FM442 Project

52826

January 20, 2023

1 Introduction

Risk and uncertainty have played a major role during the financial crisis. This project will explore the research question: "How did the risk measures and the quality of forecast models differ during the financial crisis of 2008, COVID-19, and the Russia-Ukraine War for companies such as Goldman Sachs and Amazon?" Since they are recent events, less work has been conducted on events like COVID-19 and Russia-Ukraine War. Furthermore, the majority of the work only takes into account VaR. Even though the main focus would be on VaR, other risk measures are also discussed. Thirdly, the returns of these companies have not been compared before for these crises. It will help the reader observe which financial crisis was worse for which industry and determine how the risk models behaved across companies and within the companies. Some previous research includes a paper compiled by the Lund University for the empirical evaluation of the VaR in the financial crisis of 2008. Another essay from the Hanken School of Economics evaluated the predictive performance of VaR models on Nordic Market Indices.

2 Data

2.1 Prices and Returns

This paper focuses on the logarithmic returns of Amazon and Goldman Sachs in America. The former is a multi-technological company, and the latter is a financial services company. The price chosen is the adjusted closing price which deducts the dividends to give an improved overall value. The returns are calculated by taking the difference between the logarithmic prices.

2.2 Data Collection

The data was collected from Finance Yahoo since it has the latest stocks. The reason for choosing these companies was because they are the most notable in their respective industries. The insights derived from this research would be widely applicable. The sample data is from January 2004 to January 2023 (4789 trading days).

3 Empirical Analysis

The empirical method is to compare the risk measures and models in different financial crises for the two companies. This paper conducts stress testing to check the predictive performance of the risk models. After a simple violation ratio analysis, some formal tests of the significance of violation ratios such as Coverage tests and Independence tests determine which models are most significant. The risk measures discussed are volatility, VaR, and Expected Shortfall (ES); however, the main focus will be on VaR since it is the most common risk measure, and its backtesting is much simpler than ES. The risk forecast models discussed are EWMA, GARCH, t-GARCH, and Historical Simulations. Before using the GARCH model to measure VaR, the ACF of the squared residuals were checked.

3.1 Functions in R

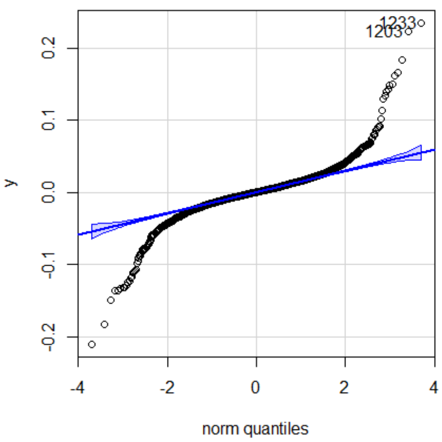
The 5% VaR is used to get a higher number of observations and to make statistical inference easier. The holding period is one day and the portfolio value is $1000. The GARCH(1, 1) model is used and it excludes the mean to de-mean so that the unconditional mean is subtracted from returns. For EWMA, the parameter λ is 0.94 according to J.P. Morgan's risk metrics approach. The parameters of the Coverage test and Independence Test are 5%, and the violations. The estimation windows are 500 and 1000 for HS and GARCH, and 1000 for t-GARCH and EWMA. Higher estimation windows were not chosen because of the 5% VaR.

The programming language used to compute the code is R. The libraries include `moments` to determine the normality of the returns, `rugarch` for implementing the GARCH model, and `car` for QQ plots. Other standard packages include `lubridate`, `car`, `reshape2`, `dplyr`, `zoo` for better time series plots. The `QRMlib` is used to calculate the degrees of freedom for Student-t.

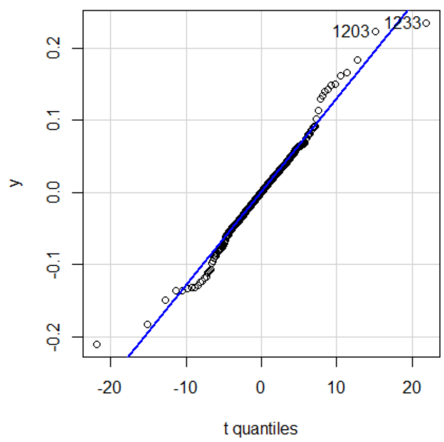
3.2 Analysis of Returns

The QQ plot shows that the returns are non-normal since fat tails exist; therefore, the returns have more extreme outcomes than a normal distribution. Normality determines if the volatility is the accurate measure of risk. The returns for both are fat relative to t(3) on the downside and upside.

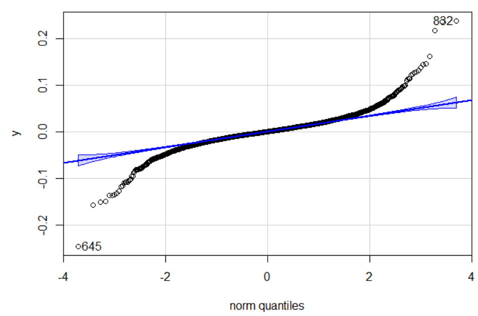
(a) Normal (Amazon)



(b) Student t(3) (Amazon)



(c) Normal (Goldman Sachs)



(d) Student t(3) (Goldman Sachs)

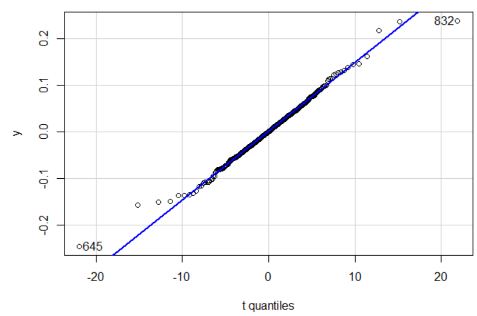


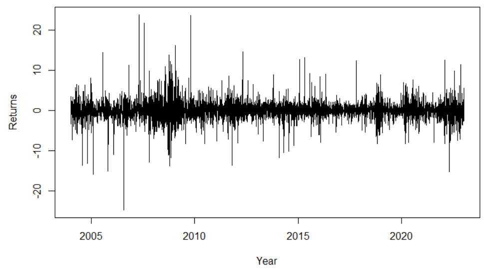
Figure 1: QQ plots for Amazon and Goldman Sachs returns

The table below shows the descriptive statistics of the returns for Amazon and Goldman Sachs. Both have means lower than their volatility. The LB test using 20 lags of returns shows a significant return predictability for both, and the volatility graphs show a few volatility clusters. The returns show high volatility during all three financial crises; however, the most volatile year for both was 2008 due to visibly wide volatility clusters.

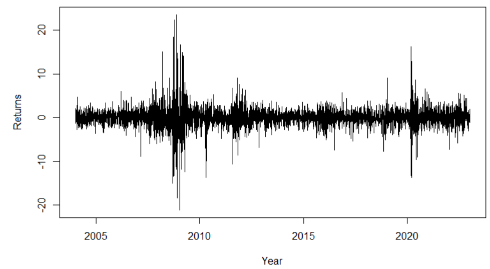
Table 1: Daily returns sample statistics 2004-2023

Stock	Mean	StDev	Min	Max	Skewness	Kurtosis	LB test
Amazon	0.075	2.431	-24.618	23.862	0.279	12.489	0
Goldman Sachs	0.033	2.204	-21.022	23.482	0.184	15.803	0

(a) Amazon Returns



(b) Goldman Sachs Returns



4 Risk Measures

4.1 Volatility

The paper will start with the most common risk measure known as volatility which is the standard deviation of returns. The values for each of the crises are given in Table 2 and 3:

The overall volatility is similar for both companies; therefore, the percentage changes depict that Goldman Sachs stocks were more volatile than that of Amazon during 2008 and COVID-19; however, for Russia-Ukraine war, Amazon stock prices were more volatile than Goldman Sachs. The reasons for these differences are discussed in Section 6.

Furthermore, the low volatility in Russia-Ukraine war for Goldman-Sachs does not mean that the risk was low due to fat tails which leads to gross underestimation of risk. The same is true for Amazon during COVID-19. Therefore, we explore other risk measures.

Table 2: Amazon

	Overall	Financial Crisis 2008	COVID-19	Russia-Ukraine War
Volatility	0.0243	0.0369	0.0202	0.0316
Percentage δ	-	51.9	-16.78	29.95

Table 3: Goldman Sachs

	Overall	Financial Crisis 2008	COVID-19	Russia-Ukraine War
Volatility	0.022	0.0439	0.0258	0.0185
Percentage δ	-	99.36	17.06	-16.11

4.2 Value at Risk and Expected Shortfall

Unlike VaR, Expected Shortfall is sub-additive, which provides more information about the tail VaR since it takes into account the tail distribution while VaR does not.

Table 4 shows the values of VaR and ES calculated using Historical Simulation. The biggest difference between VaR and ES for both the companies during the financial crisis of 2008 and this is the period where the volatility was the highest for both as shown in Table 2 and 3. There was a vast difference for Amazon during Russia-Ukraine War. Therefore, it follows a pattern, the higher the volatility, the higher the difference between VaR and ES.

Table 4: VaR and ES

	Risk Measure	Financial crisis 2008	COVID-19	Russia-Ukraine War
Amazon	VaR	51.10	29.36	53.54
	ES	77.39	44.18	71.24
Goldman Sachs	VaR	61.99	31.35	29.72
	ES	102.26	59.32	39.90

5 Stress Testing of VaR

The table below shows which model had the best and worst predictive performance for the financial crisis for each company.

Table 5: Stress Testing Amazon

Model	Estimation Windows	Financial Crisis 2008	COVID-19	Russia-Ukraine War
EWMA	1000	Acceptable	Good	Acceptable
HS	500	Bad	Good	Useless
HS	1000	Useless	Good	Useless
GARCH	500	Good	Good	Bad
GARCH	1000	Good	Good	Bad
t-GARCH	1000	Good	Good	Good

Table 6: Stress Testing Goldman Sachs

Model	Estimation Windows	Financial Crisis 2008	COVID-19	Russia-Ukraine War
EWMA	1000	Good	Good	Good
HS	500	Useless	Good	Good
HS	1000	Useless	Acceptable	Good
GARCH	500	Acceptable	Good	Good
GARCH	1000	Good	Good	Good
t-GARCH	1000	Good	Good	Good

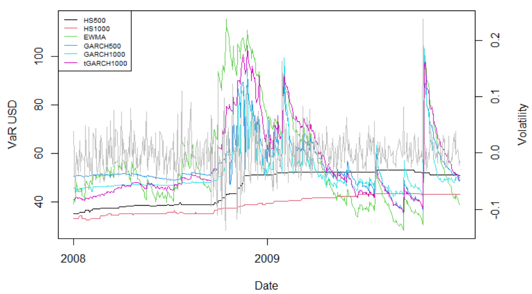
Table 7: Hierarchy of Predictive Performance

	Financial Crisis 2008	COVID-19	Russia-Ukraine War
Amazon	Best - GARCH1000 Worst - HS1000	Best - EWMA Worst - HS1000	Best - tGARCH1000 Worst - HS1000
Goldman Sachs	Best - EWMA Worst - HS1000	Best - GARCH500 Worst - HS1000	Best - HS500 Worst - EWMA

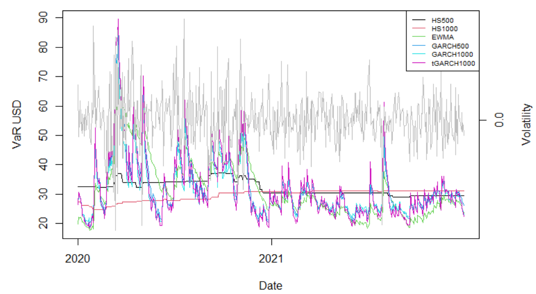
When the volatility is high, the predictive performance is bad. For example, Goldman Sachs had wider volatility clusters than Amazon in 2008; the predictive performance of VaR models for Goldman Sachs is worse than Amazon. Furthermore, the volatility for Amazon in 2022-23 was higher than Goldman-Sachs; therefore, the models were mostly bad for the models; whereas for Goldman-Sachs all models performed well since the volatility was low. However, the only model which performed well in all three stages for both companies was the t-GARCH model because it takes into the non-normality of the returns since the returns had fat tails.

The following sub-figures show the backtesting models for VaR with the grey line depicting the actual volatility. The general analysis of graphs suggests that for both the companies, the HS plots are relatively stable and less reactive to high volatility periods; whereas EWMA, GARCH and t-GARCH were reacted sharply when the volatility increased or decreased.

(a) Focus on 2008



(b) Focus on COVID-19



(c) Focus on Russia-Ukraine War

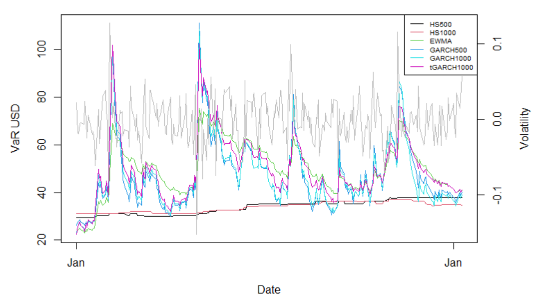
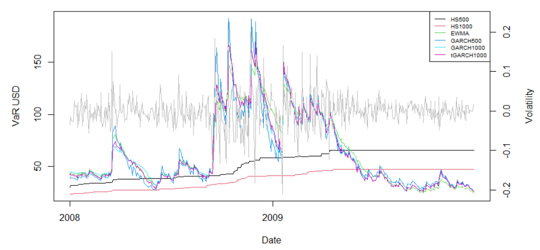
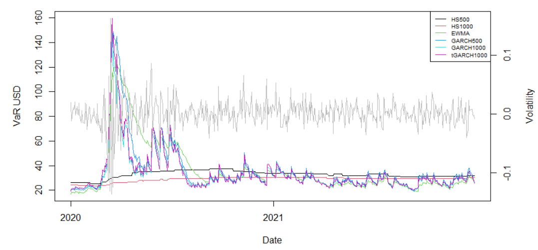


Figure 3: Amazon

(a) Focus on 2008



(b) Focus on COVID-19



(c) Focus on Russia-Ukraine War

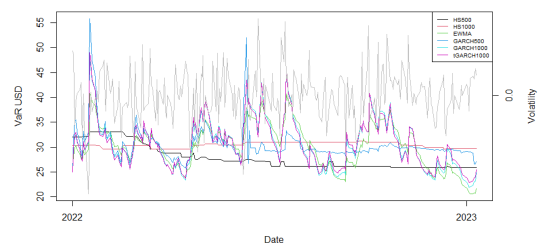


Figure 4: Goldman Sachs

6 Violation Ratios

This section will describe how the violations of Amazon and Goldman Sachs differ during these three stages. The models that under forecast risk are highlighted in green; while those that over forecast risk are in red.

6.1 Financial Crisis 2008-09

As shown in Table 8, Goldman Sachs observes much higher violations than Amazon for all the risk models. Goldman Sachs played a major role in the Financial Crisis of 2008 as it participated in converting mortgage debt into securities. Therefore, it was impacted more than Amazon which performed better since it launched Amazon Prime, Amazon Kindle, and AWS during this period. Most models are over forecasting the risk for Amazon and under forecasting the risk for Goldman-Sachs.

Table 8: Violations in Financial Crisis 2008

Model	Amazon	Goldman Sachs
EWMA	0.752	1.069
HS500	1.545	2.059
HS1000	2.02	3.05
GARCH500	0.832	1.228
GARCH1000	0.911	1.109
t-GARCH1000	0.911	1.188

6.2 COVID-19

Table 9 shows that Goldman Sachs had more violations than Amazon; however, the gap is not as intense as the previous one since the volatility cluster for Goldman Sachs is wider than that of Amazon. However, the volatility cluster of Goldman Sachs for 2008-09 is smaller than 2020-21; therefore, complementing the results.

Table 9: Violations in COVID-19

Model	Amazon	Goldman Sachs
EWMA	1.030	1.109
HS500	0.832	1.109
HS1000	1.188	1.347
GARCH500	0.95	0.911
GARCH1000	0.95	1.069
t-GARCH1000	0.95	0.99

6.3 Russia-Ukraine War

The violation ratios for Amazon are higher than Goldman Sachs unlike the previous two. Effected supply chain, higher staff and fuel costs have impacted e-commerce industry. There might be other factors than the war since the e-commerce businesses had to increase the pay to attract workers after COVID-19. Almost all the models are under forecasting the risk for both Amazon and GARCH.

Table 10: Violations in Russia-Ukraine War

Model	Amazon	Goldman Sachs
EWMA	1.318	1.081
HS500	2.558	1.004
HS1000	2.636	0.927
GARCH500	1.628	1.081
GARCH1000	1.55	1.081
t-GARCH1000	1.163	1.081

7 Significance of Back-testing

To keep it concise, this section will perform significance tests, including Bernoulli Coverage Tests and Independence Test for Goldman-Sachs only. The violations for each of the model is greater than 10; therefore, the significance tests can be performed. The tests are conducted with a significance level of 5%.

7.1 Bernoulli Coverage Test

This test tests the significance of violation ratios i.e. if they are statistically equivalent to one. Table 11 depicts the results of each financial crisis. It shows that HS models were rejected for Financial Crisis 2008 which had the highest volatility and violations. In other financial crisis, the models performed well and the violation ratios were statistically equal to 1 for all crises.

Table 11: Results for Bernoulli Coverage Tests

	Model	p-value	Result
Financial Crisis 2008	HS500	0	**reject**
	HS1000	0	**reject**
	EWMA	0.587	accept
	GARCH500	0.185	accept
	GARCH1000	0.454	accept
	tGARCH1000	0.346	accept
COVID-19	HS500	0.581	accept
	HS1000	0.089	accept
	EWMA	0.346	accept
	GARCH500	0.959	accept
	GARCH1000	0.581	accept
	tGARCH1000	0.724	accept
Russia-Ukraine War	HS500	0.989	accept
	HS1000	0.784	accept
	EWMA	0.568	accept
	GARCH500	0.767	accept
	GARCH1000	0.767	accept
	tGARCH1000	0.767	accept

7.2 Independence Test

The test above excludes the time variation in data and to test if violations cluster, the independence tests are implemented.

Table 12: Results for Independence Tests

	Model	p-value	Result
Financial Crisis 2008	HS500	0.092	accept
	HS1000	0.255	accept
	EWMA	0.616	accept
	GARCH500	0.397	accept
	GARCH1000	0.556	accept
	tGARCH1000	0.499	accept
COVID-19	HS500	0.003	**reject**
	HS1000	0.005	**reject**
	EWMA	0.122	accept
	GARCH500	0.152	accept
	GARCH1000	0.270	accept
	tGARCH1000	0.226	accept
Russia-Ukraine War	HS500	0.675	accept
	HS1000	0.573	accept
	EWMA	0.887	accept
	GARCH500	0.780	accept
	GARCH1000	0.780	accept
	tGARCH1000	0.780	accept

All models performed well except for the HS models in COVID-19.

The tables above show that the unconditional coverage and independence tests are different from each other, and if one test accepts of the null hypothesis, the other might not, for example the HS model is not rejected in the first test but got rejected in the second one during the COVID-19. This can be justified since it is possible that 5% VaR can generate 5% violations only; however, the violations are squeezed into a short period during COVID-19 but if these violations were spread out evenly, the independence test might have not failed.

8 Analysis

The predictive performance of the risk models behave differently for each financial crisis. The key takeaway is that each financial crisis affects all industries; however, each involves different political and economic factors. Therefore, the magnitude of the impact would be different for each industry. The risk models behave differently for tranquil and crisis years, suggesting that quantitative analysts should not choose the risk models just based on the overall predictive performance. The models which did not perform well, for example, HS1000, their parameters need to be reevaluated by the analysts. A good start would be to decrease the estimation windows.

The results suggest that since HS is based on fixed weights of returns hence it is not suitable for structural changes in risks. The GARCH and t-GARCH performed better for all financial crises since it considers conditional volatility; therefore, it adjusts more sharply to changes in volatility.

Furthermore, VaR is only a quantile i.e. the minimum potential loss on the profit/loss distribution. It also is not a coherent risk measure since it does not satisfy the subadditivity axiom. The problem with backtesting is that it does not take into account the structural breaks in the data; whereas, due to external factors such as technology and politics, the financial markets are continuously changing. This suggests that the predictive performance of risk models changed during these 15 years since the past 15 years have been the years with the most technological advancements.

The limitation of the significant tests is that it relies on asymptotic distributions. The sample size of violations is small since the violations occur infrequently and the test was split into three small data sets; therefore, the results might not be accurate. To further research on the topic and reach better results; we can carry out the confidence bounds through simulations.

9 Conclusion

In conclusion, its not the risk measures that fail during the crises but the implementation of these risk measures. The distribution of returns is the key to forecasting risk during unforeseen events; however, it is impossible to do so. Therefore, a distribution-free risk would be better for forecasting purposes. VaR provides a better approach; however, it is not sub-additive. For further research, Expected Shortfall is a better risk measure which takes into account the tail distribution of VaR. However, due to technicalities, it is harder to implement in real life situations especially crises. For backtesting t-GARCH is the best model since it takes into account the fat tails. The initial value of the conditional volatility can be initialized by EWMA to reach better predictive performance.

10 References

Financial Risk Forecasting. Jon Danielsson. 2011. www.wiley.com/finance

Extreme Value at Risk and Expected Shortfall during Financial Crisis. L. 2011. Kourouma, Denis Dupré, G. Sanfilippo, O. Taramasco. https://shs.hal.science/halshs-00658495

Analysis of Risk using Value at Risk (VaR) After Crisis in 2008 Study in Stocks of Bank. 2011. Hasna Fadhila, Nora Amelda Rizal, Institute Management Telkom

Abdullah Jobayed. Evaluating the Predictive Performance of Value-at-Risk (VaR) Models on Nordic Market Indices. 2017. https://www.researchgate.net/publication/324687617