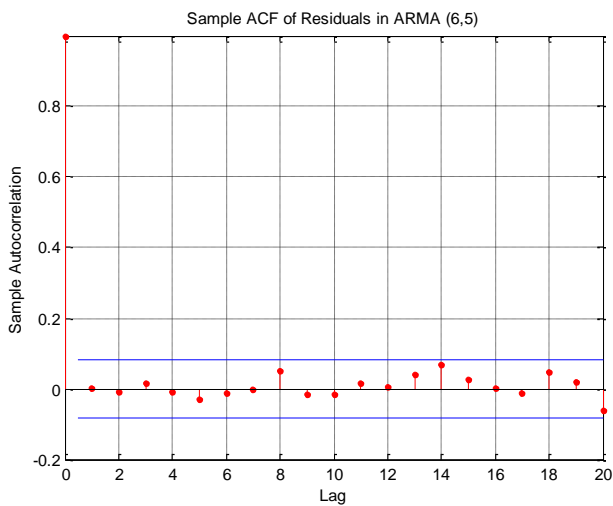
$$|t| \geq 2 \rightarrow \text{statistical significant coefficient}$$

In our case the constant terms, the respective coefficients before the 1,2,3 lags of AR, as well as the respective coefficients before the 1,2,3 lags of MA are statistically insignificant. But due to the fact that the aim of our model is forecasting, the problem of insignificance is not so essential.

Let's check the reliability of the model by testing the residuals. They should be normally distributed with no heteroscedasticity and autocorrelations. If we plot the residuals, they look the same ways as the returns and the squared returns respectively. The figures below show that we do not cure the volatility clustering:



By plotting the ACF of the residuals, we will see that they are white noise. This does not hold for the squared residuals:



Here we can apply Ljung-Box Test to check for joint statistical significance of the correlation coefficients. The null hypothesis of the test is that all ACF correlations are insignificant i.e.:

$$\rho_1 = \rho_2 = \dots = \rho_k = 0$$

The H_0 can be rewritten as white noise process. The p-value of the conducted test is 0.9513 which completely supports the statement under the null hypothesis. The p-value of the squared residuals, however, is utterly the opposite $1.0000e-03 \approx 0$ so the null hypothesis should be rejected. As a result the residuals are white noise while their squared forms are not. This is due to the volatility clustering – the variance exhibit variation. Therefore, another more accurate model is needed. This is ARMA-GARCH.

ARMA – GARCH Approach

By stepping on ARMA modeling, we upgrade the traditional ARMA model by allowing for hetescedasticity. In other words, we incorporate varying variance in the model. That’s why we used **autoregressive conditional heteroscedasticity (ARCH)** model or **generalized autoregressive conditional heteroscedasticity (GARCH)**, depending on how exactly we model the variance – by using AR or MA modeling technique. The general formula for GARCH is:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i u_{t-i}^2 + \sum_{i=1}^q \beta_i \sigma_{t-i}^2$$

As it can be seen, this is a typical ARMA equation (having both AR and MA part) but with one significant difference. The dependent variable Y is not the process itself but the variance of the residuals σ_t^2 . As a result, the independent variables are also the lagged values of the variance. Since the variance is not always known and easy to obtain, the squared residuals can be used as a proxy. That’s why we plotted the squared residuals in order to check whether we have detected varying variance. The squared residuals are usually used in many heteroscedasticity and autocorrelation tests in regression analysis for cross-sectional data. The reason for this is again the fact that they are good proxy for the variance. These tests for heteroscedasticity and autocorrelation involve auxiliary regressions by which heteroscedasticity and autocorrelation are detected. Some classic examples are Breusch-Pagan-Godfrey Test, White Test, Breusch-Godfrey (Lagrange Multiplier) Test etc.

In some case the formula can be different. If $q = 0$, we fall into a specific case where we do not have MA part but only AR part. On such occasions, the model through which we estimate the variance is just ARCH (p):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i u_{t-i}^2$$

As a conclusion it can be noted that ARMA- GARCH modeling have some superiorities over ARMA modeling since it has constant **unconditional and conditional mean and constant unconditional variance but time varying conditional variance**. These properties originate directly from the formulas above. Moreover, these features are very appropriate due to the fact that the process has a tendency to return to its mean no matter the additional information available denoted as ξ . This additional information forms the condition. In our case this is the volatility clustering. So the feature for time varying conditional variance is simply $V(u_t | \xi_{t-1})$. Since volatility clustering is the phenomenon “*in calm market periods, you expect that the stock market variation of the next days will not be dramatically high, whereas in highly volatile periods (e.g., after a crash), you expect large market movements*”⁵ the time varying conditional variance give us a hand in modeling by taking advantage of what we have previously noted.

⁵ Rachev Sv., Mann, Ch., Fabozzi Fr. (2005) *Fat-Tailed and Skewed Asset Return Distributions*; John Wiley & Sons, Inc.; p. 130.

Therefore, we can use ARMA-GARCH as a specification of our overall model due to its better features in comparison to ARMA. In order to be more specific we add another modeling element to the overall model and it is GJR-GARCH. GJR is abbreviation of Glosten-Jagannathan-Runkle Model and it includes leverage into the overall model. If negative shocks (in our market risk case, these negative shocks are losses) contribute more than positive shocks of to the volatility, then we can use GJR to model the innovations. The term “*leverage*” in time series econometrics is synonym of “*asymmetry*”. The higher the coefficient, the greater the asymmetry. GJR model catches exactly such leverage points since they determine the overall behavior of the model. Finally we specify our model this way:

```
model      = arima('AR', NaN, 'Distribution', 't', 'Variance', gjr(1,1));
nIndices   = size(Data,2);
residuals  = NaN(T, nIndices);
variances  = NaN(T, nIndices);
fit        = cell(nIndices,1);
cv=NaN(nIndices,nIndices);

options    = optimset('fmincon');
options    = optimset(options, 'Display' , 'off', 'Diagnostics', 'off', ...
                          'Algorithm', 'sqp', 'TolCon'      , 1e-7);
```

We approximate the ARMA model as AR(1) structure for all six indices and we use GJR(1,1) in other words $p = 1$ and $q = 1$ because GJR(P, Q).

Estimation of the Model and Hypothesis Testing

According to Box-Jenkins approach the second step in developing a model is its estimation. This is the most crucial part since it pre-determines the results as well as the hypothesis testing. That's why here we combine them into one section. Both stages go hand in hand. The reason for the significance of the estimation is finding the technique which is used in other to obtain the parameters of the model. They should possess specific features: being unbiased, being consistent, being efficient. Each of these properties determines the compliance of the parameters to testing hypothesis.

- 1). Being unbiased means that the mathematical expectation of the estimate equals its real value:

$$E(\hat{\alpha}) = \alpha$$

- 2). Being consistent means that in large sample, the estimate converges to its real value:

$$\lim_{n \rightarrow \infty} \hat{\alpha} = \alpha$$

- 3). Being efficient means that its variance is the smallest possible in the whole class of estimates. In other words, it is essential to be on the Cramer-Rao bound.

In the classical regression analysis, OLS technique satisfy these conditions and the estimates are BLUE (Best Linear Unbiased Estimates)⁶. However, there are situations in which the BLUE feature does not hold anymore. Such occasions are heteroscedasticity, autocorrelation, incorrect specification of the model etc. In time series modeling, OLS does not have adequate sustainability. Therefore, **Maximum Likelihood (ML)** technique is applied.

Following the specification of our model, we expect the estimation report to consist of two parts – estimation of ARIMA part (in our case it will be ARIMA (1, 0, 0)) and GJR part consisting of GARCH and ARCH estimations as well as estimation of the leverage. The estimation output for HSBC is as follows:

ARIMA (1,0,0) Model:

Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic

Constant	-0.000332801	0.000726839	-0.457874
AR{1}	0.296223	0.0392186	7.55313
DoF	5.76052	1.75683	3.27892

⁶ "Best" implies that they are efficient and have the smallest possible variance.

GJR(1,1) Conditional Variance Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	1.57011e-05	5.96474e-06	2.63232
GARCH{1}	0.835435	0.0336959	24.7934
ARCH{1}	0.0687258	0.0514637	1.33542
Leverage{1}	0.161669	0.0726655	2.22484
DoF	5.76052	1.75683	3.27892

Here again we stick to the following practical rule:

$$|t| \geq 2 \rightarrow \text{statistical significant coefficient}$$

Let's start with the ARIMA part. The model follows the structure:

$$Y_t = -0.000332801 + 0.296223 Y_{t-1} + u_t$$

The constant terms has very little impact which practically approaches 0. Moreover, it is statistically significant since $|-0.457874| < 2$. In addition, we do not need a constant term. The coefficient 0.2962 is the slope and is also known as AR (1) coefficient since it is the slope of the dependent variable at lag 1. It is highly significant. This coefficient is actually $\hat{\rho}_1$ from the sample PACF. If we return to the correlogram we can see that graphically this coefficient is indeed close to 0.30. Furthermore, the same holds for the other indices. There are no statistically insignificant slopes at lag 1 but there are no statistically significant constant terms.

The same type of analysis is conducted for the GJR part.

$$\sigma_t^2 = 1.57011e-05 + 0.0687258 u_{t-1}^2 + 0.835435 \sigma_{t-1}^2$$

The constant term here is statistically significant but has little effect on the variance because $1.57011e-05 \approx 0$. The same holds for the other indices. The GARCH part is highly significant but the ARCH part is not. The leverage estimation is significant as well. For indices 4 (UNI) the situation is different:

GJR(1,1) Conditional Variance Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	0.000570515	0.000168736	3.38111
GARCH{1}	0.440383	0.0828612	5.3147
ARCH{1}	0.426995	0.160361	2.66271
Leverage{1}	0.13221	0.210999	0.626591
DoF	3.50189	0.441503	7.93176

Both GARCH and ARCH part are statistically significant. The constant has small impact again. The only insignificant estimation is the one of the leverage. So we can conclude that there is no leverage significant at lag 1. So here again we can conclude that index 4 is an outlier – its leverage is significant. All other indices have such leverage effect and it is statistically significant.

In order to check whether the model is accurate enough, we can conduct a specification error test by which we can compare our GJR (1,1) model with another of this type e.g. GJR (2,2). We perform the **Likelihood Ratio Test**. Why do we need this test for? Usually in time series modeling, we do not know how many lags to include. Therefore, we use criteria like **Akaike** and **Schwarz** or specification tests like **LR Test**.

The LR Test compares GJR (1,1) with GJR (2,2) by using log-ratios. The first model contains less parameters than the second model. Furthermore, the first model can be characterized as **restricted** while the second one – **unrestricted**. The discrepancy between them is calculated by taking into account the number of included parameters (degrees of freedom). The so obtained statistic $\lambda \sim \chi^2$ with df = number of restrictions. The formula is the following:

$$\lambda = 2 [\text{Log Likelihood (unrestricted)} - \text{Log Likelihood (restricted)}]$$

The null hypothesis is that the newly added parameters do not belong to the model. Having estimated both models, we add logL1 and logL2 for the restricted and unrestricted model respectively. For each model there is a row vector with corresponding logL statistic. For the restricted model it is:

LogL1 =

```
1.0e+03 *
1.3689    1.1812    1.3725    0.9729    1.2117    1.2075
```

For the unrestricted the row vector is:

LogL2 =

```
1.0e+03 *
1.3695    1.1814    1.3727    0.9767    1.2138    1.2113
```

We perform the test by loop function and get the respective p-values. In the command, we add the unrestricted model first:

```
for i =1:6
    [H,p]=lratiotest(LogL2(:,i),LogL1(:,i),2)
end
```

The resulting p-values are ambiguous:

<u>Index</u>	1	2	3	4	5	6
<u>p-value</u>	0.5308	0.8158	0.7754	0.0236	0.1232	0.0235

For indices 1, 2, 3 and 5 we cannot reject the null hypothesis that the restricted model is better. Therefore, we stick to it. If we use Information Criteria (Akaike Information Criteria and Bayesian Information Criteria) for the first index in both models we obtain:

aic = -2.7297e+03 bic = -2.7124e+03

For the second model:

aic = -2.7250e+03 bic = -2.6946e+03

For index 1 we see that the respective criteria from the unrestricted model are smaller than unrestricted but they are too close. Therefore, we comply to the restricted model i.e. GJR (1,1).

Just for the check, we can have a look at the estimation output of GJR(2,2). The ARIMA part of both models is the same. All indices have significant AR(1) coefficient. However, they are differences in the GJR part. For some indices GJR(1,2) is estimated, for other – GJR(2,1) and for the rest GJR(2,2). For example, BNP Paribas GJR part estimation output is:

GJR(1,2) Conditional Variance Model:

```
-----
Conditional Probability Distribution: t
```

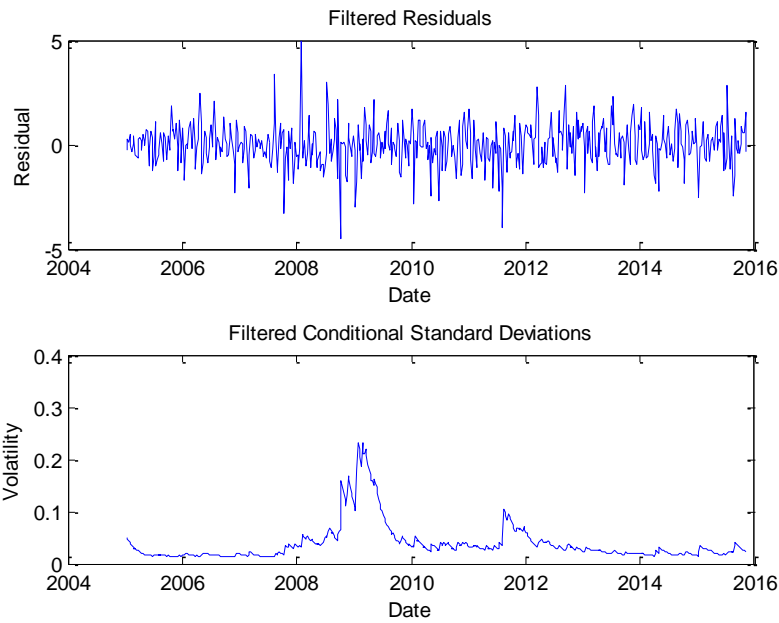
Parameter	Value	Standard Error	t Statistic
Constant	3.92426e-05	2.54181e-05	1.54388
GARCH{1}	0.811171	0.613657	1.32186
ARCH{1}	0.00684409	0.0540217	0.126691
ARCH{2}	0.0128294	0.0561049	0.228668
Leverage{1}	0.0683123	0.0870947	0.784345
Leverage{2}	0.212878	0.189214	1.12507
DoF	11.225	5.09855	2.2016

There are not significant coefficient in this part except for the degrees of freedom (they are significant everywhere). Due to the fact that it is GJR(1,2) there are two ARCH parts (at 2 lags). For HSBC the ARCH(1) is significant in GJR(1,1) while in GJR(2,2) it is not

Testing the Residuals

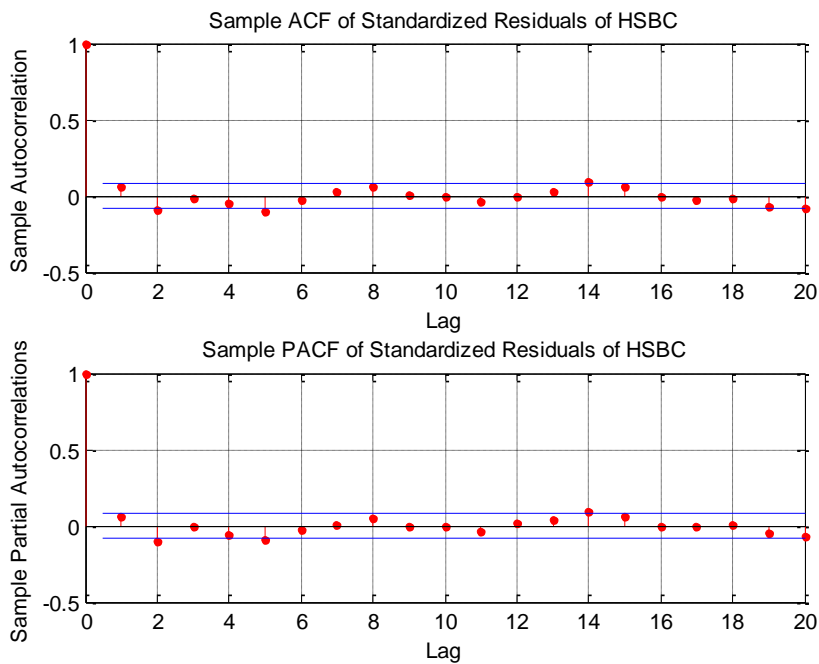
We plot the residuals and the standardized residuals (the latter are formed as rescaled by the standard deviation of the model). The second graph show that there is volatility clustering while the first plots a process which looks like white noise. We should make some test in order to check this.

```
residuals = residuals ./ sqrt(variances)
```



Now we should make some **tests for the normality of the standardized residuals**. We can apply the above mentioned tests and some others.

The first intuitive test is to plot the **ACF and PACF** of the standardized residuals (from now on we use only the standardized residuals).



Judging from the plot, all coefficients lie within the confidential interval so they can be considered insignificant. Therefore, we can conclude that the innovations follow the desired white

noise process. In order to be sure, we perform the Ljung-Box Test. The p-value is 0.341 which implies that the process is not clearly white noise. Another test which we can conduct is the two-sample K-S Test where we can compare the residuals with a random process. Initially, we generate the empirical process. They should be of the same size:

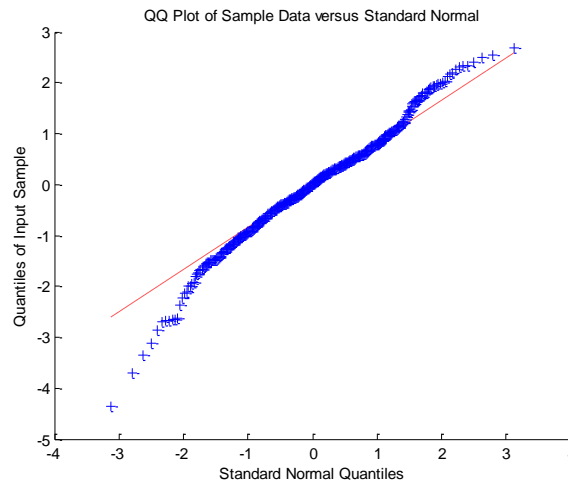
```
random = rand(566,1);
```

The results show that all residuals but index 4 originate from the same process as “random”.

Index	1	2	3	4	5	6
p-value	0.1500	0.5793	0.3950	0.0072	0.6295	0.3950

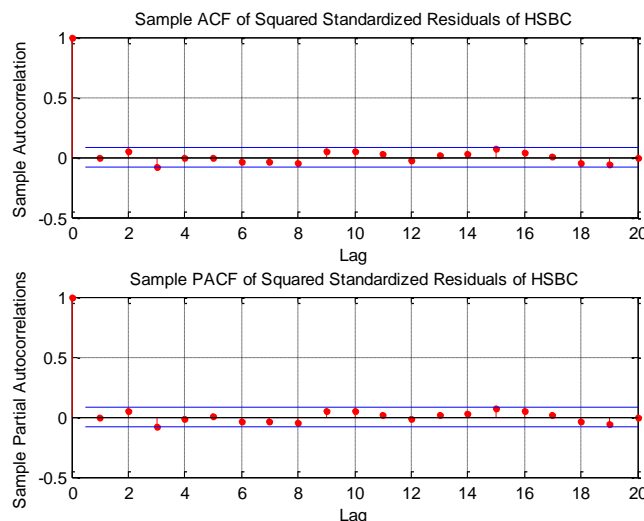
Therefore, we can tell that they are identically independently distributed i.e. i.i.d. and support the null hypothesis behind the K-S Test.

Another supportive test which we can do is the QQ plot. The function qqplot (X) displays a quantile-quantile plot of the sample quantiles of X versus theoretical quantiles from a normal distribution. Again we use the HSBC as an example:

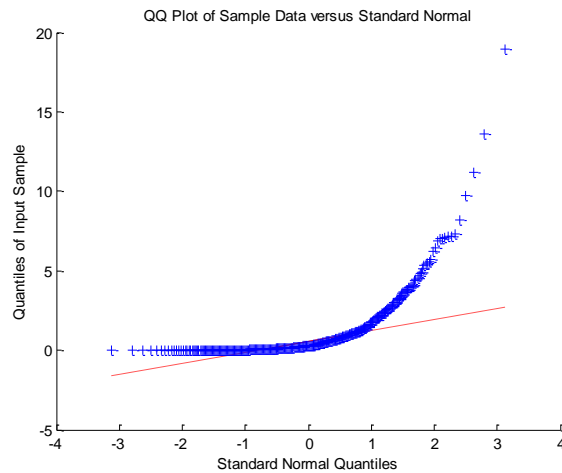


Since most residuals lie on the red line we can conclude that they are normally distributed. This QQ plot supports the K-S Test.

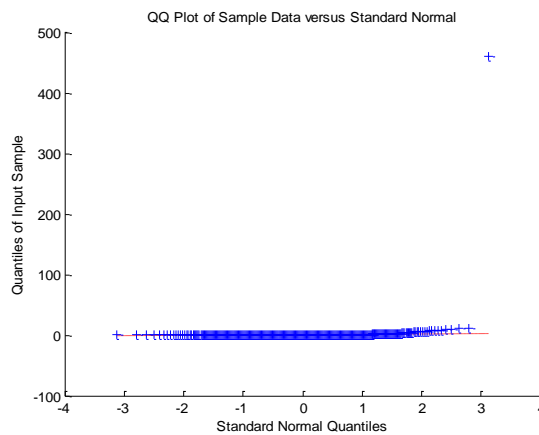
Now, we should make the same tests for the squared residuals. If they follow the same manner, we can conclude beyond any doubt that we cope with white noise.



Obviously we cope with white noise. There are no statistically significant squared coefficients. Therefore, the model suits the volatility clustering perfectly. Moreover, we continue with the QQ plot of squared residuals:



The QQ plot shows that the squared innovations are not so close to the normal distribution as the ACF and PACF suggested. The squared innovations follow quadratic trend as any quadratic (squared) function. The situation with the other indices is the same except for index 4 where the QQ plot fits perfectly to the red line:



Nonetheless, there is one outlier in the sample which characterized the whole index as an outlier.

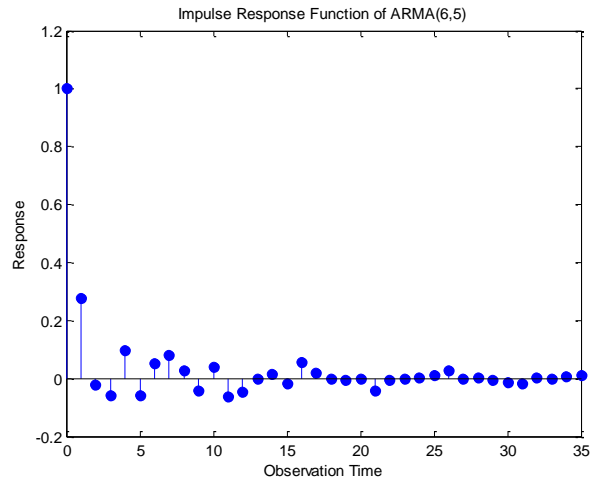
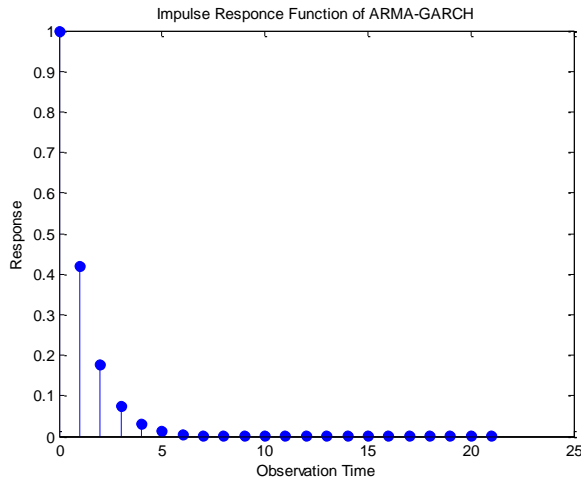
Another interesting illustration is the **impulse response function**. It informs us about the changes in the dependent variable due to changes in the innovations. The impulse response function is closely related to the polynomial lag operator:

$$(1 + \psi_1 L + \psi_2 L^2 + \dots).$$

Since it gives the relative changes in the future realization of the dependent variable in terms of one-unit change in the residuals, mathematically it can be expressed as derivatives:

$$\frac{\partial y_{t+j}}{\partial \varepsilon_t} = \psi_j$$

A much widely spread realization of the impulse response function is done by illustrations. For the ARMA-GARCH model and ARMA(6,5) the graph are the following:



The ARMA-GARCH impulse response function is exponentially decaying while the ARMA(6,5) it is like sinusoid. For the ARMA model the sinusoid shape implies that the AR part is much stronger and prevails (in fact it is 6 lags while the MA part is 5):

$$y_t = \mu + \frac{\theta(L)}{\phi(L)} \varepsilon_t,$$

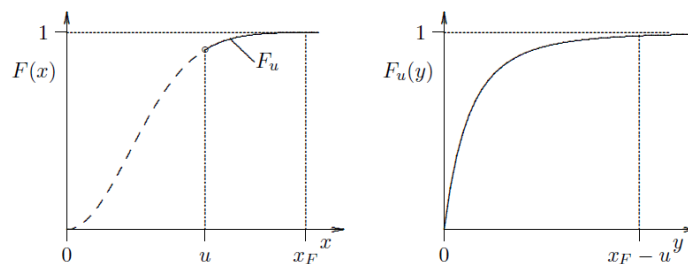
The formula implies that impulse function is linked to the coefficients of the model and their behavior.

In conclusion, we can say that the model fits the returns very well on the base of several tests and plots. As a consequence of this, we can continue with the principles of Extreme Value Theory (EVT) and Estimation of the Semi-Parametric CDF.

Extreme Value Theory. Estimation of the Semi-Parametric CDF

Extreme Value Theory (EVT) is widely used in evaluation of market risk because in this field we are interested mainly in the extreme values which the random process can take. The reason behind this is the fact that market risk is associated with capital losses. As mentioned above, the standard deviation is closely connected to market risk but its main shortcoming is that it gives both directions – positive and negative while the risk is emphasizing only on the negative deviations. Furthermore, EVT will help us in estimation of extreme negative values which the random process can take. In addition to that, we can use the estimated random process of the returns with the help of ARMA-GARCH and GJR (1,1) and model the extreme values. How can we make this and which distribution to use?

Extreme values to the fact that they are rare have specific distribution features and logically should be modeled separately from the non-rare values. This is such because these low density areas of the distributions includes small number of observations. These low density regions are called **tails**. Usually a model embraces quite well the middle regions of the distribution (the ones situated around the first central moment or the mode), but poorly captures the tails. Therefore, we need special attention there i.e. special distribution. In this situation we apply the peak over threshold approach of the EVT. This implies that we try to model the values above certain value u which we have named as threshold. The figure below illustrates the case:



The function F_u is the **conditional excess distribution function**. There is a theorem which approximates F_u with Generalized Pareto Distribution, given $u \rightarrow \infty$. **Generalized Pareto Distribution (GDP)** is a relatively good solution to the problem.⁷

GDP is a right-skewed distribution with the following probability density function:

$$y = f(x|k, \sigma, \theta) = \left(\frac{1}{\sigma}\right) \left(1 + k \frac{(x-\theta)}{\sigma}\right)^{-1-\frac{1}{k}}$$

There are three parameters: k , σ and θ where k is known as the **shape parameter**, σ is the **scale parameter** and θ is the **threshold (exceedance)**. The shape parameter k , also known as **tail index**, can take any values – positive, negative and zero. Its behavior determines the shape of the GDP: the higher the k , the slower the decay of the distribution. If we model a distribution whose tails decay exponentially (like the normal distribution), theory implies $k = 0$. Unlike k , σ is relatively stable and kept at $\sigma = 1$.

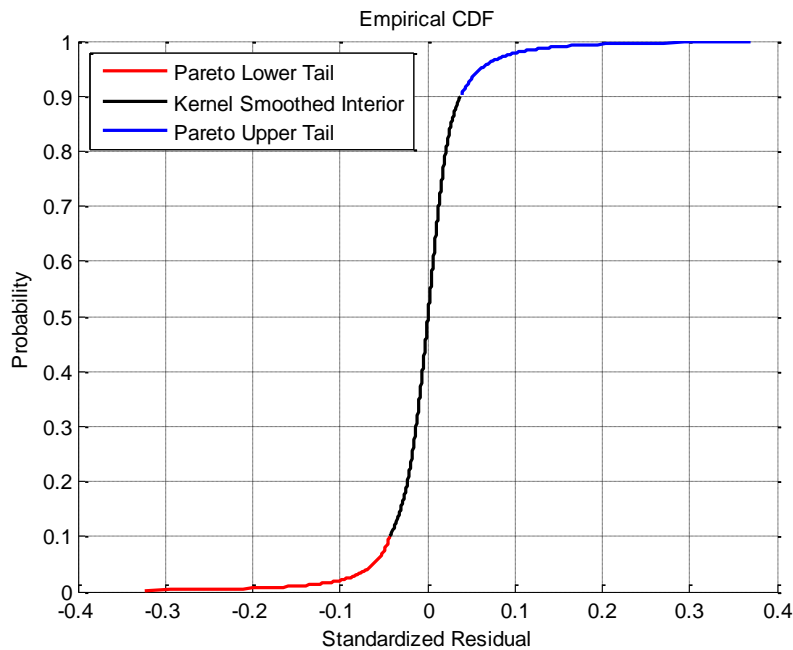
As mentioned above, we will use GPD for the lower and upper tails of the distributions. The only thing left to be considered is the middle part. It should be smoothed with Gaussian Kernel Distribution so that its initial staircase is eliminated. This is achieved by:

⁷ M. Gili, K llezi Ev. (2006) *An Application of Extreme Value Theory for Measuring Financial Risk*; p.7

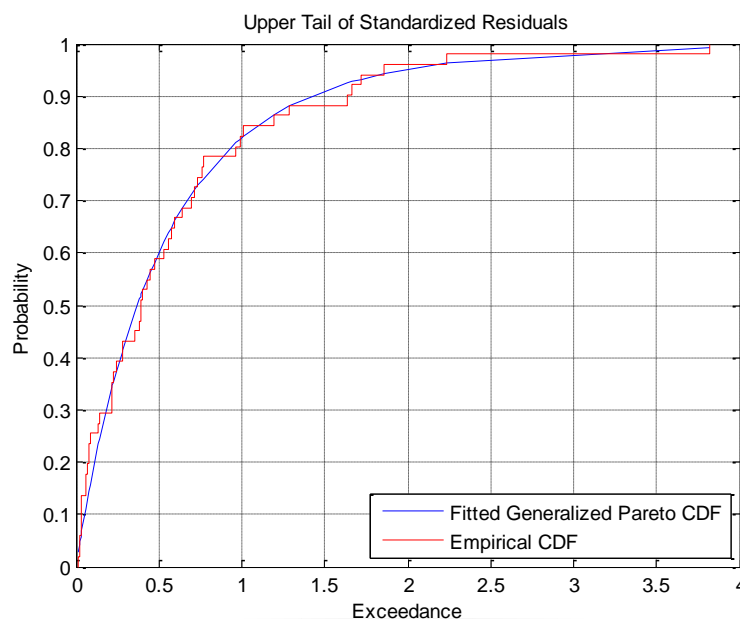
```
OBJ{i} = paretotails(residuals(:,i), tailFraction, 1 - tailFraction, 'kernel');  

    i= 1,2,3...6
```

We model the **standardized filtered residuals**. We know that they are i.i.d. and are considered white noise. They are extremely easy to be modeled by GDP. The variable **tailFraction** indicates the threshold form which we would like to model. In our case it is 0.1. Moreover, since we model both lower and upper tail, we calculate the upper tail as **1- tailFraction**. The last element of the command is the interior smoothing function. This is Gaussian Kernel. Owing to the fact that we do this for each of the indices, we develop a **loop function** again. Furthermore, sometimes this procedure is named as “smoothly truncating the tails” from the empirical distribution. Actually they are not cut but replaced with another tails from different stable distribution. This process can be visualized as below:



We can examine GDP in even further details. The aim is to see how well the GPD converges to the empirical distribution of the tails of the filtered standardized residuals in terms of the exceedance.



Obviously the cumulative distribution function of GPD is quite good approximation of the staircase empirical cumulative distribution function of the empirical distribution. Since “the stable Paretian distributions are heavy-tailed, and they have the potential to describe the heavy-tails and the asymmetry of the empirical data”.⁸ Stability in this context means that time invariant statistical features. It is important to reveal how the exceedances are obtained. The critical value above which an observation is considered part of the upper tail is 1.1497. Then they are sorted as exceedance vector having subtracted the same value of 1.1497. As a result we can see that GDP is relatively good choice of approximation. Therefore, the goodness of fit is very high.

⁸ Fabozzi Fr., Stoyanov St., Rachev, Sv. (2008) *Advanced Stochastic Model, Risk Assessment and Portfolio Optimization*;p.123

Estimation of the t Copula

So far we have come across only **univariate distributions** like normal distribution, Generalized Pareto Distribution (GPD), Gaussian Kernel, etc.. They are univariate because they show the probability density function of only one random variable. In our case we have estimated different GPD distributions for the tails of each index and respectively different Gaussian Kernel for the interior parts. These six indices corresponding to the returns of the shares of each bank actually are traded in the same stock exchange and as it has been seen in the very beginning, some of them share the same trend. Using only univariate distribution will not be enough owing to the fact that we cannot catch the mutual effects and correlations. In portfolio theory and the principles of risk diversification implies that a bunch of indices should be considered simultaneously. Moreover, this comes from the diversification principle “*not to put all the eggs in one basket*”. As a result, multivariate distributions should take part of the play as well as the univariate ones. In effect, only **multivariate distribution** can be used in order to reveal the dependence structure in which all six indices are incorporated. In addition to that, it is typically the case that the dependence is influenced by the scales of the random variables.

In the context of multivariate distribution the univariate distribution is called **marginal (probability) distribution** while the multivariate distribution is named as **joint (probability) distribution** because it reveals the *joint, mutual* probability density function. There is a mathematical link between the two distributions and it is as follows:

$$f_i(x) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_d) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_d$$

“The marginal density f_i of X_i is obtained by integrating the joint density over all variables which are not taken into consideration”⁹.

Joint distribution is just one solution to the problem. Dependence in statistics can be revealed through correlations. Denoted with ρ , the correlation measure obtains values in the interval $[-1, 1]$ and it is measure of **linear association**. If $\rho = 0$, this does not mean lack of relationship. It means lack of linear relationship. It could be non-linear (quadratic, cubic, exponential etc.). In order to circumvent the shortcomings of the correlation as a measure of dependence, copulas are widely introduced and used in applied finance. **Copula** originates from Latin and means “link” or “connections”. In fact, a copula is nothing else but a **probability distribution function** which is used so as to join or combine several marginal distributions into a joint distribution. Unlike correlations, copulas are universal tool for modeling. Among their most strong features are that they allow modeling extreme values situated in the tails. This is not possible in terms of correlations because the variance there can be infinite. In conclusion, copulas are widely used in stress-tests because of the above mentioned features.

Copulas are probability distribution functions and should be estimated and calibrated. One possible approach is by using the **Maximum Likelihood (ML)** estimation technique. The copulas is to be presented in MatLab as a matrix with the same size and rank as the one of the filtered residuals. Therefore, we use the command:

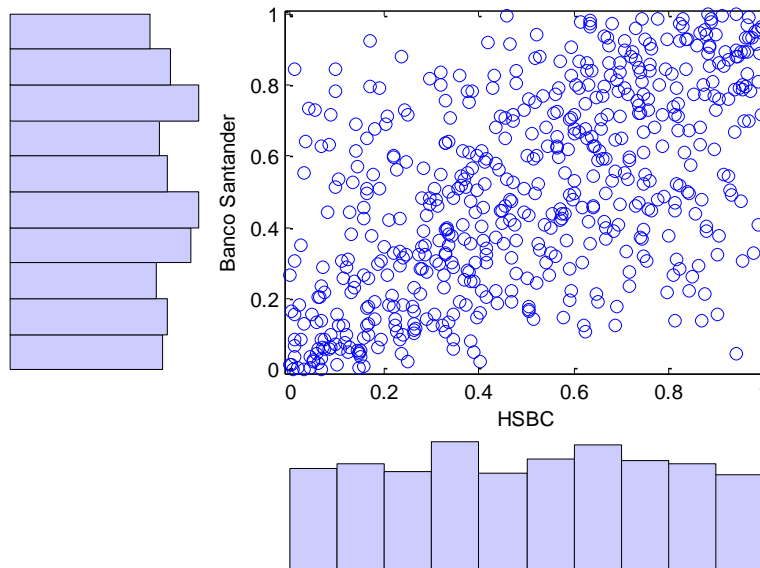
⁹ Rachev Sv., Mann, Ch., Fabozzi Fr. (2005) *Fat-Tailed and Skewed Asset Return Distributions*; John Wiley & Sons, Inc.; p. 58.

```
U = zeros(size(residuals));
```

Now the matrix contains only zeros as elements but we have fixed its size. The estimation itself is done in two stages. The first stage is:

```
for i = 1:nIndices
    U(:,i) = cdf(OBJ{i}, residuals(:,i));
end
```

The command `cdf()` returns the vector `U` for each of the indices for a certain distribution at given the filtered residuals. In other words, it computes the cdf values for the OBJ (GPD used for the tails and Gaussian Kernel for the middle part) at the values of the filtered residuals. Therefore, there is a correspondence of each element from the matrix `U` to each element of the filtered residuals matrix. The figure below illustrates what happens between `U1` and `U2` corresponding to HSBC and Banco Santander respectively:



Both histograms imply uniform distribution and the scattergram is concentrated around the diagonal which imposes positive relationship although not precisely (proportionally) positive due to the variation. The same pattern is shown for the other couples.

The second final stage actually fits the t copula:

```
[R, DoF] = copulafit('t', U, 'Method', 'ApproximateML');
```

where the matrix `R` stands for the **correlation matrix** and `Dof` – **degrees of freedom**. It shows positive correlations because all the correlation coefficients are greater than zero. The matrix `R` being results of the estimation of the t copula are denoted as $\hat{\rho}$ and are obtained by ML estimation technique. The `R` matrix as the usual correlation matrix contains two exactly the same triangles. This is due to the fact that the correlation is a two-directional function. Matlab calculates the `R` (ρ) matrix:

```
R =
    1.0000    0.5945    0.9812    0.4535    0.5623    0.5219
    0.5945    1.0000    0.7294    0.5414    0.6083    0.4306
    0.9812    0.7294    1.0000    0.5045    0.6149    0.5346
    0.4535    0.5414    0.5045    1.0000    0.6226    0.4043
    0.5623    0.6083    0.6149    0.6226    1.0000    0.5105
    0.5219    0.4306    0.5346    0.4043    0.5105    1.0000
```

The highest correlation coefficient is for the couple HSBC and Deutsche Bank (indices 1 and 3). We can further investigate this matter by using the simple correlation function and calculate the respective p-value at the default $\alpha = 5\%$. The result of the indices 1 and 3 is the following:

```
rho13 = 0.9810    p = 0
```

This implies that the correlation coefficient is **statistically significant** at $\alpha = 5\%$. Let's change the level of significance to $\alpha = 1\%$. The same results appear. It is important to note that the results are obtained by using the `corr` function with its default option of **Pearson linear correlation coefficient**. In this context for the hypothesis testing, t distribution is used.

It is good to take into account the fact that the copula gives us the opportunity to calculate the **Kendall tau τ correlation coefficient**. As the Pearson correlation coefficient it measures the association between two random variables (or vectors) but here they are ranked in terms of quantity.

```
r = copulastat('t', R, DoF)
```

```
r =
    1.0000    0.4053    0.8764    0.2997    0.3802    0.3496
    0.4053    1.0000    0.5204    0.3642    0.4163    0.2834
    0.8764    0.5204    1.0000    0.3366    0.4216    0.3591
    0.2997    0.3642    0.3366    1.0000    0.4278    0.2650
    0.3802    0.4163    0.4216    0.4278    1.0000    0.3410
    0.3496    0.2834    0.3591    0.2650    0.3410    1.0000
```

The coefficients are much smaller, due to the different calculation technique. Nonetheless, the correlation between the first and the third index is kept the largest. Kendall tau τ is used mainly in testing statistical dependence.

Degrees of freedom sometimes denoted as `nu` has the value of 13.2025. The function `copulafit` gives us the opportunity to construct a 95% (the default option) **confidence interval** for the degrees of freedom. In our case this interval is named `DoFci` and is equal to:

```
DoFci =
    13.2025 - 1.7124i    13.2025 + 1.7124i
```

where `1.7124i` is a complex number. As it is obvious the confidence interval is build around the ML estimate for the DoF and it says that in the long run, 95% of the so built intervals [`13.2025 - 1.7124i`; `13.2025 + 1.7124i`] will contain the real value of DoF.

Having obtained the copula, we can use the so structured model for simulations and forecasting.

Simulation of Portfolio Returns using t Copula

Simulation and forecasting involves the same number of step which we have already made but in opposite direction – from the last step onwards.

Having estimated the parameters of the t copula (the R matrix and the DoF), it is time to make a simulation by using this structure as a joint distribution. This step is achieved by the function `copularnd`:

```
U = copularnd('t', R, DoF, horizon * nTrials);
```

The matrix U contains the simulated standardized residuals by using t copula for each of the indices. The function `copularnd` specifies the type of copula (t copula or *Gaussian* copula) as well as the previously estimated parameters (R and DoF). Moreover, it includes another parameter and this is the number of rows in the matrix U. Usually in the function, it is given as single number n . In our case, however, it is represented as a multiplication operator of the number of independent trials (`nTrials`) and the time horizon of the forecast (`horizon`). The time horizon is usually one month ahead but since we are working with weekly data, it will be **4-period horizon**. i.e. **one month ahead**. For each of these days we make `nTrials` independent random generations. Therefore the result is the multiplication of both parameters:

```
horizon * nTrials = 8000
```

The aim of our simulation is not the matrix U because it contains the results of the t copula simulation for each index. Our focus is on obtaining the residuals which are simulated in another matrix named Z. By using the `reshape` function we extrapolate the GPD lower and upper tails and interpolate the smoothed middle part for each index at the time horizon of the projection. It is important to note that, we use the inverse semi-parametric cumulative distribution function for each index. The outcome is the simulated standardized residuals ready for incorporation into the ARMA-GARCH model where we should use the estimated parameters from the respective model.

The simulated returns are obtained by using the parameters of the estimated ARMA-GARCH model. They have the same matrix as the residuals. For the simulation we take the last values of the returns, residuals and the variance as a starting point of our projection. The main assumption which we make is that **the dependence structure is kept the same** throughout the horizon period (4 weeks ahead) and **the structure of the portfolio is not to be altered** (the weights). Besides, we assume that the portfolio consists of six **equally weighted indices** (assets):

```
weights = repmat(1/nIndices, nIndices, 1);
```

Therefore each index comprises of $\frac{1}{6}$ of the whole portfolio. The current project circumvents the issue of optimization of portfolio weights.

So far we have worked with the weekly logarithmic returns. Now is the time to transform them into cumulative returns since this is the object of our interest. Cumulative returns are actually the sum of the capital gains/losses throughout a period. In other words, it is the balance of the account covering a certain financial operation. We should make a transformation so that the

logarithmic returns (`price2ret` function) is now cumulative ones (`cumulative ret2price`). The formula is:

$$\text{Cumulative returns} = \sum(\log(1 + \text{weights} * e^{\text{simulatedReturns}-1}))$$

This transformation in fact breaks the correlation structure of the returns if any. The correlation operator is not able to support such exponential transformation and that is one of its main weak sides. This does not hold for the copula, nonetheless. It sustains this transformation and still work as a link between the indices.

Having obtained the cumulative returns now is time to measure the risk using Value-at-Risk (VaR) and Average Value-at-Risk (AVaR) measures.

Risk Measure. Value-at-Risk. Average Value-at-Risk.

There are many ways of measuring risk. So far we have concern this issue by mentioning part of this problem in the very beginning. Risk is usually associated with variance and dispersion. There is, however, crucial distinction between **risk** and **uncertainty**. The former is associated mainly with certain distribution and we simply do not know at which part of it the random variable will be realized while the latter is completely unknown – we neither know the probability distribution nor the realization of the random variable. There is a huge debate whether we exist under risk or uncertainty. The classical economic theory assumes that we live under risk.

One of the most popular measures of spread is the **standard deviation**. But as it was mentioned above it is not a measure of risk rather than of uncertainty due to the fact that it shows the realization in both directions – positive and negative.

Another measure of spread is **mean absolute deviation (MAD)**. As its name has it, it measures the absolute deviation from the mean:

$$MAD_X = \sum_{k=1}^n |x_k - EX|p_k,$$

It actually takes the absolute value of the deviation from the mean. This is an alternative to squaring the deviations as it is done in the formula of the variance. Here we have the mathematical expectation operator applied to the absolute value. One of the weaknesses of this measure is that again it is a notion concerning uncertainty. In addition, sometimes the absolute value operator is not so easy to be used.

A third measure of spread is the **semi-standard deviation (SSD)**. It upgrades the features of the classical standard deviation as follows:

$$\begin{aligned}\sigma_X^+ &= (E(X - EX)_+^2)^{1/2} \\ \sigma_X^- &= (E(X - EX)_-^2)^{1/2}\end{aligned}$$

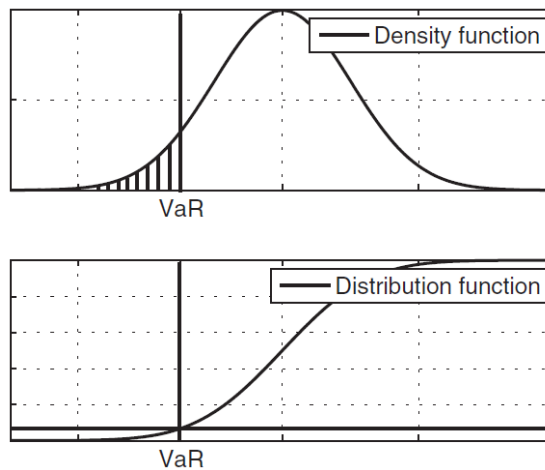
This measure treats in a different way the positive and the negative deviations from the mean. Therefore, we can characterize this measure as a measure of risk. Moreover, this statistic is not symmetric as the previous two indeed owing to the different approach to the direction of the deviations. Furthermore, this measure should be closely analyzed alongside with the probability distribution. If the distribution is symmetric (**skewness** = 0), the positive and negative deviations are equal. The same does not hold if there is skewness.

The most popular measure of risk is the **Value-at-Risk (VaR)**. It is the officially recognized measure of risk from the Basel Committee and is adopted in **Basel III requirements**. “VaR is defined as the minimum level of loss at a given, sufficiently high, confidence level for a predefined time horizon. The recommended confidence levels are 95% and 99%”¹⁰ In other words, VaR informs us that a loss greater than certain number or percentage denoted by X_ϵ will occur in the predefined time horizon with probability of ϵ . Mathematically this is written as:

$$VaR_\epsilon(X) = -\inf\{x | P(X \leq x) \geq \epsilon\} = -F_X^{-1}(\epsilon)$$

¹⁰ Fabozzi Fr., Stoyanov St., Rachev, Sv. (2008) *Advanced Stochastic Model, Risk Assessment and Portfolio Optimization*, p.182

Its definition as well as the formula implies that a useful technique of calculating VaR is by using the **inverse cumulative distribution function**. In fact, the result of the calculation of VaR is a negative number because we focus on the **lower (left) tail** of the probability density function. The minus before the above formula indicates that we are interested in the **absolute value of the loss**. Indeed VaR is not used as a measure of capital gain because this is not considered risk. The Figure below illustrates exactly the above mentioned ideas of calculating VaR:



The strong features of VaR are associated with **scaling**. If we multiple a random variable with a **constant λ** , we multiple its respective VaR with the same constant:

$$VaR_{\epsilon}(\lambda X) = \lambda VaR_{\epsilon}(X)$$

Another very strong feature of VaR is the ability to automatically reduce risk if **constant (riskless asset) C** is added to the portfolio:

$$VaR_{\epsilon}(X + C) = VaR_{\epsilon}(X) - C.$$

The above mentioned features of VaR depict it as a strong and supreme measure of risk. There are, nonetheless, some shortcomings to this measure. One of them concerns the diversification effect which the investors to achieve by constructing a portfolio. Moreover, the investors would like to have smaller VaR of the joint effect of several assets than the VaR of them separately. Unfortunately, this is not the case of VaR where synergy is experienced:

$$VaR_{\epsilon}(X + Y) > VaR_{\epsilon}(X) + VaR_{\epsilon}(Y)$$

This synergy is in conjunction with the diversification effect sought. To be precise, it does not violate or reject the existence of diversification. VaR is not capable of catching this effect. The reason is that the risk is categorized as **systematic** and **unsystematic** and in VaR we use only one type of it.

This **subadditivity problem** can be overcome by modifying the VaR measure into Average Value-at-Risk (AVaR). It is also known as **expected shortfall** or **conditional VaR**. It gives us the information about the risk above certain threshold level. It is more precise than VaR because it captures **the whole tail** and takes into account this weight. The name “average” VaR comes from the feature that this statistic **takes the average of all VaRs associated with greater values than VaR at probability ϵ** . Geometrically AVaR is the whole area under the probability density function and the VaR at the certain level. Therefore, it is calculated as average of definite integral:

$$AVaR_{\epsilon}(X) := \frac{1}{\epsilon} \int_0^{\epsilon} VaR_p(X) dp$$

MatLab does not support built-in function for VaR and AVaR so we should develop scripts for them by following the formulas from above. For one month horizon the respective measures of risks are:

Maximum Simulated Loss: 30.4438%
Maximum Simulated Gain: 22.2537%

Simulated 90% VaR: 7.8678%
Simulated 95% VaR: 10.3744%
Simulated 99% VaR: 16.3225%

Simulated 95% AVaR: 14.2094%

The results show that our portfolio has maximum loss greater than the maximum gain which may be indication of not so diversified portfolio. 90% VaR shows us that there is a probability of 10% that we can lose more than 7.8678% of our yield for the one-month period. In addition, the probability of losing more than 10.3744% is 5% and losing more than 16.3225% is 1%. The AVaR is indicative for the average (diversification) effect. The average of all VaRs larger than 95% VaR is 14.2094% of the yield in one-month period. Logically 95% VaR is smaller than 95% AVaR because the latter take the average of all VaRs greater than the former.

The results for a longer period than one month are streaking. They are not reliable due to the long-term projection. The results of 22-week forecast are:

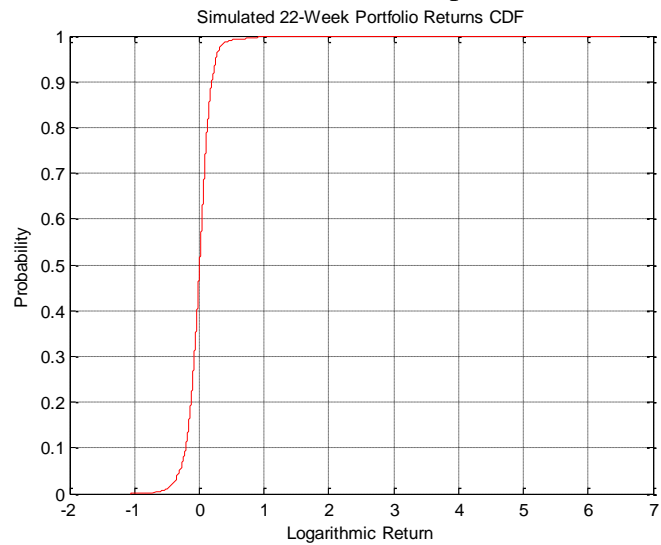
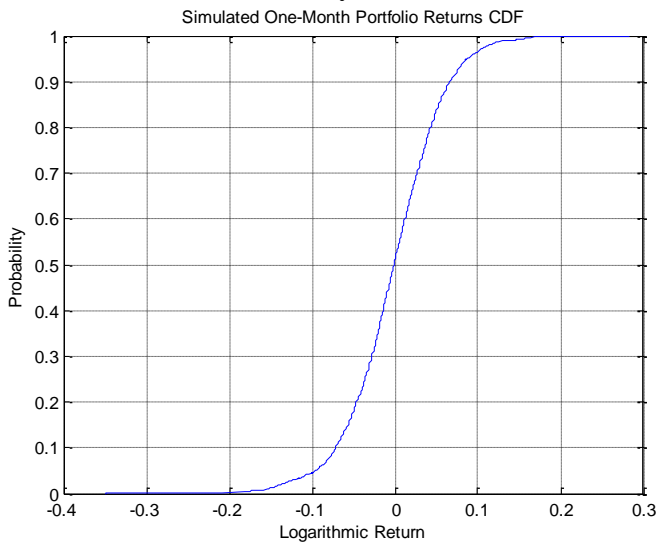
Maximum Simulated Loss: 103.0301%
Maximum Simulated Gain: 1595.0664%

Simulated 90% VaR: 22.1148%
Simulated 95% VaR: 30.5123%
Simulated 99% VaR: 49.9917%

Simulated 95% AVaR: 42.8157%

The picture in terms of maximum gain and maximum loss is exactly the opposite in comparison to the one-month forecast. But the VaR values are nearly 3 times larger. The reason again is hidden in the long-term period.

Geometrically the forecast for the cumulative returns is illustrated for both periods is below:



The blue graph represents the cumulative distribution function of the logarithmic returns under **one-month projection** while the red – **for 22-week projection**. The differences are obvious. The cumulative returns are in the interval $[-0.37; 0.26]$ for the one-month forecast while for the 22-week forecast the interval is $[-1.10; 3.66]$. Much larger interval is captured by the second forecast. Its cumulative distribution function is much steeper in comparison to the shorter forecast. The reason is again the period.

VAR Model

Vector Autoregressive models like the ones proposed by the Box-Jenkins approach require that the processes be stationary. The Augmented Dickey-Fuller Test shows that the returns which we generate by `price2ret` function are stationary. Therefore, we can use them directly into the VAR model. The only problem concerns the number of lags to be included. We can arbitrary choose 1 lag VAR model. We can also specify 2 lags or higher lags VAR model and compare their Information Criteria – the smaller the value, the better the model. It is important to note that the higher the number of lags the greater consumption of degrees of freedom and the harder the calculation in MatLab. Therefore, we start with 2 lag VAR with a constant term:

Model : 6-D VAR(2) with Additive Constant
 Conditional mean is AR-stable and is MA-invertible

a Constant:

-0.0010649
 -0.00105839
 -0.00105978
 0.00174001
 3.1864e-05
 -0.00182125

AR(1) Autoregression Matrix:

-0.625735	-0.0674629	0.902006	0.0076204	-0.0250737	0.0482043
1.07925	0.730725	-1.58118	-0.0129477	-0.0815619	0.0748889
-0.622028	0.00792832	0.823784	0.00389896	-0.034349	0.0536607
-1.66891	0.153063	1.76701	0.172774	-0.134669	0.165222
-0.534149	0.112979	0.404662	-0.0361658	0.316829	0.113084
-0.445532	0.221138	-0.236939	0.0500423	-0.0760349	0.614124

AR(2) Autoregression Matrix:

-2.64667	-0.471961	2.98506	0.0308236	-0.0577014	0.070461
-3.06618	-0.721279	3.60228	0.0158733	-0.00782926	0.0844332
-2.60712	-0.490605	2.95679	0.0284572	-0.0505454	0.0730272
-0.0801799	0.0543106	-0.303607	-0.13655	-0.098878	0.128575
-0.824608	-0.193504	0.87277	0.0594376	-0.191459	-0.0154254
-2.04344	-0.49275	2.49075	0.0373926	0.0935978	-0.214395

Q Innovations Covariance:

0.000932674	0.000772725	0.000904797	0.0007222	0.000691915	0.00103131
0.000772725	0.00136629	0.000867632	0.00097374	0.000850148	0.00103765
0.000904797	0.000867632	0.000897564	0.000760789	0.000716685	0.00102951
0.0007222	0.00097374	0.000760789	0.00667428	0.00106015	0.00108381
0.000691915	0.000850148	0.000716685	0.00106015	0.00116994	0.00103178
0.00103131	0.00103765	0.00102951	0.00108381	0.00103178	0.0023172

The estimation output includes only the coefficients and not the t-statistics corresponding to them. This is so because we do not emphasize on statistical significance rather than on forecasting. In VAR models the most significant part is played by the innovations (residuals) which are called **shocks**. The last part of the estimation output is actually their covariance structure. The numbers are rather small which implies little influence. The VAR model shows that the first index (HSBC) is negatively related to its own lags 1 and 2 and negatively related to all other indices' lags except for

lag 1 of Banco Santander. All constants are negligibly small. By using this matrix we can explain the mutual trends of the returns by using the lag structure of the VAR.

In order to check the adequacy of the model we must take the Information Criteria as a benchmark. LR ratio test is also a good solution to the problem. However, we should compare two VAR models with different number of lags – we compare our 2 lags model to a 5 lags model. The results are the following:

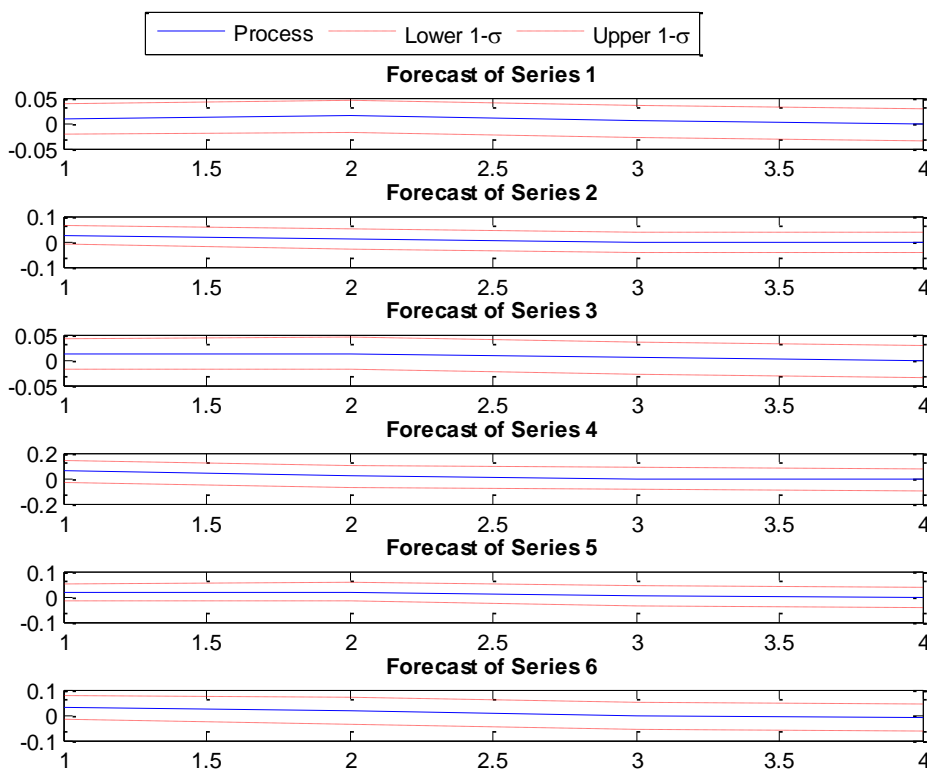
```
[n1,n1p] = vgxcount(EstSpec)
[n2,n2p] = vgxcount(EstSpec1)
aicbic([logL, logL1], [n1p, n2p])
ans =

    1.0e+04 *

    -1.7337    -1.7396
```

Therefore, the 2 lags model is a better solution to the problem since its Information Criteria is smaller to the other. As a consequence, we can stick to it.

Forecasting involves simulation as well. We generate 4 observation for each path (index) and they correspond to one-month projection. The forecast is illustrated below:



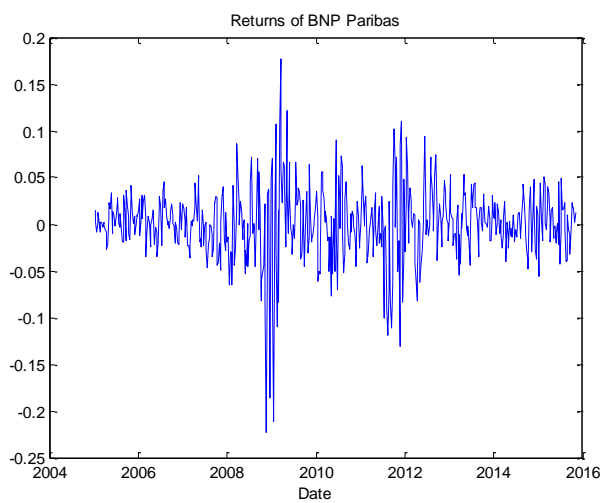
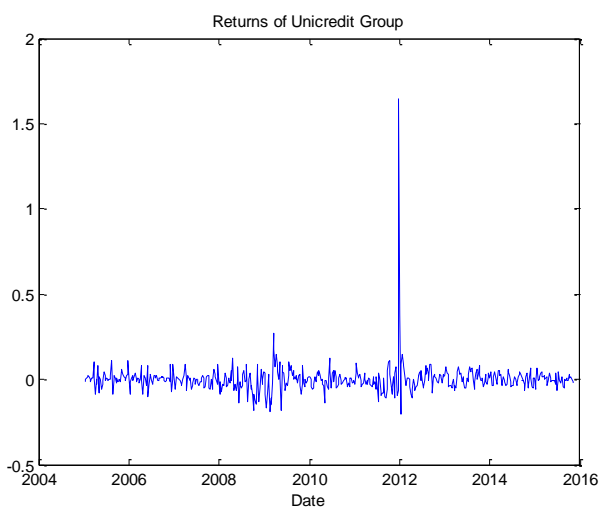
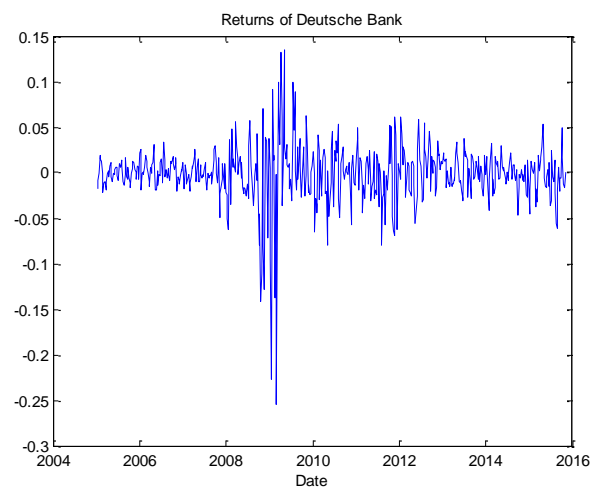
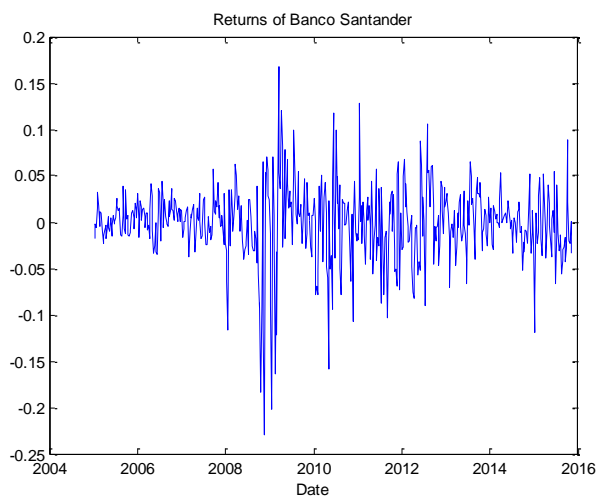
The graph shows us the same as the impulse response function. Here, we trace the unit-change in the standard deviation rather than the innovations themselves. Due to the short-term forecast the returns are relatively stable. Moreover, the six indices are not likely to deviate much from their usual path. Still there is a slight downturn of them all. The illustration can help us in prediction when the process will returns back to its starting point. Nonetheless, 4 periods are not enough.

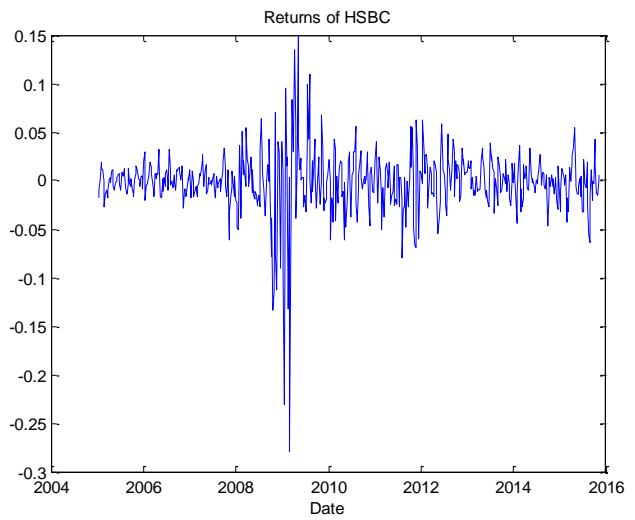
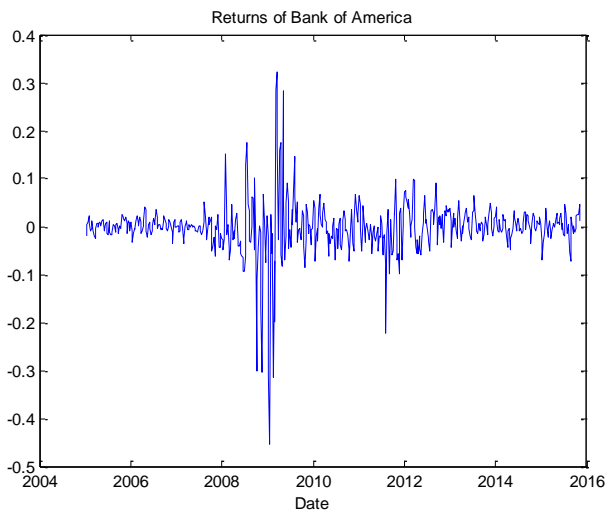
Appendix

<i>Variables</i>	HSBC	SAN	DB	UNI	BNP	BOA
HSBC	-	0.0048	0.8657	5.1108e-07	0.0039	0.0188
SAN	0.048	-	0.0020	0.0130	0.8657	0.2511
DB	0.8657	0.0020	-	9.8177e-07	0.0016	0.0157
UNI	5.1108e-07	0.0130	9.8177e-07	-	0.0020	0.0130
BNP	0.0039	0.8657	0.0016	0.0020	-	0.7299
BOA	0.0188	0.2511	0.0157	0.0130	0.7299	-

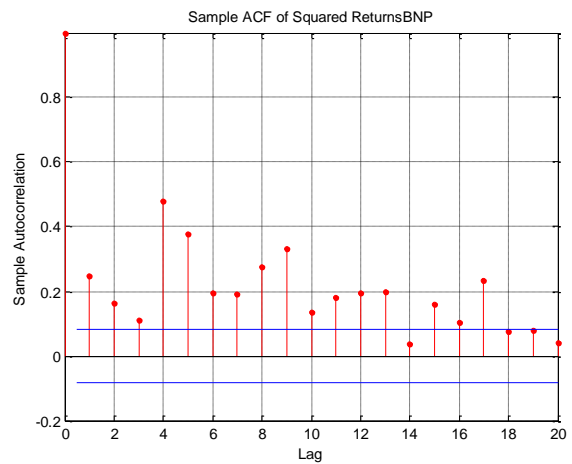
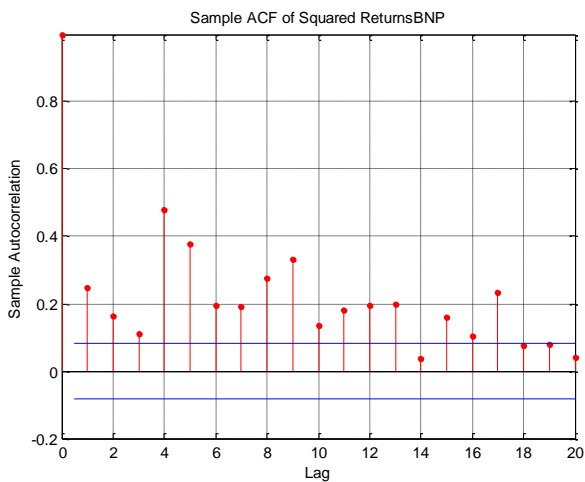
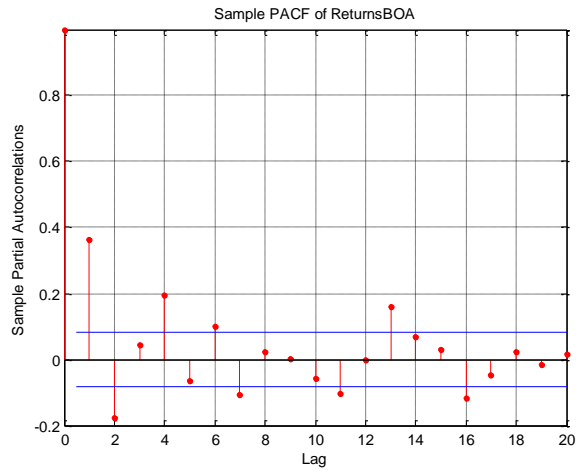
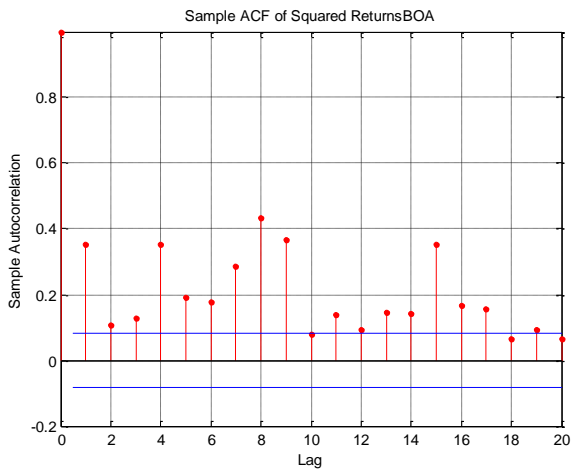
Table 1. Two sample K-S Tests results. p-values.

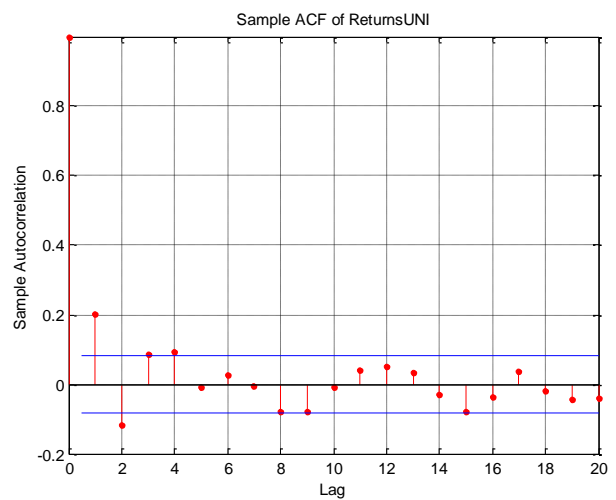
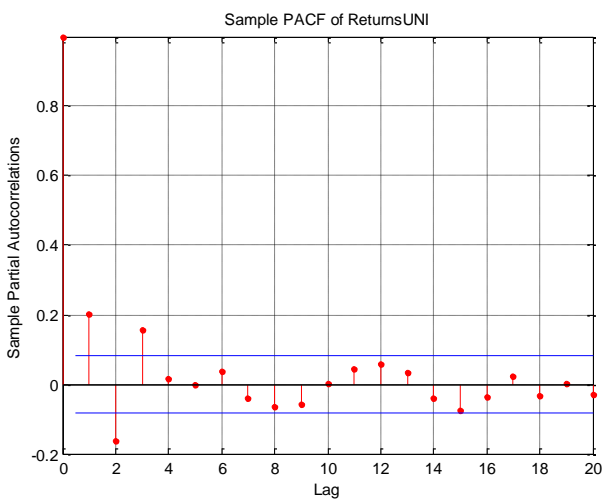
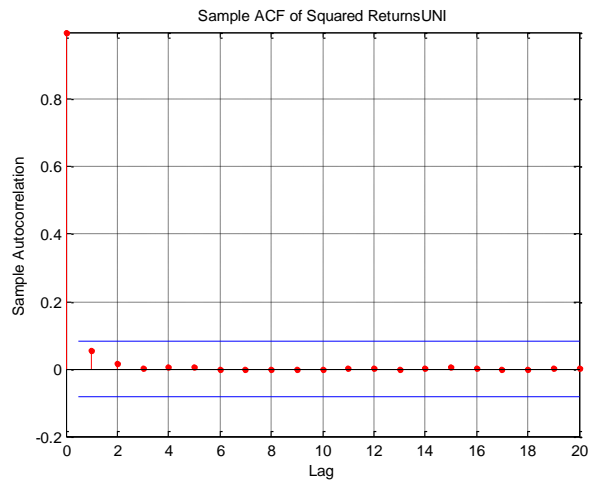
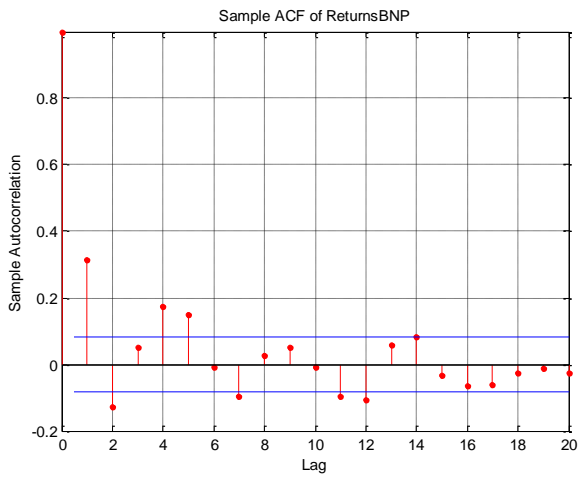
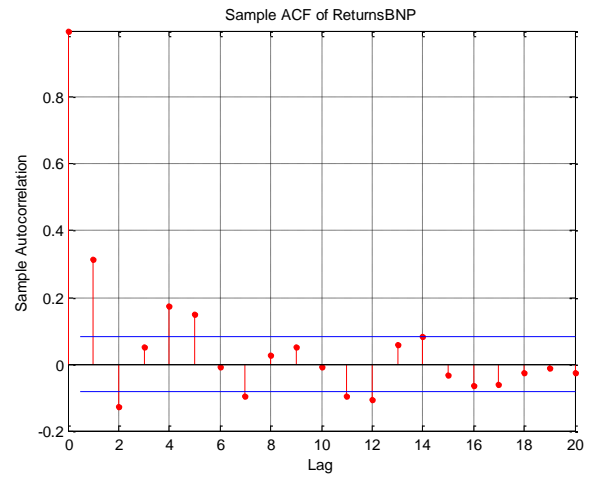
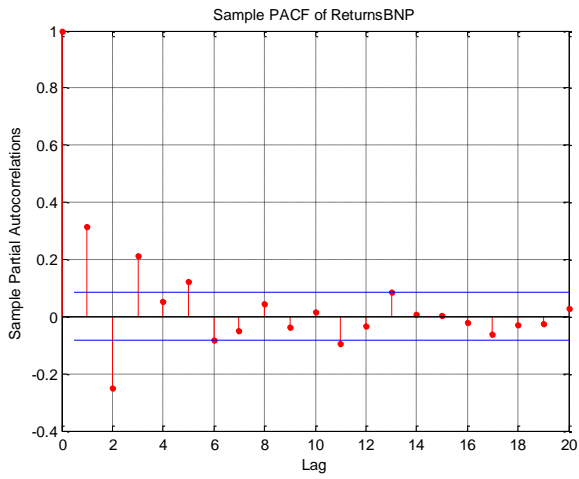
Plotting the processes for each index:

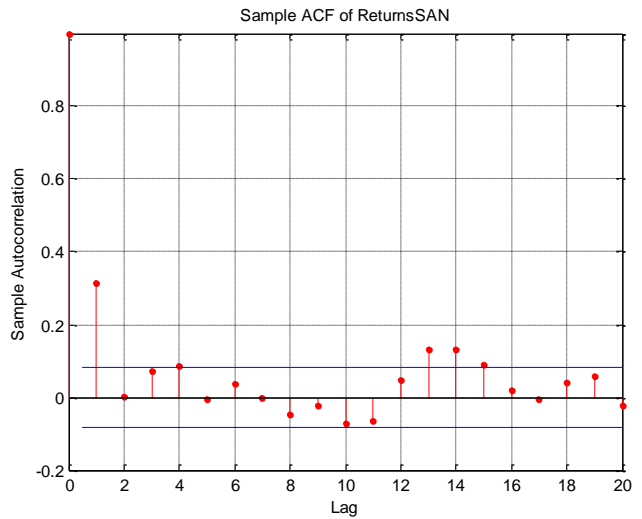
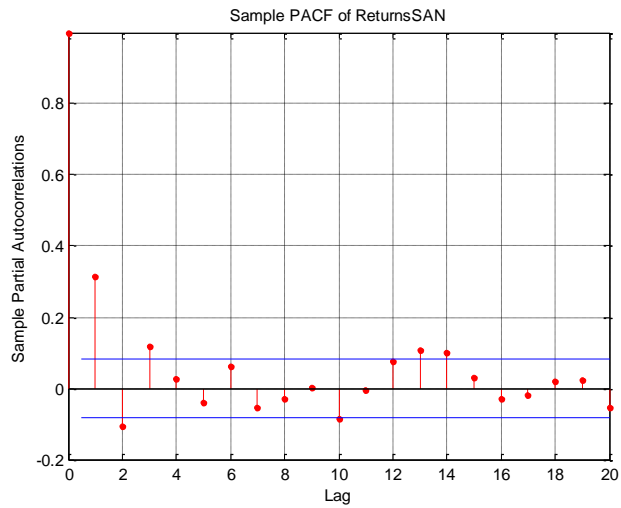
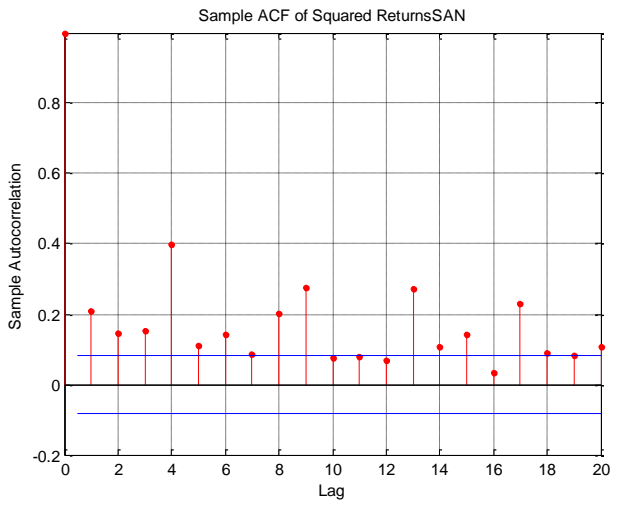
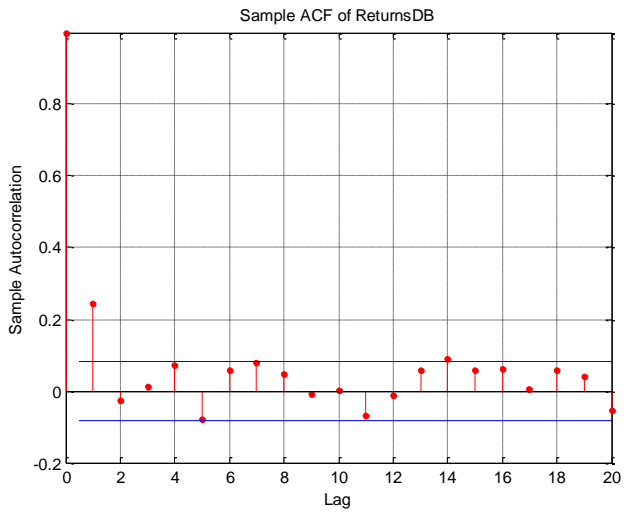
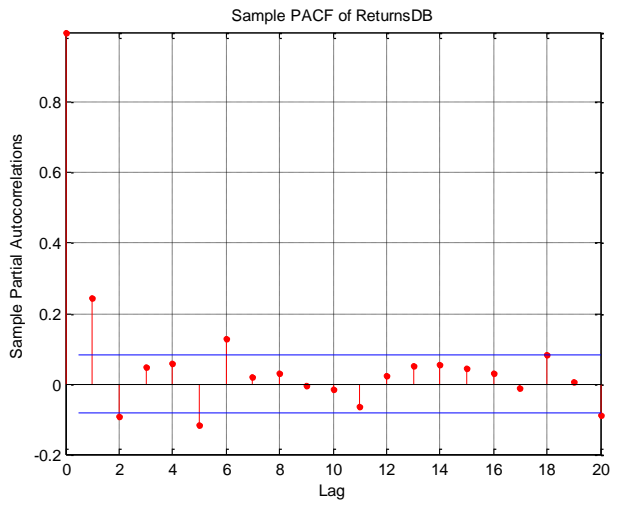
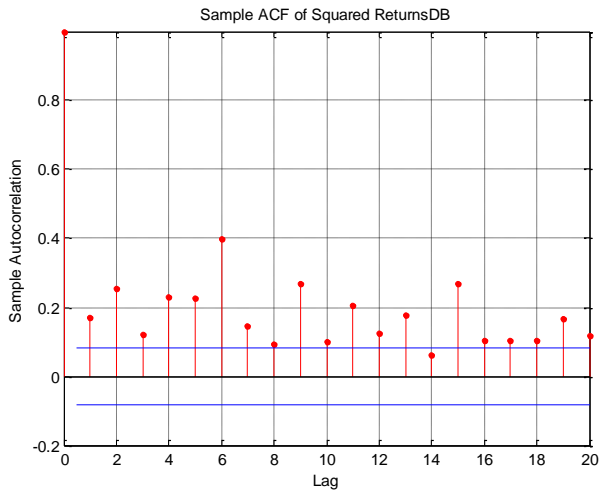


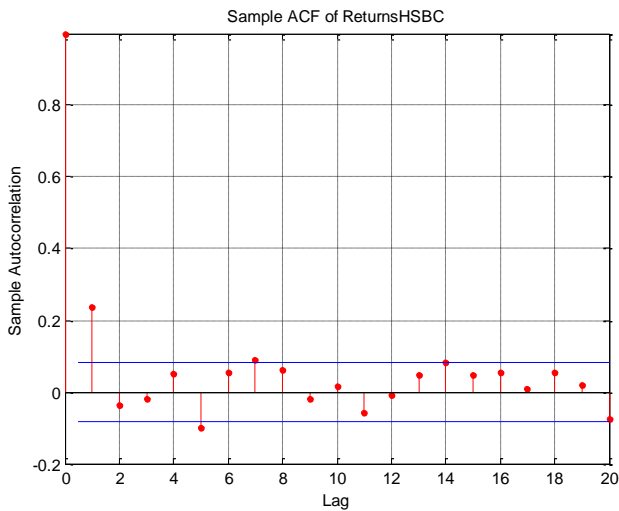
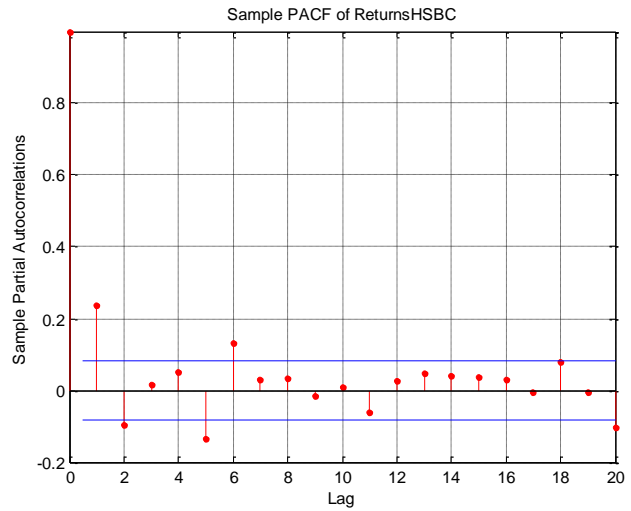
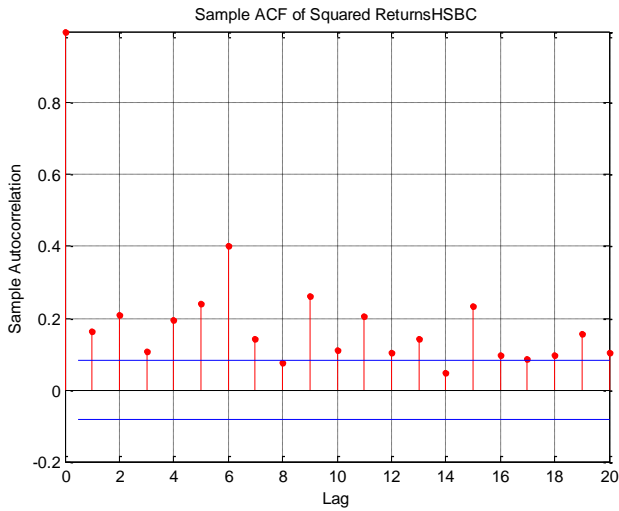


Plotting correlograms for the 6 indices:









Estimation Output for ARMA-GARCH:

i = 2

ARIMA(1,0,0) Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	0.000299509	0.00102096	0.293361
AR{1}	0.336453	0.0400553	8.39971
DoF	7.72527	2.98445	2.5885

GJR(1,1) Conditional Variance Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	2.9828e-05	9.78627e-06	3.04795
GARCH{1}	0.843356	0.0353842	23.8342
Leverage{1}	0.279752	0.0635906	4.39927
DoF	7.72527	2.98445	2.5885

i = 3

ARIMA(1,0,0) Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	-0.000336061	0.000730355	-0.460133
AR{1}	0.296687	0.0398286	7.4491
DoF	6.86777	2.40547	2.85506

GJR(1,1) Conditional Variance Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	1.52365e-05	5.33726e-06	2.85475
GARCH{1}	0.834405	0.0332344	25.1066
ARCH{1}	0.0487533	0.0466352	1.04542
Leverage{1}	0.198109	0.0695652	2.84782
DoF	6.86777	2.40547	2.85506

i = 4

ARIMA(1,0,0) Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	-0.000542832	0.00144921	-0.374571
AR{1}	0.328996	0.0446456	7.36907
DoF	3.50189	0.441503	7.93176

GJR(1,1) Conditional Variance Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	0.000570515	0.000168736	3.38111
GARCH{1}	0.440383	0.0828612	5.3147
ARCH{1}	0.426995	0.160361	2.66271
Leverage{1}	0.13221	0.210999	0.626591
DoF	3.50189	0.441503	7.93176

i = 5

ARIMA(1,0,0) Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	0.000230938	0.00105017	0.219904
AR{1}	0.351894	0.0428995	8.20276
DoF	11.5989	5.45041	2.12808

GJR(1,1) Conditional Variance Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	3.29083e-05	1.20365e-05	2.73404
GARCH{1}	0.844849	0.0320757	26.3393
ARCH{1}	0.0101179	0.0300225	0.337012
Leverage{1}	0.236835	0.0565821	4.18569
DoF	11.5989	5.45041	2.12808

i = 6

ARIMA(1,0,0) Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	-5.30295e-05	0.0008647	-0.061327
AR{1}	0.420048	0.0384343	10.929
DoF	5.16593	1.19434	4.32536

GJR(1,1) Conditional Variance Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	1.8813e-05	8.12909e-06	2.31428
GARCH{1}	0.830919	0.0325759	25.5071
ARCH{1}	0.0781019	0.0469026	1.66519
Leverage{1}	0.181957	0.0712804	2.5527
DoF	5.16593	1.19434	4.32536

ARMA-GARCH GJR(2,2):

i = 1

ARIMA(1,0,0) Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	-0.000269506	0.000733355	-0.367496
AR{1}	0.301916	0.0383525	7.87215
DoF	5.72086	1.78855	3.19861

GJR(2,2) Conditional Variance Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	3.15146e-05	1.2387e-05	2.54417
GARCH{1}	0.141681	0.305591	0.463629
GARCH{2}	0.561725	0.266392	2.10864
ARCH{1}	0.00981973	0.0678002	0.144833
ARCH{2}	0.103285	0.0762773	1.35408
Leverage{1}	0.153067	0.103601	1.47746
Leverage{2}	0.140914	0.116148	1.21323
DoF	5.72086	1.78855	3.19861

i = 2

ARIMA(1,0,0) Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	0.000266482	0.00102655	0.259589
AR{1}	0.340827	0.0408004	8.35353
DoF	7.75634	3.01111	2.57591

GJR(2,2) Conditional Variance Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	5.4829e-05	2.1086e-05	2.60026
GARCH{2}	0.697889	0.326654	2.13648
ARCH{1}	0.0189579	0.0525032	0.361081
Leverage{1}	0.24366	0.0944406	2.58003
Leverage{2}	0.267201	0.100954	2.64676
DoF	7.75634	3.01111	2.57591

i = 3

ARIMA(1,0,0) Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	-0.000326116	0.000731772	-0.445652
AR{1}	0.294046	0.0409284	7.18439
DoF	6.94142	2.50711	2.7687

GJR(2,1) Conditional Variance Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	1.74905e-05	3.37703e-05	0.517925
GARCH{1}	0.635056	2.27069	0.279675
GARCH{2}	0.168878	1.88278	0.0896958
ARCH{1}	0.0619717	0.07707	0.804096
Leverage{1}	0.229111	0.132838	1.72474
DoF	6.94142	2.50711	2.7687

i = 4

ARIMA(1,0,0) Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	-0.000740276	0.0014423	-0.51326
AR{1}	0.31964	0.0426919	7.48713
DoF	3.64715	0.508238	7.17607

GJR(1,2) Conditional Variance Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	0.000717924	0.000240848	2.98081
GARCH{1}	0.246924	0.305516	0.80822
ARCH{1}	0.334696	0.147695	2.26612
ARCH{2}	0.297259	0.204564	1.45313
Leverage{1}	-0.0297199	0.19855	-0.149685
Leverage{2}	0.102197	0.218927	0.46681
DoF	3.64715	0.508238	7.17607

i = 5

ARIMA(1,0,0) Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	8.60117e-05	0.00102826	0.083648
AR{1}	0.357514	0.0403252	8.86578
DoF	11.225	5.09855	2.2016

GJR(1,2) Conditional Variance Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	3.92426e-05	2.54181e-05	1.54388
GARCH{1}	0.811171	0.613657	1.32186
ARCH{1}	0.00684409	0.0540217	0.126691
ARCH{2}	0.0128294	0.0561049	0.228668
Leverage{1}	0.0683123	0.0870947	0.784345
Leverage{2}	0.212878	0.189214	1.12507
DoF	11.225	5.09855	2.2016

i = 6

ARIMA(1,0,0) Model:

 Conditional Probability Distribution: t

Parameter	Value	Standard Error	t Statistic
Constant	0.000248394	0.000880114	0.282229
AR{1}	0.429857	0.0389968	11.0229
DoF	5.42928	1.26875	4.27926

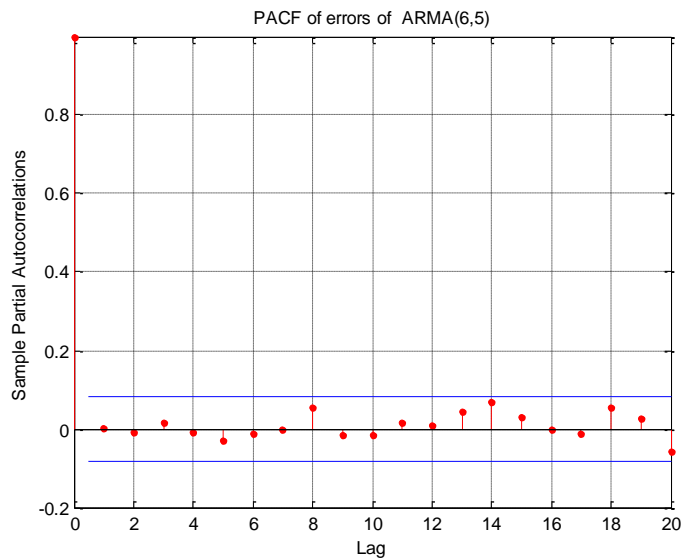
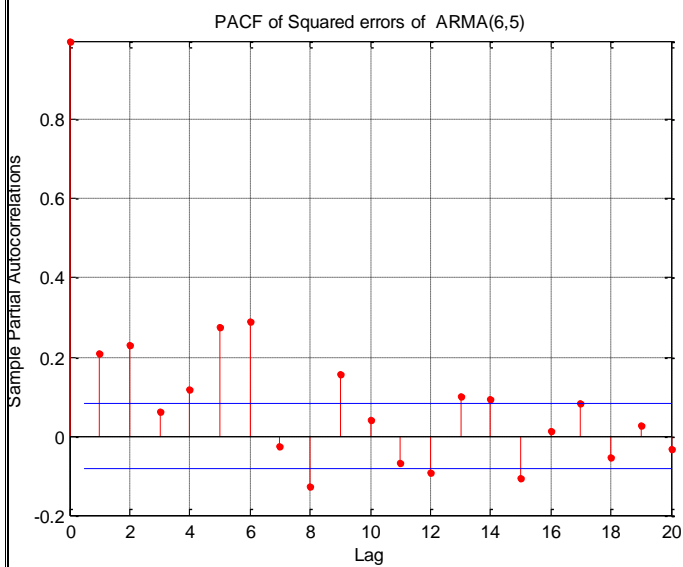
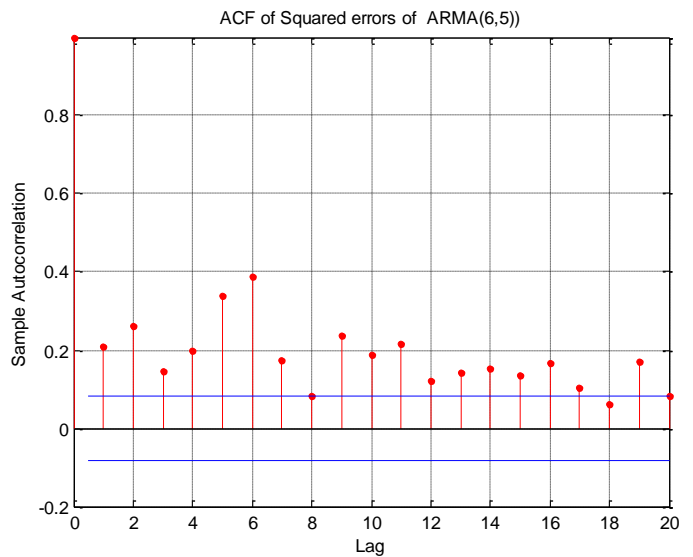
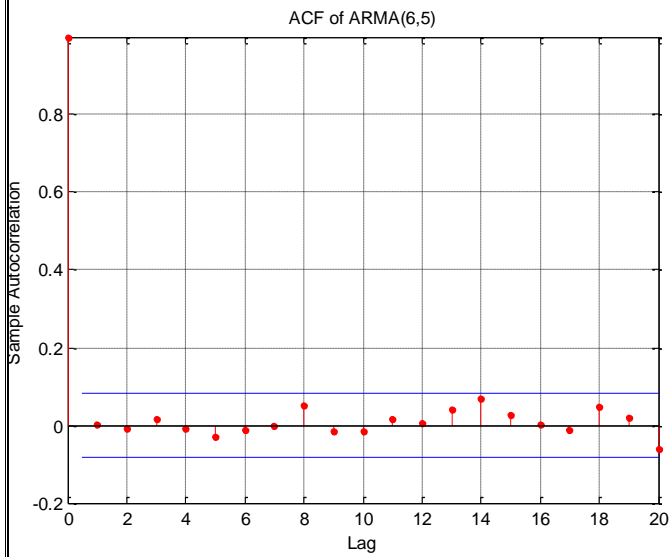
GJR(2,2) Conditional Variance Model:

 Conditional Probability Distribution: t

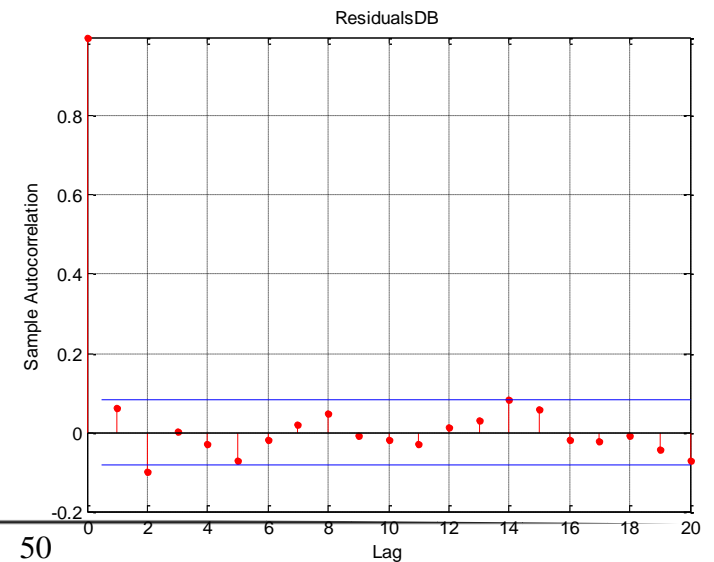
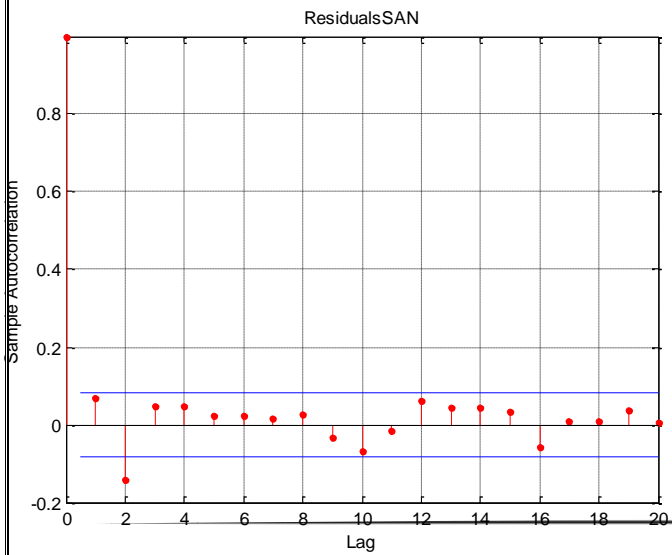
Parameter	Value	Standard Error	t Statistic
Constant	2.94651e-05	1.38462e-05	2.12802
GARCH{1}	0.150524	0.195337	0.77059
GARCH{2}	0.564597	0.161641	3.49291
ARCH{1}	0.0370351	0.0578995	0.639645
ARCH{2}	0.100984	0.0767991	1.31491
Leverage{1}	0.394702	0.130543	3.02354
Leverage{2}	-0.100984	0.142253	-0.709893
DoF	5.42928	1.26875	4.27926

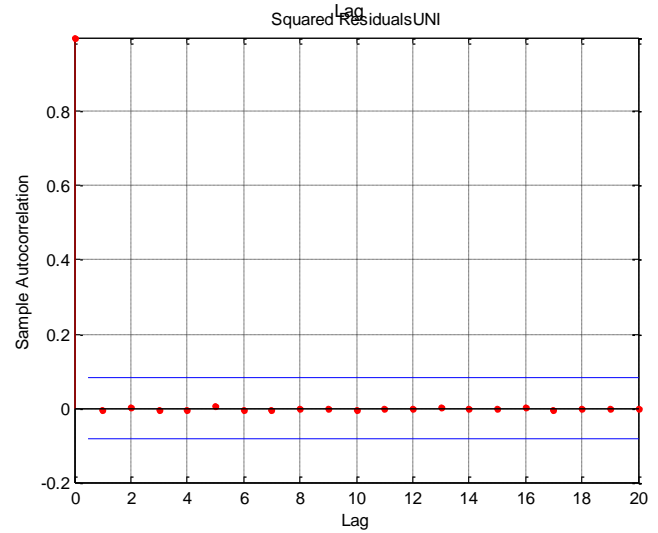
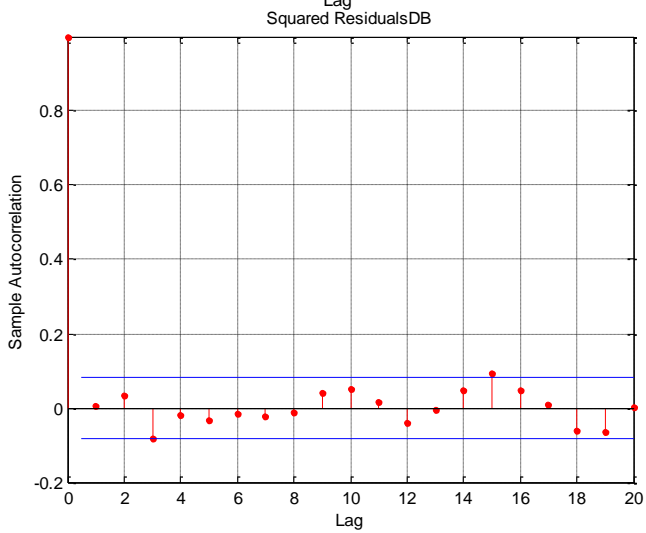
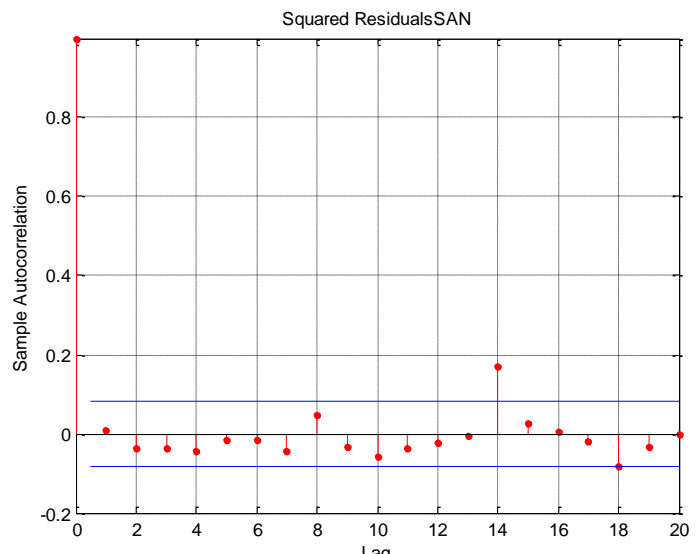
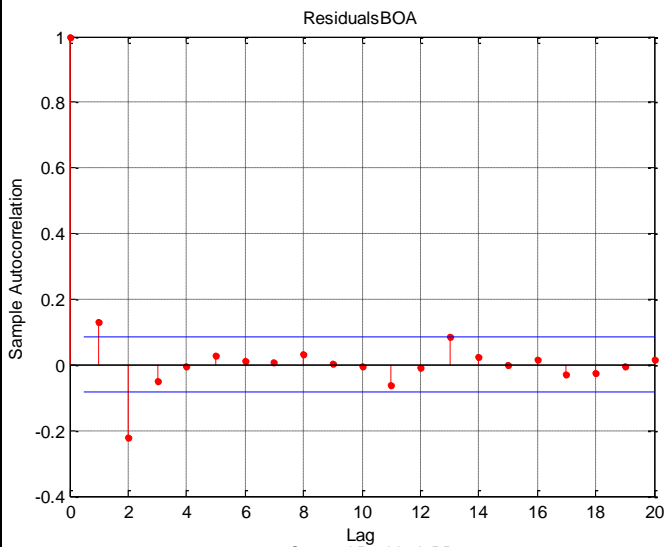
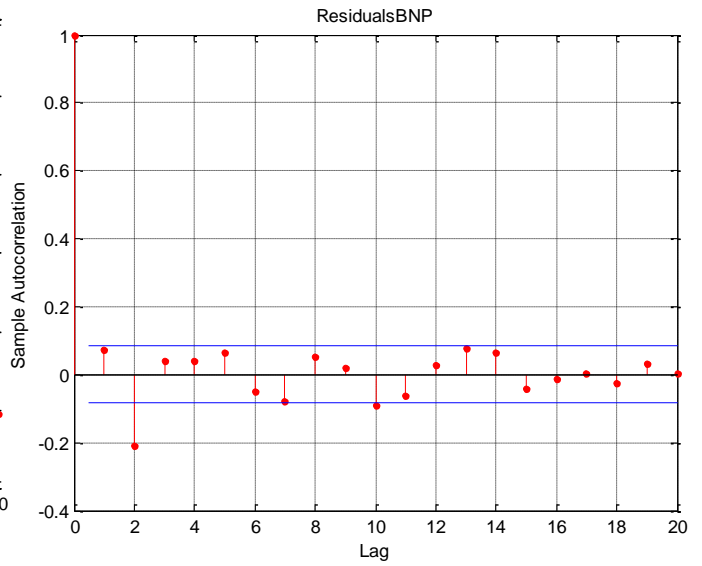
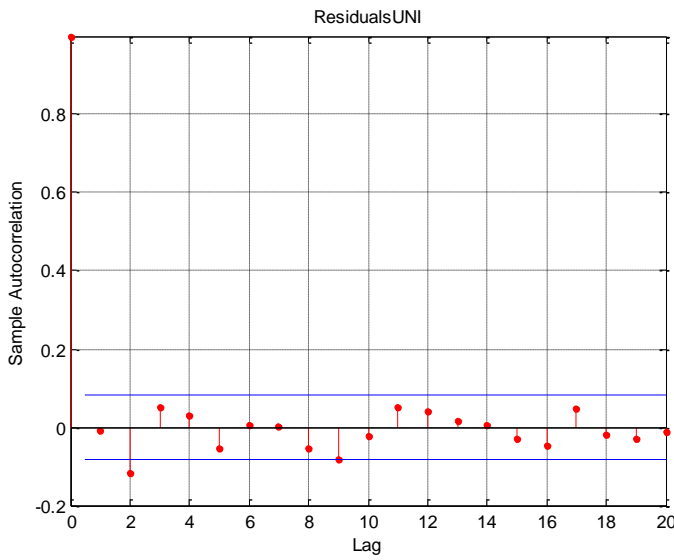
Testing the Residuals:

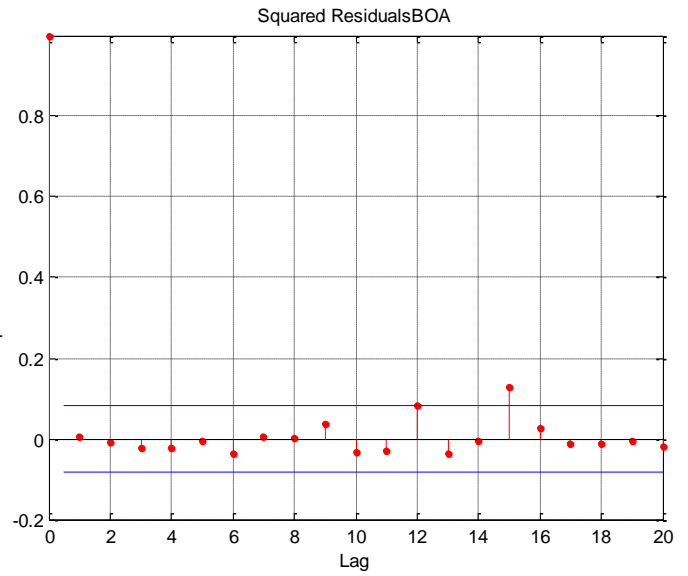
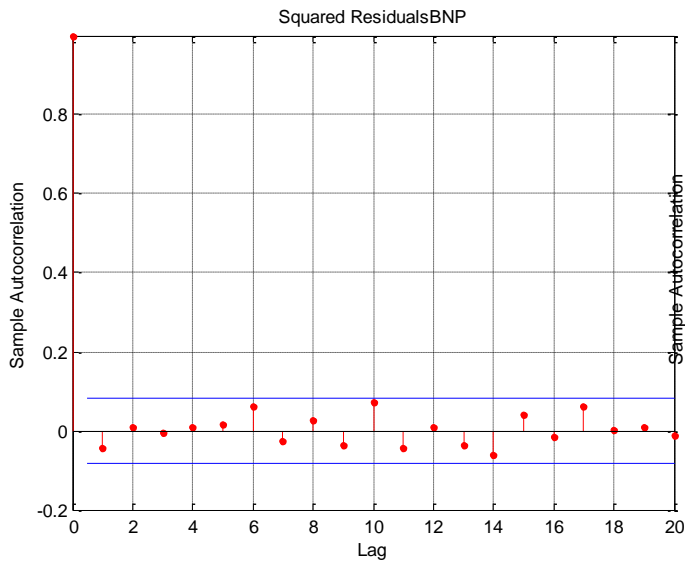
ARMA (6,5)



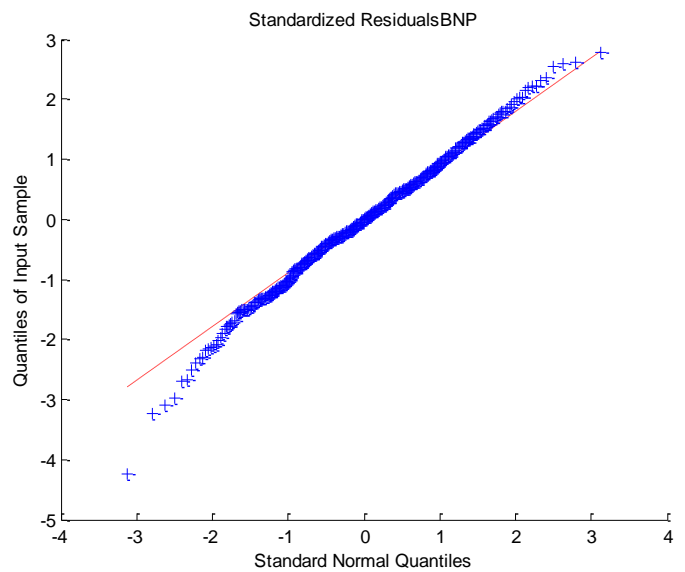
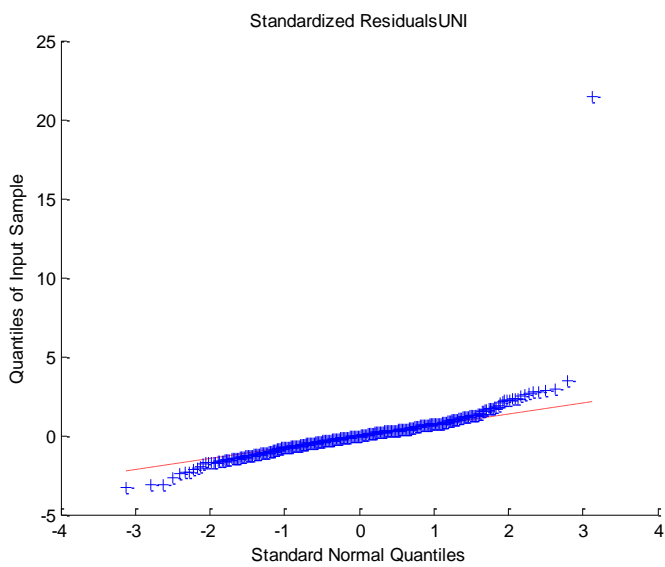
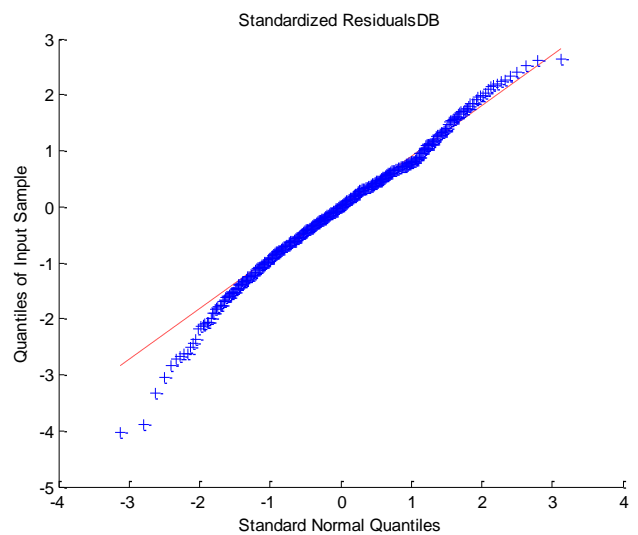
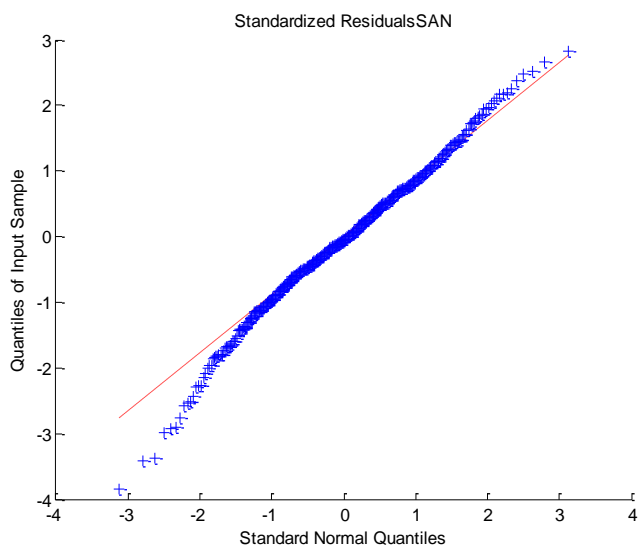
ARMA-GARCH:

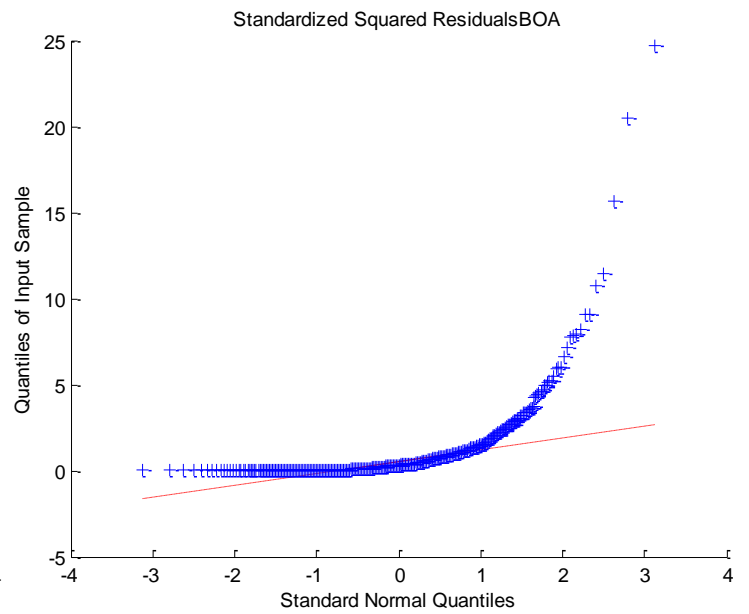
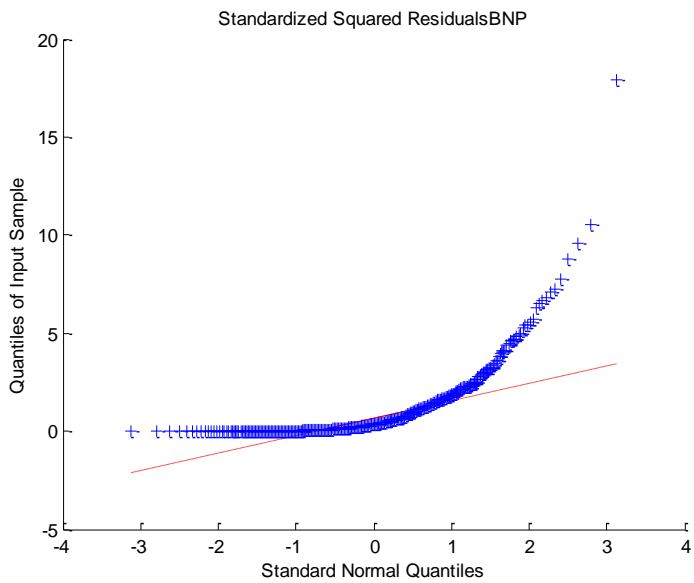
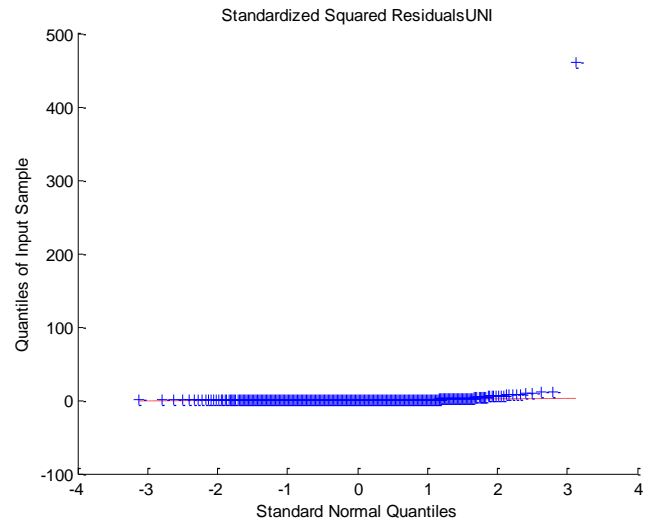
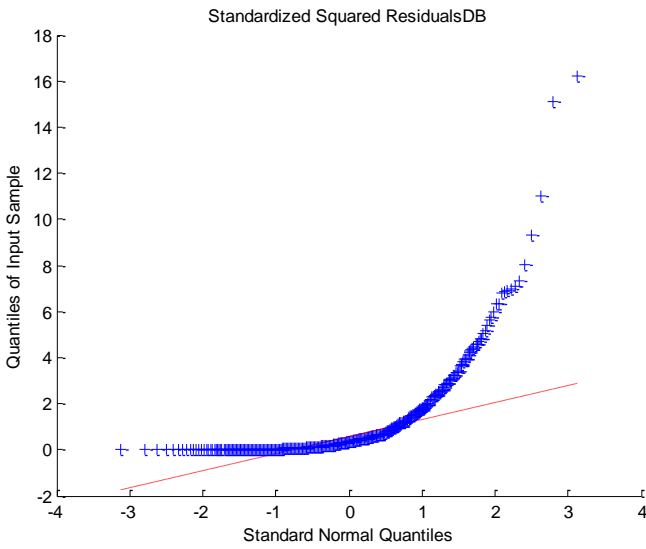
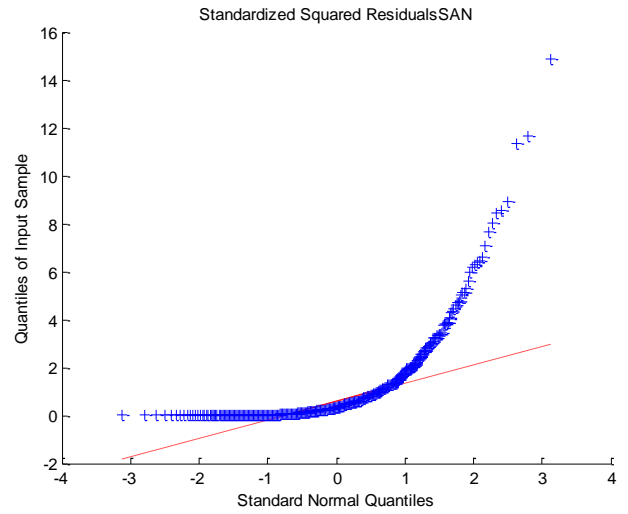
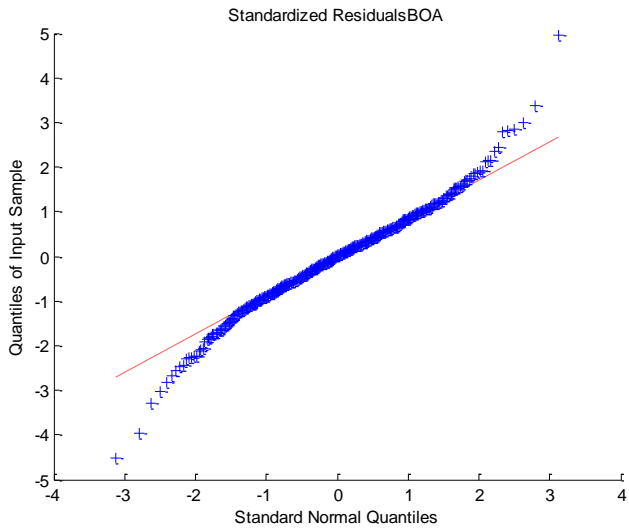




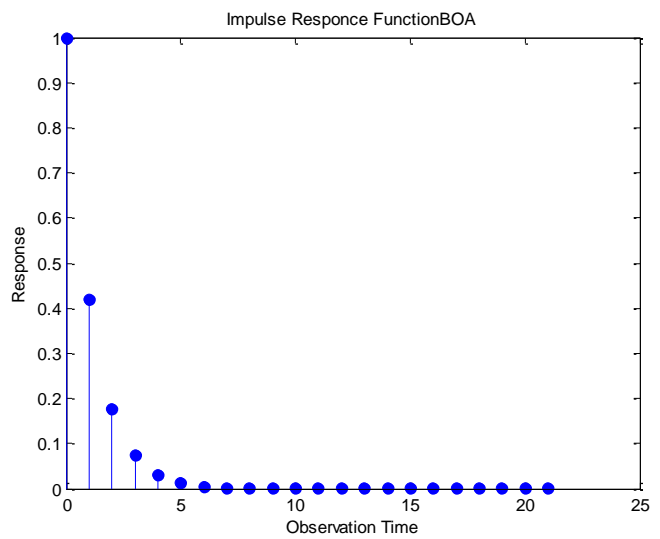
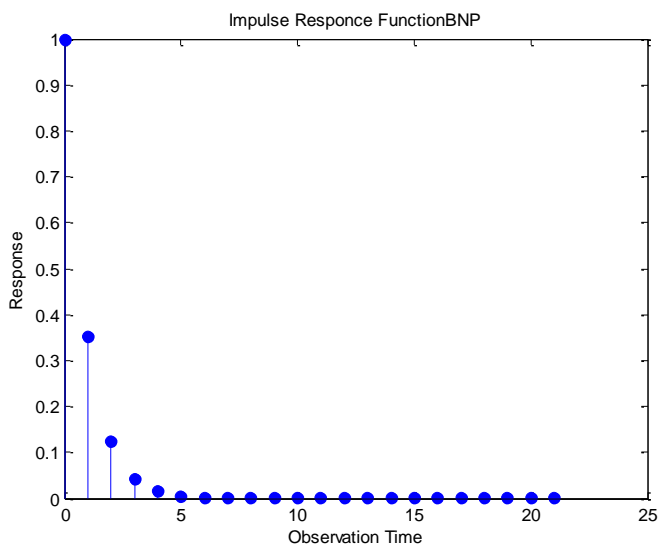
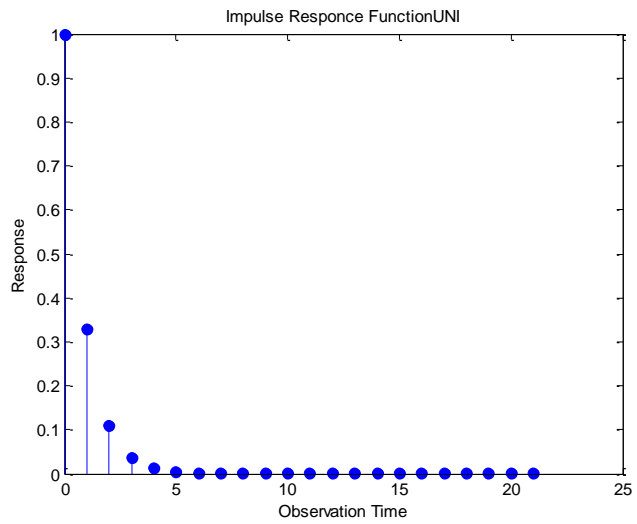
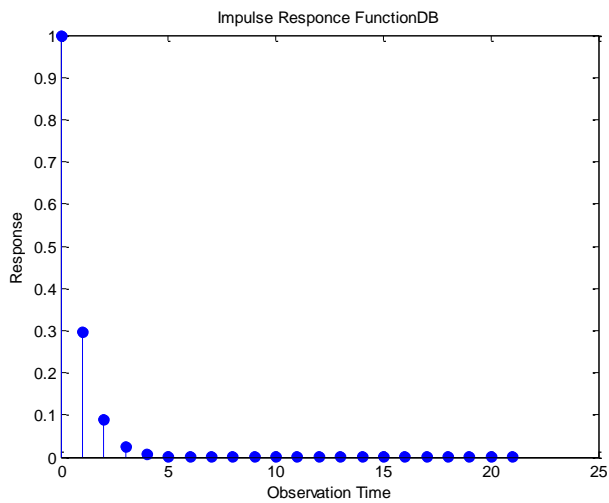
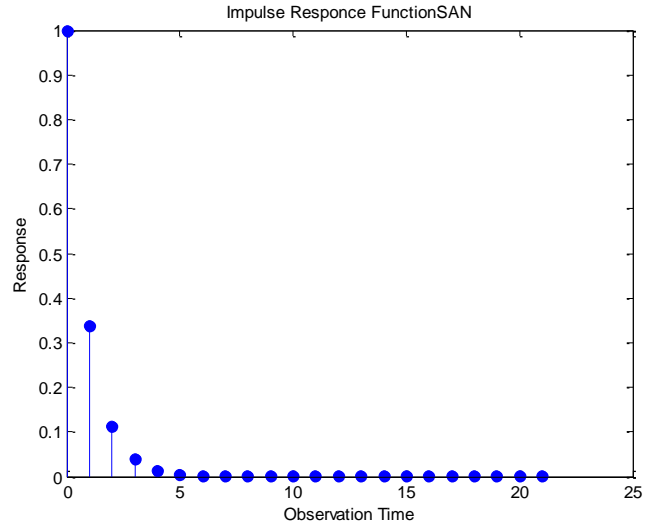
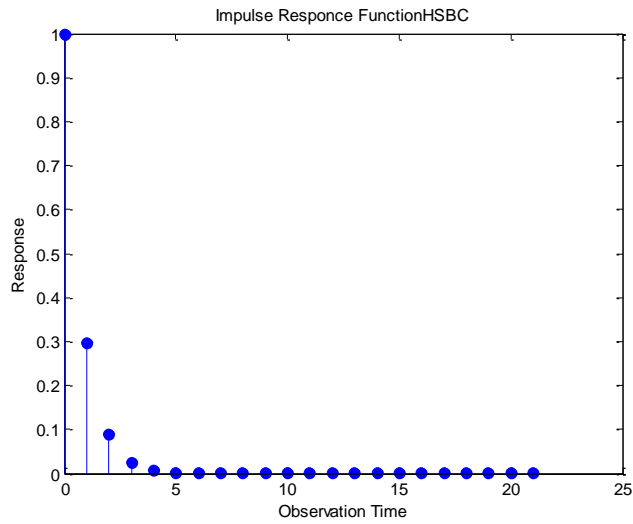


QQ plot of the Standardized Residuals and Squared Standardized Residuals:

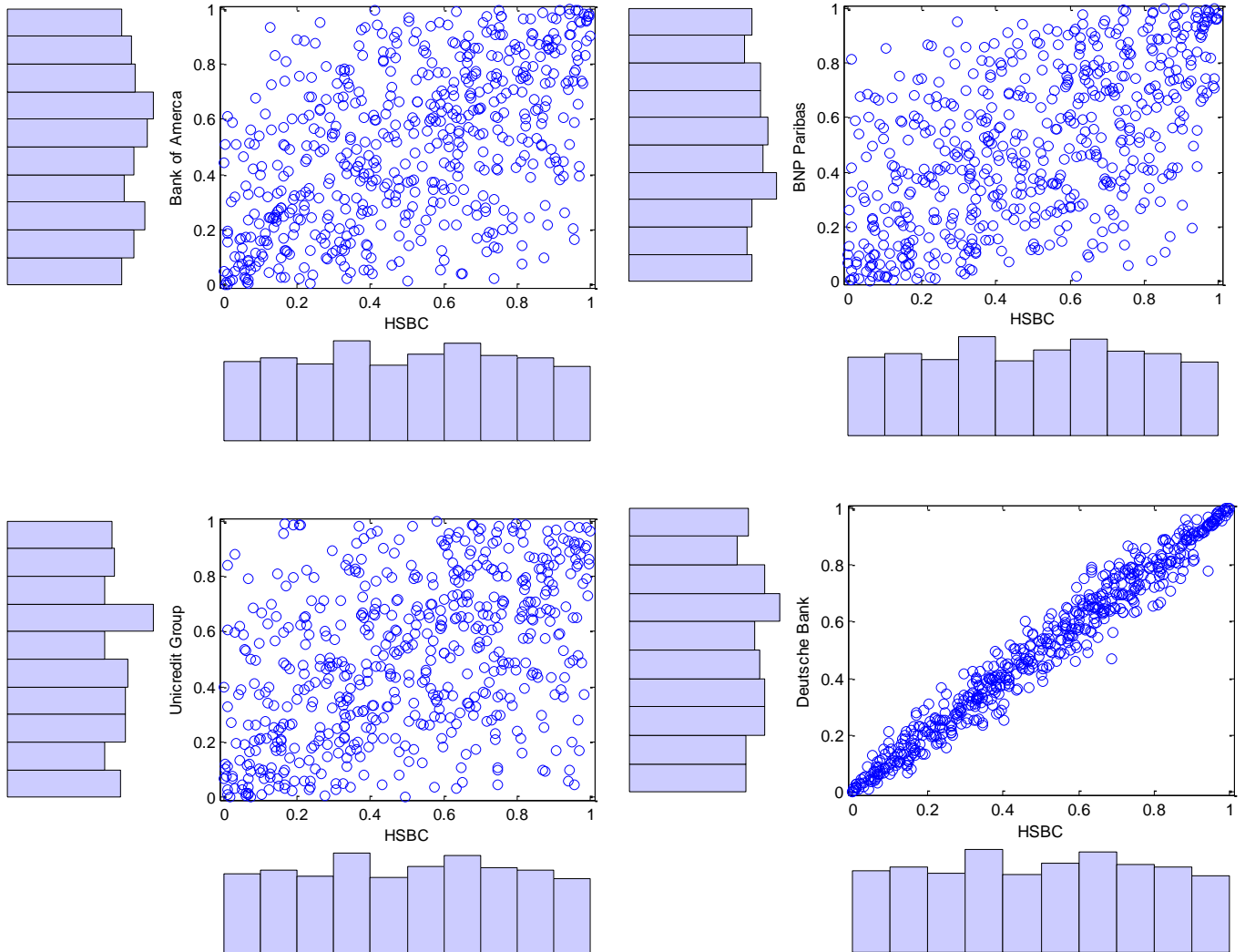




Impulse Response Function for each index:



Copula dependence illustration:



Bibliography

Rachev Sv., Mann, Ch., Fabozzi Fr. (2005) *Fat-Tailed and Skewed Asset Return Distributions*; John Wiley & Sons, Inc.;

Fabozzi Fr., Stoyanov St., Rachev, Sv. (2008) *Advanced Stochastic Model, Risk Assessment and Portfolio Optimization*;

M. Gili, Këllezi Ev. (2006) *An Application of Extreme Value Theory for Measuring Financial Risk*;

Gujarati D. (2004) *Basic Econometrics*, McGraw-Hill Company;

Maddala G.S, (1992) *Introduction to Econometrics*, 2nd edition, Macmillan Publishing Company;

Hamilton J. (1994) *Time Series Analysis*, Princeton University

MatLab Documentation:

<http://www.mathworks.com/help/econ/examples/using-extreme-value-theory-and-copulas-to-evaluate-market-risk.html>

<http://www.mathworks.com/help/econ/introduction-to-vector-autoregressive-var-models.html>