UKRAINIAN CATHOLIC UNIVERSITY
APPLIED SCIENCE FACULTY
INTRODUCTION TO FINANCES

KOSTYANTYN HRYTSYUK
KHRYSTYNA KUBATSKA

APPLICATION OF GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY (GARCH) MODELS FOR ANALYZING INVESTMENT FUNDS VOLATILITY

LVIV
2020

1 Abstract

The paper explores the volatility predicting model for stock returns. Data collected for six investment companies, which are some of the largest worlds. 24 GARCH and GJR models were made and analyzed to predict the volatility of the stock returns. The conclusion was made using the results of the analysis, comparing the real volatility and predicted.

The analysis was performed using R software.
All visualizations, tables, and R-code are attached.

2 Introduction

Return is the main characteristic of any financial asset. Its volatility that describes the spread plays an essential role in many financial applications. Its main use is to assess the value of market risk. Volatility is crucial for financial institutions not only to know the current meaning of variability managed assets but to assess their future value. Accurate forecasting of financial indicators is complicated and usually non-intuitive.

Stock market volatility is usually associated with investment risk. Volatility is usually measured as standard deviation, which indicates how spread out the price of a stock is around the mean or moving average.

So, the purpose of the research is to predict the volatility of stock prices of the largest investment companies in the world using the GARCH model.

3 Data description

For the analysis, we used stock returns for six large investment companies:

- Vanguard Total Stock Market Index Fund
- Blackrock
- State Street Corporation
- JPMorgan Chase & Co.
- The Bank of New York Mellon Corporation
- Allianz SE

All data are daily for the period between 2001-06-18 and 2020-05-08.
Prices for stocks were gotten from the one source **yahoo!finance** and converted in returns.

Summary statistics for the stock returns and their volatility are shown in the table in the Appendix.

From Table 1, we can see that the JPMorgan company had the highest mean of stock returns, and Allianz SE had the lowest one. The lowest volatility had Vanguard, while the highest was in State Street.

On the plots for returns per share and their volatility, we can see that returns of all funds had higher volatility during the next periods:
- 2000-03 (Dotcom bubble)
- 2008-09 (World Economy Crisis)
- 2011-12 (We assume it is consequences of World Economy Crisis)
- 2020 (Crisis caused coronavirus pandemic)

On the histograms (see Appendix), we can see that the returns of all funds are closer to skewed student t distributions (with gamma = 1.2) than to normal ones.

4 Methodology explanation

4.1 Theoretical concepts

In our research, we used the Generalized AutoRegressive Conditional Heteroskedasticity model (**GARCH**).

It is a statistical model used in analyzing time-series data where the variance error is serially autocorrelated. GARCH model assumes the variance of the error term follows an autoregressive moving average process. That is, the error term is heteroskedastic.

The goal of GARCH models is to provide a volatility measure that can be used in financial decisions like risk analysis, portfolio selection, and derivative pricing.

In financial markets analysis, it is important to predict how stocks will behave in the future. Particularly, what value of their returns will be. We will denote return as follows

$$R_{t+1} = \frac{P_{t+1} - P_t}{P_t}$$

In appended graphs of the daily return, we can see that for each company return series has 0 mean and volatility σ_t. For modeling returns, we can compute the mean value $\boldsymbol{\mu}$ and use it as a prediction for future returns. But with such an approach, we will receive normally distributed series of errors $\{e_t\}$ of prediction with the value of the difference between $\boldsymbol{\mu}$ and actual return:

$$e_t = R_t - \mu, \quad e_t \sim N(0, \sigma_t^2)$$

Formula
A time series $\{e_t\}$ is given at each case by:

$$e_t = \sigma_t \omega_t$$

Where $\{\omega_t\}$ is discrete white noise signal, with zero mean and unit variance, and is given by:

$$\sigma_t^2 = \omega_0 + \sum_{i=1}^{q} \alpha_i e_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

Where α_i and β_j are parameters of the model. There is a number of restrictions for parameters:

- $\omega, \alpha, \beta > 0$. It is required by the fact that variance has only non-negative values;
- $\alpha + \beta < 1$. It is required by the fact that sample variance always tends to the population variance in the long run.
- $\sigma^2 = \frac{\omega}{1-\alpha-\beta} \longrightarrow \omega = \sigma^2 * (1 - \alpha - \beta)$. ω parameter depends on population σ^2.

$\{\sigma_t^2\}$ is a generalised autoregressive conditional heteroskedastic model of order p,q, denoted by GARCH(p,q).

4.2 Program realization of GARCH model

In our project R package `rugarch` was used for creating GARCH models.
We can generalize three main steps in GARCH modeling:

1. Specification GARCH settings such as a model for mean, variance and errors distribution by the function `ugarchspec`;
2. Evaluating GARCH model by the function; `ugarchfit`
3. Predicting future volatility with fitted model by the function `ugarchforecast`.

4.2.1 Mean models

AR

Autoregressive models(AR) predict future values based on past values. In the AR(1) autoregressive process, the current value is based on the immediately preceding value, while in the AR(2) current value is based on the previous two values.

The AR(p) model is defined as

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \epsilon_t$$

where c is a constant, $\varphi_i, \ldots, \varphi_p$ are the parameters of the model, and ϵ_t is error term.

4.2.2 Variance models

Standard GARCH(1,1)

The most common realization of the GARCH model is (1,1), which means that we are using auto-regressive processes of order 1 to estimate α and β coefficients:

$$\sigma_t^2 = \omega_0 + \alpha\ e_{t-1}^2 + \beta\ \sigma_{t-1}^2$$

Glosten, Jagannathan and Runkle GARCH

There is an effect by the sign of the error called "leverage effect". For example:
The company stock price had a big fall. Thus, the debt-to-market ratio decreased, the company became riskier and the volatility also increased.

Therefore, we need two different equations to handle both positive and negative value of a previous return or, better said, a previous error. When value of e_{t-1} is positive we are using the standard GARCH(1,1) equation. When this value is negative, we have to increase the multiplier of α with constant $\gamma \geq 0$:

$$\sigma_t^2 = \begin{cases} \omega_0 + \alpha\ e_{t-1}^2 + \beta\ \sigma_{t-1}^2 & , e_{t-1} > 0 \\ \omega_0 + (\alpha + \gamma)\ e_{t-1}^2 + \beta\ \sigma_{t-1}^2 & , e_{t-1} \leq 0 \end{cases} \tag{1}$$

4.2.3 Distribution models

Normal distribution A normal (or Gaussian or Gauss or Laplace–Gauss) distribution has a bell-shaped density curve described by its mean and standard deviation. The density curve is symmetrical and centered about its mean with its spread determined by its standard deviation.

The general form of its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where μ is the mean and σ it the standard deviation. [5]

Skewed Student t Distribution Skewed Student t Distribution belongs to family of continuous probability distributions. The general form of its probability density function is

$$f_{SGT}(x; \mu, \sigma, \lambda, p, q) = \frac{p}{2v\sigma q^{1/p} B(\frac{1}{p}, q) \left(\frac{|x-\mu+m|^p}{q(v\sigma)^p(\lambda sign(x-\mu+m)+1)^p} + 1 \right)^{\frac{1}{p}+q}}$$

Where

- B is the beta function
- μ is the location parameter
- $\sigma > 0$ is the scale parameter
- $-1 < \lambda < 1$ is the skewness parameter
- $p > 0$ and $q > 0$ are the parameters that control the kurtosis [4]

Plots, showing changing the distributions depending on the parameters, are attached in the Appendix.

5 Modeling

In this chapter, we will present outcomes from our modeling. For each subsection we will use data for the 6 selected funds.

In our research, we made four models for every fund: sGARCH and GJR models for normal and student t distributions each.

The difference between sGARCH and GJR was described in chapter 4, so here we will explain difference in choice of error distribution models.

From the plots of returns distribution, we can see that this distribution tends to the normal one. Also, we can see that the distribution of returns is more peaked than the normal one and often is skewed a little. To handle such features of the return distribution we can use the skewed Student t distribution.

With a parameter of degrees of freedom (in GARCH modeling we will call it *shape*), a model can handle peak around mean value. With a parameter of skewness λ (in GARCH modeling we will call it *skew*), a model can handle a bigger amount of positive or negative returns.

The effectiveness of these two distributions of errors in GARCH modeling will be discussed in the next subsections.

5.1 Model comparing

5.1.1 Total error sum

The total error sum (TES) is calculated as the sum of the difference between predicted and real volatility. The lower is TES, the higher is fitting of a model.

For stock returns of Vanguard Total Stock Market Index Fund, Blackrock, and JPMorgan Chase & Co., the lowest total error sums are in GARCH models with the skewed student t distribution.

For stock returns of State Street Corporation, the lowest total error sums are in the GJR model with the skewed student t distribution.

For stock returns of The Bank of New York Mellon Corporation and Allianz SE, the lowest total error sums are in the GJR model with the normal distribution.

5.1.2 Information criterion

To find the best predicting model for our cases, we compared models by their likelihood and information criteria. The higher the likelihood is, and the lower the information criterion is, the better model is. However, there is a risk that a model will be overfitting since the likelihood is calculated in the-sample. So, the information criterion is more significant.

Having compared these parameters for our four predicting models, we can sum up that the best likelihood and information criteria are in GJR with skewed Student t distribution. It means that this type of model is the best among the others above for the 6 selected funds.

5.2 Model validation

5.2.1 Ljung-Box test

One of the important features of a good GARCH model is zero correlation among absolute standardized predicted returns. This feature verifies that the distribution of future returns isn't affected by the past.

To test whether there is or no correlation in predicted returns, we can use two different approaches:

1. With `acf()` function we can plot correlation of our R_t for each of previous returns in our rolling window (in our case, it is 22 days $\sim$ 1 working month).

2. We can use Ljung-Box test to check if there is a correlation between our predicted returns in rolling window. The H_0 is next

   H_0 : There is no correlation of the current return with previous in our rolling window.

   The α for hypothesis testing is 5%. If the p-value is greater of equal to 0.05, then we failed to reject H_0 and should consider that our model is valid.

5.3 Model evaluation

5.3.1 Obtained coefficients

Tables with coefficients for each model are added in the Appendix.

The robust coefficients of the Standard GARCH with the normal distribution of errors are the same as usual ones; however, their standard errors are higher.

The robust coefficients of Standard GARCH are the same as coefficients of GJR GARCH with the normal distributions of errors in both cases; however, their standard errors are higher as well.

The robust coefficients of Standard GARCH with the normal distribution of errors are higher than coefficients of the same model, but with skewed Student t distribution of errors. However, robust‘ standard errors are higher as well.

All robust coefficients of GJR GARCH with the normal distribution of errors are the same as of the same model with skewed Student t distribution of errors. However, robust coefficients have higher standard errors.

5.4 Predicting returns

The end date of our data sample is 2020-05-08. To test the quality of our predictions we evaluated volatility with a rolling window technique with a window size of 22 days up to end date. After we cut our data since 2020-04-30, to be able to compare our predictions with the actual values of volatility.

In the R file attached, there is code to predict volatility for the period after 2020-05-08. The accuracy of these predictions will be discussed on the project presentation.

There are graphs for each investment fund that show predicted volatility together with the actual. We can see that they generally coincide. It is the one more evidence of the effectiveness of our models.

Also, below them there is a comparison table with the value of predictions for each day in the specified prediction period.

6 Conclusion

In this project, we explored the volatility prediction for stock returns of 6 investment companies, which are the biggest in the world.

We used GARCH and GJR models since they are powerful tools for volatility prediction. These types of models were performed for r normal and student t distributions each and for all the funds. We compared all models by different significant parameters. So, we can conclude there is not the best model that shows the lowest total error sum of return volatility for all six selected funds. However, the GJR model with skewed Student t distribution shows the best likelihood and information criteria for all cases above.

The important part of the GARCH modeling is the proper choose of model specification. We presented a bunch of criteria such. But ,still, often the best model is the most intuitive and simple.

As the outcome of this project, we developed an effective customizable tool for modeling the volatility of investment funds. This solution can also be used with any other historical returns data.

References

1. Robert Engle 2001, GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics
2. QUANTSTART, Generalised Autoregressive Conditional Heteroskedasticity GARCH(p, q) Models for Time Series Analysis
3. Sergiy Ladokhin 2009, Volatility modeling in financial markets
4. Wikipedia, Skewed generalized t distribution
5. Wikipedia, Normal distribution
6. Alexios Ghalanos, `rugarch` R package reference manual
7. Alexios Ghalanos, Introduction to the `rugarch` R package

7 Appendix

Summaries

Table 1: Summary Statistics for the stock returns

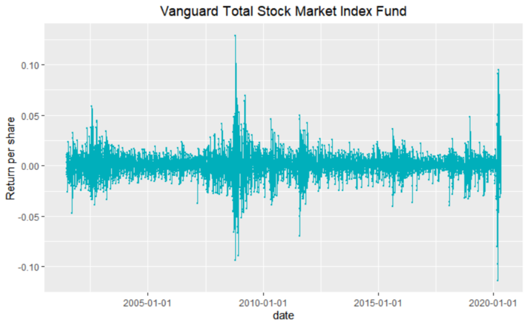
Company	**Observations**	**Mean**	**St.Dev.**	**Min**	**Max**
Vanguard	4753	0.0002815	0.01236052	-0.1138085	0.1282977
Blackrock	4753	0.0007982	0.02132793	-0.1552417	0.1956095
State Street	4753	0.0004141	0.02670451	-0.5903714	0.3134556
JPMorgan	4753	0.0004467	0.02436172	-0.2072743	0.2509673
BNY Mellon	4753	0.0001846	0.02333283	-0.2715778	0.2480650
Allianz SE	4685	0.00009	0.02193001	-0.15328	0.26245

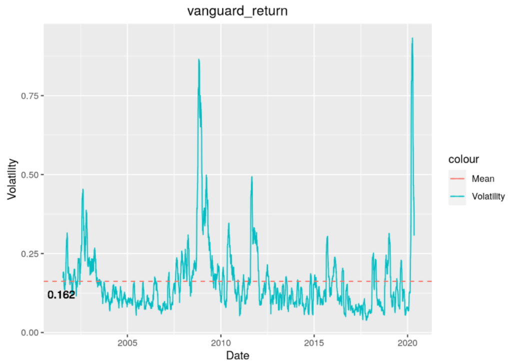
Table 2: Summary Statistics for volatility of the returns

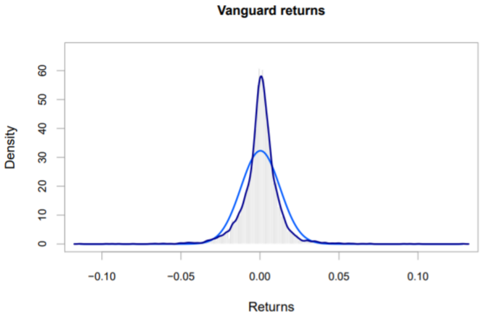
Company	**Observations**	**Mean**	**Min**	**Max**
Vanguard	4753	0.16197	0.03868	0.93240
Blackrock	4753	0.28848	0.09856	1.30542
State Street	4753	0.32310	0.08795	2.71482
JPMorgan	4753	0.30113	0.05013	1.63060
BNY Mellon	4753	0.29544	0.07755	2.0076
Allianz SE	4753	0.28552	0.06762	1.56906

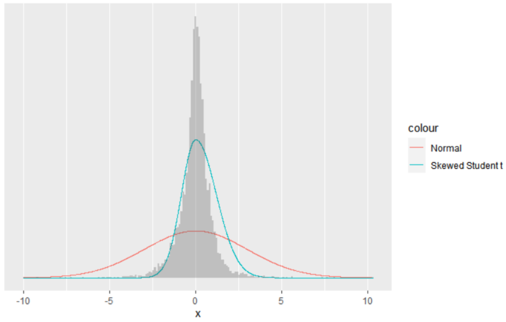
Visualization

Vanguard Total Stock Market Index Fund

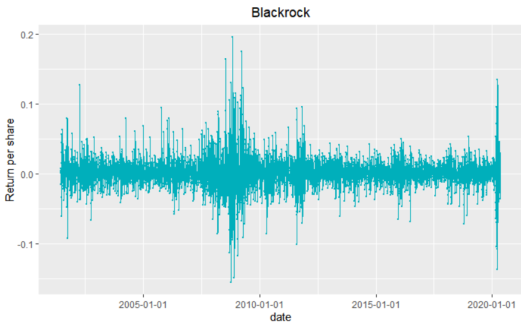


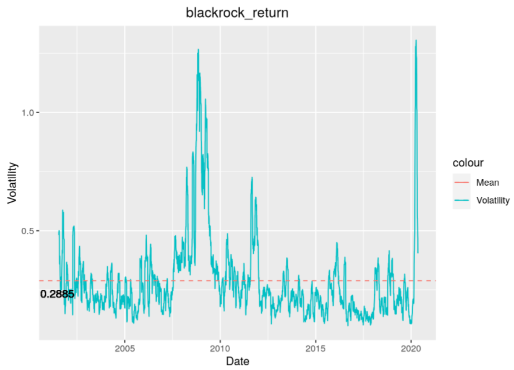


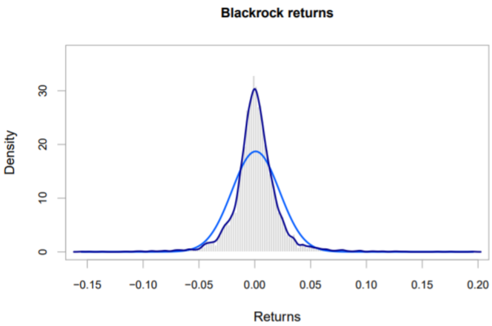


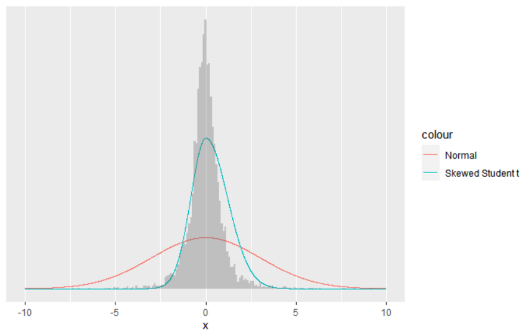


Blackrock

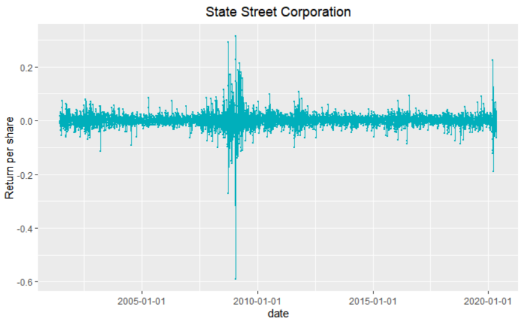


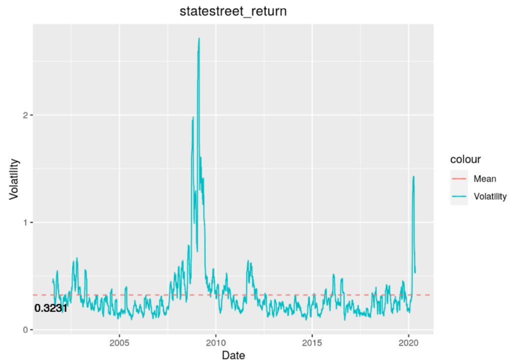


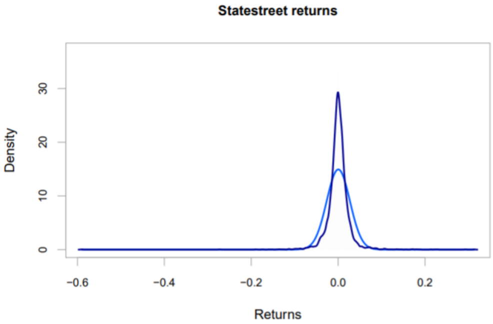


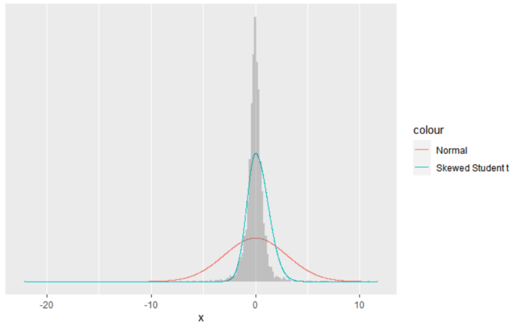


State Street Corporation

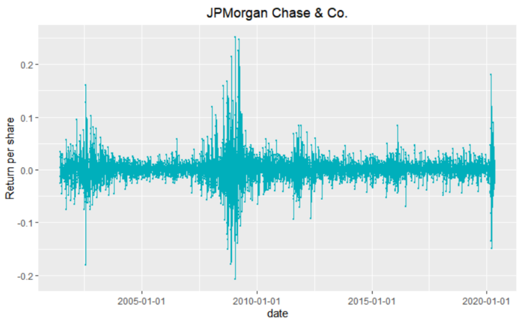


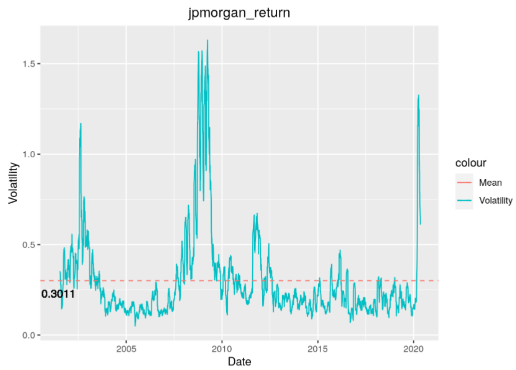


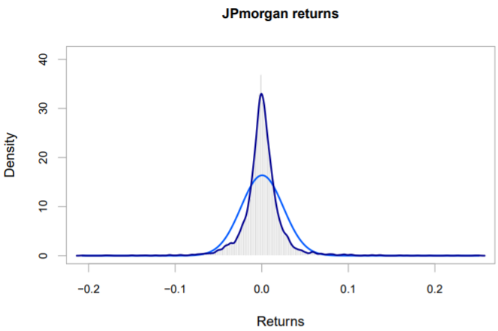


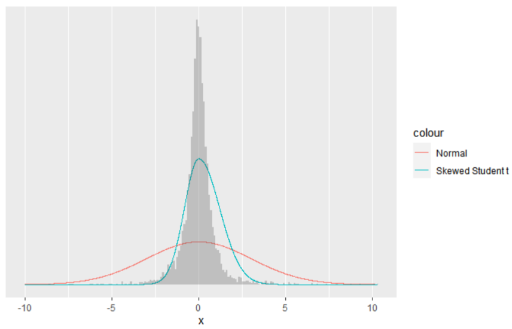


JPMorgan Chase & Co.

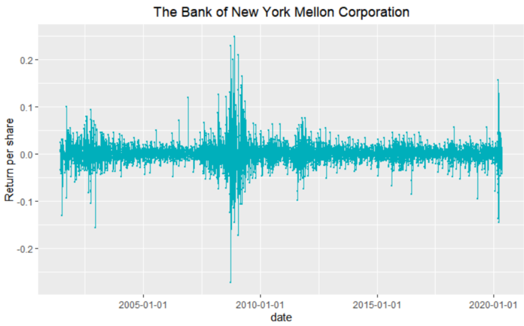


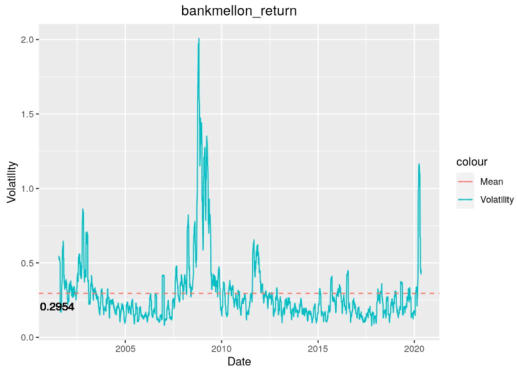


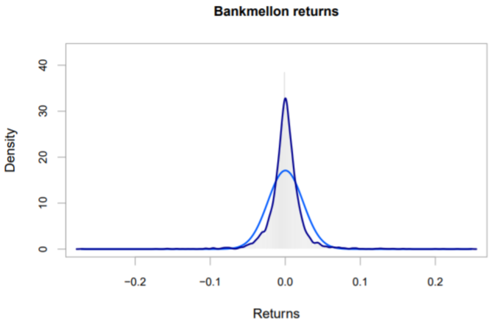


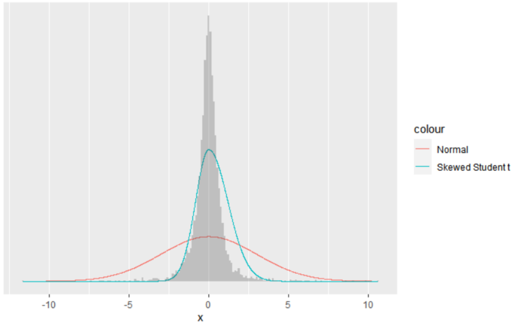


The Bank of New York Mellon Corporation

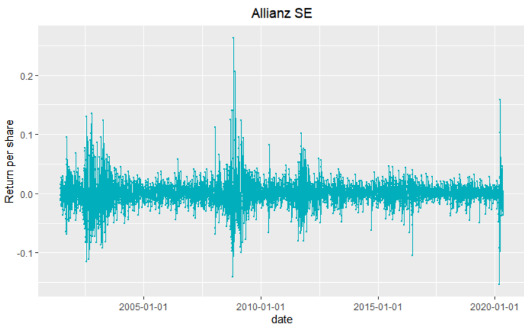


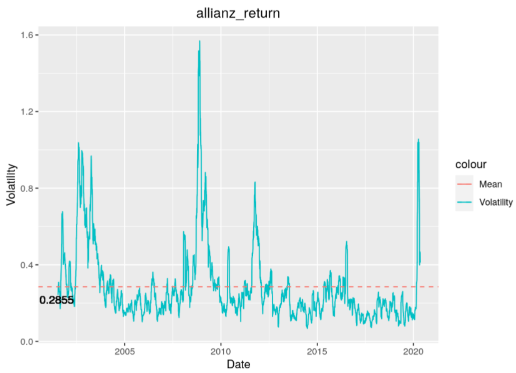


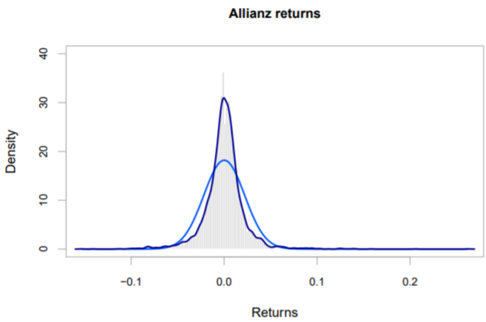


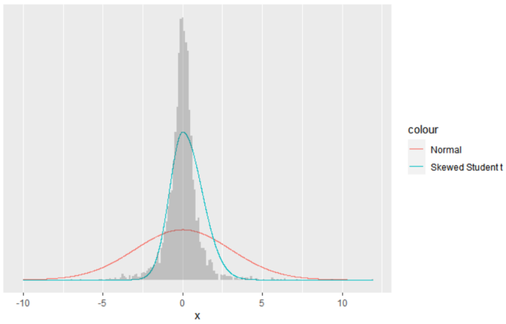


Allianz SE

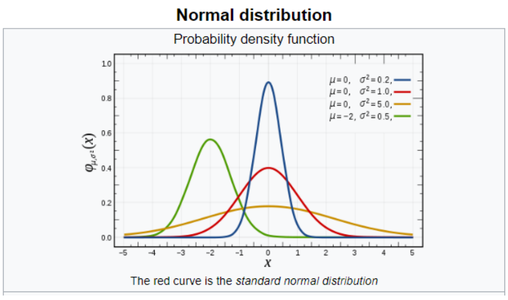


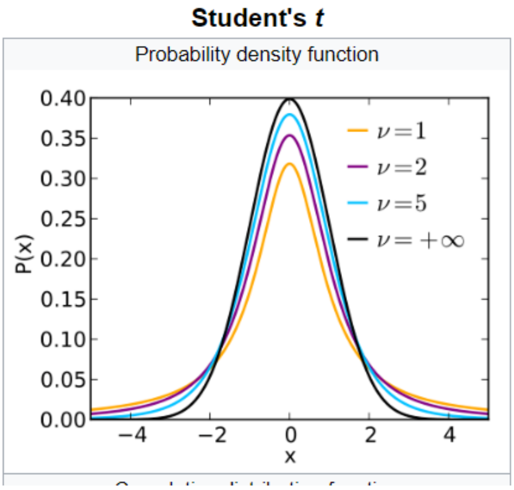






Changing the distributions depending on the parameters





Coefficients in Models

Vanguard

Table 3: Coefficients of Standard GARCH with normal distribution of errors

Variable	Coefficient	Std. Error	T-value	P-value
ω	0.0000024796	0.0000007827	3.167999	0.0015349183
α_1	0.1236562746	0.0098739025	12.523546	0.0000000000
β_1	0.8564012304	0.0106465927	80.438996	0.0000000000
μ	0.0006660551	0.0001134153	5.872709	0.0000000043

Table 4: Robust coefficients of Standard GARCH with normal distribution of errors

Variable	Coefficient	Std. Error	T-value	P-value
ω	0.0000024796	0.0000043291	0.5727754	0.5667967733
α_1	0.1236562746	0.0242583805	5.0974662	0.0000003442
β_1	0.8564012304	0.0397101847	21.5662868	0.0000000000
μ	0.0006660551	0.0000945864	7.0417614	0.0000000000

Table 5: Coefficients of GJR GARCH with normal distribution of errors

Variable	Coefficient	Std. Error	T-value	P-value
ω	0.0000024736	0.0000002385	1.037265e+01	0.000000000
α_1	0.0000001769	0.0033173146	5.333410e-05	0.999957445
β_1	0.8729915992	0.0066767202	1.307516e+02	0.000000000
μ	0.0002599302	0.0000981960	2.647054e+00	0.008119637
γ_1	0.2067021938	0.0129895853	1.591292e+01	0.000000000

Table 6: Robust coefficients of Standard GARCH with normal distribution of errors

Variable	Coefficient	Std. Error	T-value	P-value
ω	0.0000024736	0.0000005885	4.203159e+00	0.0000263216
α_1	0.0000001769	0.0140922669	1.255480e-05	0.9999899827
β_1	0.8729915992	0.0072802743	1.199119e+02	0.0000000000
μ	0.0002599302	0.0001309420	1.985079e+00	0.0471356725
γ_1	0.2067021938	0.0315575960	6.549998e+00	0.0000000001

Table 7: Coefficients of Standard GARCH with skewed Student t distribution of errors

Variable	Coefficient	Std. Error	T-value	P-value
ω	0.0000014778	0.0000018608	0.7941576	0.4271037050
α_1	0.1209330654	0.0313502725	3.8574805	0.0001145618
β_1	0.8717911859	0.0295432800	29.5089504	0.0000000000
μ	0.0006062493	0.0001117549	5.4248140	0.0000000580
$skew$	0.8888211922	0.0185481764	47.9195998	0.0000000000
$shape$	7.1084616197	1.1187769377	6.3537792	0.0000000002

Table 8: Robust coefficients of Standard GARCH with normal distribution of errors

Variable	Coefficient	Std. Error	T-value	P-value
ω	0.0000014778	0.0000155483	0.0950453	0.9242788484
α_1	0.1209330654	0.2496181390	0.4844723	0.6280507427
β_1	0.8717911859	0.2373673347	3.6727513	0.0002399530
μ	0.0006062493	0.0001642235	3.6916119	0.0002228373
$skew$	0.8888211922	0.0575212007	15.4520626	0.0000000000
$shape$	7.1084616197	6.9887168178	1.0171340	0.3090896792

Table 9: Coefficients of GJR GARCH with skewed Student t distribution of errors

Variable	Coefficient	Std. Error	T-value	P-value
ω	0.0000020107	0.0000010712	1.876978e+00	0.06052114
α_1	0.0000000450	0.0127386126	3.535900e-06	0.99999718
β_1	0.8732336384	0.0109951208	7.942010e+01	0.00000000
μ	0.0002495716	0.0001175544	2.123031e+00	0.03375129
$skew$	0.8546648715	0.0168886550	5.060586e+01	0.00000000
$shape$	8.2593642262	0.9887244708	8.353555e+00	0.00000000
γ_1	0.2246589215	0.0309933212	7.248624e+00	0.00000000

Table 10: Robust coefficients of GJR GARCH with normal distribution of errors

Variable	Coefficient	Std. Error	T-value	P-value
ω	0.0000020107	0.0000056600	3.552445e-01	0.72240643
α_1	0.0000000450	0.0445173275	1.011800e-06	0.99999919
β_1	0.8732336384	0.0338930298	2.576440e+01	0.00000000
μ	0.0002495716	0.0002846399	8.767975e-01	0.38059664
$skew$	0.8546648715	0.0164777668	5.186776e+01	0.00000000
$shape$	8.2593642262	1.0175550758	8.116872e+00	0.00000000
γ_1	0.2246589215	0.1280897966	1.753917e+00	0.07944466

Dependence of variance on errors in different models