Comparison of GARCH and Data Science Models in Financial Times Series Forecasting: An analysis for the financial market volatility in the banking and automobile industry.

Mohammed Khalil EL BIYAALi and Diana Patricia Carvajal Rozo

Mentor: Xuehai Zhang

Jacobs University

Master's degree in Data Science for Business and Society.

This work was carried out with the assistance of Xuehai Zhang, consultant Ernst & Young. It is supervised by Adalbert Wilhem and Hilke Brockman from the Data Lab course.

Comparison of GARCH and Data Science Models in Financial Times Series Forecasting: An analysis for the financial market volatility in the banking and automobile industry.

Mohammed Khalil El Biyaali and Diana Patricia Carvajal Rozo

Emails: m.elbiyaali@jacobs-university.de; d.carvajalrozo@jacobs-university.de

I. INTRODUCTION

Volatility is commonly used in financial investment to gauge the dispersion of asset returns over time. In statistics, volatility is calculated as either a standard deviation or variance of price returns. Volatility rises as the risk of a financial asset increases. Investors/traders can use such information to distinguish between low-risk and high-risk assets. (Bee Guan , 2021)

In this paper, we try to compare the performance of GARCH models with data science models by trying to forecast returns of 4 companies. The objective is to find the model that gives the best result in the case of a very high volatility. Indeed, we have chosen to compare the evolution of returns between the German and American markets in 2 different industries, namely:

- The banking sector by comparing Deutsche Bank AG (DBK) and Bank of America Corp (BAC)
- The automotive industry by comparing BMW and Ford Motors

The machine learning models we have chosen to make this comparison are:

- GARCH
  - ARCH, GARCH, GARCH-t, E-GARCH and GJR-GARCH.
- Data Science Models
  - LSTM and Random Forest

The data source is presented in section II. The methodology incorporating 7 forecasting models in this work is studied in section III. The comparison and discussion are exposed in section IV. Finally, the conclusion is presented in section V.

II. DATA SOURCE

For the analysis we have been using yearly data of four international stock markets, including two different industries (Banking and automobile) collected from YahooFinance. Is important to state that YFinance is an open- source Python Library that allowed us to obtain stock data for any given period, in this case 20 years. In other words, from 2000-01-01 to 2020-12-31.

III. METODOLOGY

The term time series refers to a set of observations made at different points in time about the values of a (quantitative) variable. It is possible for the data to behave differently over time, there may be a trend, a cycle; they may not have a defined shape or be random; in addition, they may be subject to seasonal variations (annual, semi-annual, and so forth). The observations of a time series will be denoted by $Y1; Y2..., YT$, where Yt is the value taken by the process at time t. (Rios & Hurtado, 2008)

In this analysis, by one hand, the most recent developed models, GARCH, have been used to overcome high volatility problem in forecasting financial time series. In the GARCH process, we additionally model the volatility (Conditional variance...) (Altaf, 2008).

And by the other hand, with Data Science models like LSTM and Random Forest we expect to forecast simultaneously the stock returns and movements by this multi-tasking model. (Park, Youngjun, & Ha Young, 2022).

A. *GARCH models*

Eagle (1982) proposed the ARCH model as one of the early attempts to model volatility. In ARCH, historical asset returns are analyzed in a univariate model. Thus, the ARCH(p) model has the following form:

$$\sigma_t^2 = \omega + \sum_{k=1}^{p} \alpha_k (r_{t-k})^2$$

According to those equations, ARCH is a univariate and nonlinear model in which volatility is estimated based on the square of past returns. Additionally, distinctive property of ARCH is its time-varying conditional variance, which means, that ARCH model can model the volatility clustering mechanism, where large changes tend to follow large changes of either sign, and small changes tend to follow small changes, as described by Mandelbrot (1963). (O´Reilly, s.f.)

The GARCH model is an extention of ARCH Integrating lags into conditional variance. Therefore, ARCH has been improved by incorporating p-delayed conditional variances. This makes the GARCH model multivariate, since it consists of an autoregressive moving average model for conditional variance with p lagged square returns and *q* lagged conditional variances. GARCH (*p*, *q*) can be formulated as:

$$\sigma_t^2 = \omega + \sum_{k=1}^{q} \alpha_k r_{t\,k}^2 + \sum_{k=1}^{p} \beta_k \sigma_{t-k}^2$$

Through the use of the GARCH model, it is possible to reduce the number of estimated parameters from an infinite number to a small number and while at the same time, GARCH models are better at capturing unusual volatility because they model and forecast both mean and conditional variance. (Altaf, 2008).

Moreover, it is important to acknowledge that GARCH model with t-Student innovations is particularly useful for defining excess kurtosis[1] in the conditional distribution of returns, which is a stylized reality in most financial data sets. (Espinosa Acuña, 2016)

As a continuation of the classification, In the GJR-GARCH model, bad news has a greater. impact than good news. This is an example of asymmetric effects of announcements. To put it another way, when there is asymmetry in the distribution of losses, the tail is fatter. The equation of the model includes one more parameter, γ, and it has the following form:

$$\sigma_t^2 = \omega + \sum_{k=1}^{q} \left(\alpha_k r_{t-k}^2 + \gamma r_{t-k}^2 I(\epsilon_{t-1} < 0)\right) + \sum_{k=1}^{p} \beta_k \sigma_{t-k}^2$$

In this case, γ contain measures for the asymmetry of the announcements if
$\gamma = 0$ There is no difference in response to the previous shock.
$\gamma > 0$ Negative shocks elicit a stronger response than positive ones.
$\gamma < 0$ Positive shocks have a stronger response than negative shocks.

Lastly, another tool for controlling for asymmetric announcements is the EGARCH model, proposed by Nelson (1991). As a result, no restrictions are required to prevent negative volatility, since the specification is in logarithmic form:

$$\log(\sigma_t^2) = \omega + \sum_{k=1}^{p} \beta_k \log \sigma_{t-k}^2 + \sum_{k=1}^{q} \alpha_i \frac{|r_{k-1}|}{\sigma_{t-k}^2} + \sum_{k=1}^{q} \gamma_k \frac{r_{t-k}}{\sigma_{t-k}^2}$$

[1] Statistically, kurtosis measures the degree of concentration of values around the central zone of the frequency distribution. (San Juan, s.f)

The main difference in the EGARCH equation is that logarithm is taken of the variance on the left-hand side of the equation. A negative correlation exists between past asset returns and volatility, which indicates the leverage effect. If $\gamma < 0$, it implies leverage effect, and if $\gamma \neq 0$, it indicates volatility asymmetry. (O´Reilly, s.f.)

B. LSTM and Random Forest for predict the volatility in banking and automobile industry

An LSTM model overcomes vanishing gradient/exploding gradient problems that commonly occur when learning long-term dependencies, even after minimal time lags. In resume, A LSTM architecture consists of a set of recurrently connected sub-networks, called memory blocks and by using nonlinear gates, the memory block maintains its state over time and regulates the flow of information. (Napoles, Van Houdt, & Mosquera, 2020)

Conversely, Random Forest is an algorithm developed by Leo Breiman and Adele Cutler for combining the outputs of multiple decision trees. Since it can handle both classification and regression problems, it has been widely adopted due to its ease of use and flexibility. How it works?
Before training, the random forest algorithm needs to be configured with three main hyperparameters. A number of factors are considered in this process, including node size, number of trees, and number of features sampled. Once the random forest classifier is constructed, it can be used to solve regression or classification problems. (IBM Cloud Education, 2020)

C. Performance criteria

Although there are different measures for assess the results of each model, we decided to choose RMSE. Having in mind that RMSE has a double purpose: to serve as a part of the algorithm for the training models and analyze the accuracy of the model that has been implemented, it´s necessary to point out that RMSE gives higher weight to a large error which result that this measure is more helpful when large errors are unwanted. (JJ, 2016)

IV. RESULTS AND DISCUSSION

A. Explorative Data Analysis (EDA)

We examined the four companies in the two different industries (Banking and automobile). In this stage, the calculation of daily returns and historical prices per company and per industry were made.

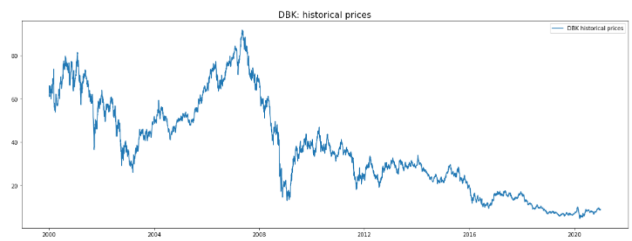

Figure 1 Historical Prices DBK

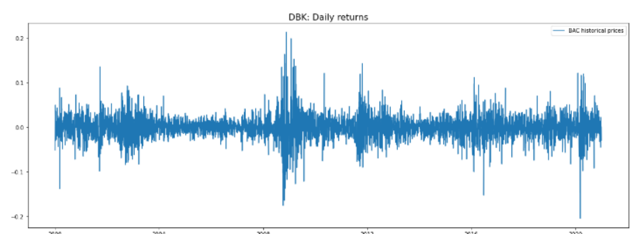

Figure 2 Daily returns DBK

We can see in Figure 1 that DBK prices have fluctuated significantly. Indeed, the financial crisis of 2003 (new market) which arrived in Germany strongly impacted the prices of the DBK. It is only from about 2004 that the prices started to increase (with some decreases) to fall again in a drastic way because of the world financial crisis of 2007 - 2008. Afterwards the prices increased a little bit but started to fall again because of a 3rd crisis between 2011 and 2012 (Euro debt crisis). Prices continued to fall throughout the period shown on the graph. And we can justify the sharp decline in prices in 2020 with the COVID-19 health crisis that strongly impacted the global banking sector.

We can confirm this analysis with the high volatility of returns on graph 2 (Daily returns) during the periods 2003-2004, 2007-2010, 2011-2013 and 2016-2020.

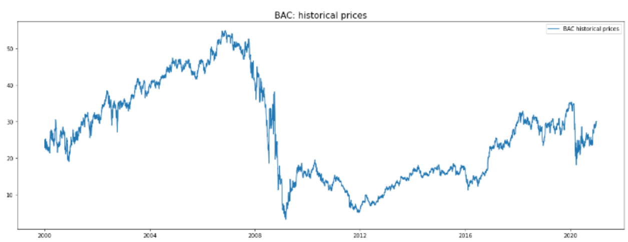

Figure 3 Historical prices BAC

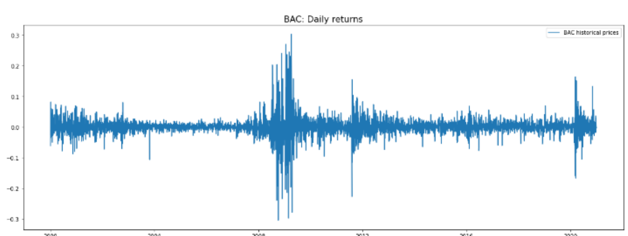

Figure 4 Daily returns BAC

We can apply the same analysis to the BAC case. However, in the United States, we can judge that the effects of the 2003 financial crisis were minimal. In fact, on graph 3, we can see that prices continued to rise during this period. The same thing for the period 2011-2012. The 2 major events where we notice a strong decrease in prices, as well as a high level of volatility (graph 4) are the global crisis of 2007 and the COVID-19 health crisis.

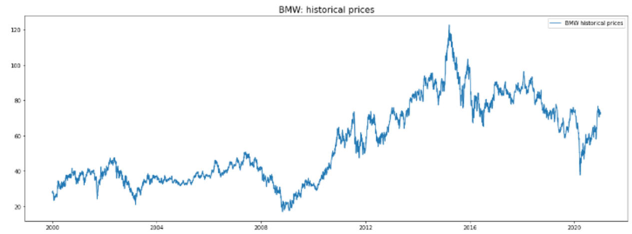

Figure 5 Historical Prices BMW

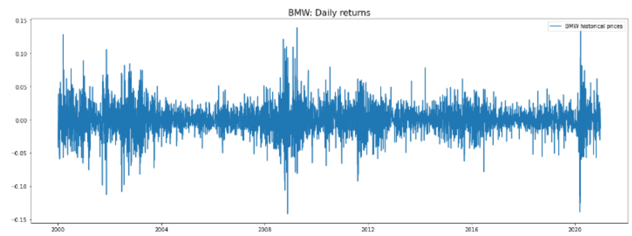

Figure 6 Daily returns BMW

According to graph 5, we can say that in the case of BMW the prices were impacted by the crises of 2003 and 2008 to reach their minimum in 2009-2010. Then, despite the crisis of 2011-2012, we can say that the prices have increased regardless of some decreases until reaching a maximum just before 2016, to start falling again. And we can well see the effect of the COVID-19 crisis, however, the prices started to increase again after 2020.

In graph 6, we can clearly see the high volatility of returns during the same periods mentioned above.

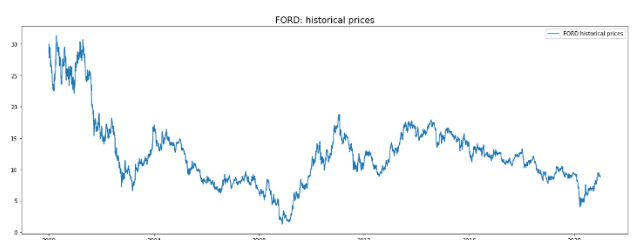

Figure 7 Historical prices Ford

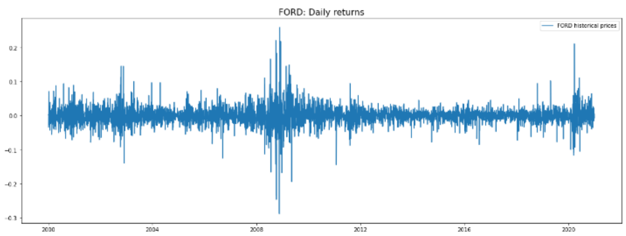

Figure 8 Daily returns Ford

According to graph 7, Ford's prices dropped drastically from 2002-2003 until they reached their minimum in 2009 due to the global crisis in 2007. After this period, the prices started to increase and then decreased in 2011, which means that the company was impacted by the 2011-2012 crisis that occurred in Germany. Thereafter, the prices continued to fall.

On graph 8, we can clearly see the high volatility of returns during these periods.

We can also add that according to the daily return plots, the global financial crisis of 2008 is the one that caused the most volatility.

Figure 9 Daily returns DBK/ BAC

From Figure 9, we can see that DBK was affected by the 2003 crisis unlike BAC. However, the financial crisis of 2008 had a much greater impact on BAC than on DBK

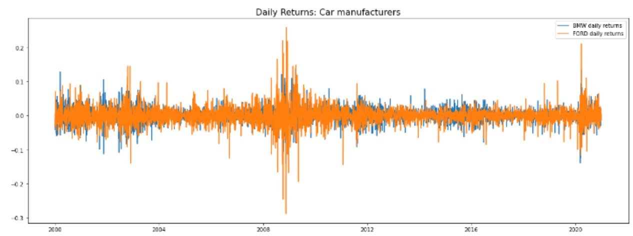

Figure 10 Daily returns BMW/ FORD

We can see that the level of volatility of Ford is higher than that of BMW. We can explain this by the fact that Ford was more affected by the crises that occurred during the period studied.

B. Comparison and discussion

For each company and industry, we designed and applied GARCH, LTSM and Random Forest models to demonstrate its importance of forecast the volatility of those markets. The forecast results of GARCH and Data Science models can be seen in table 1. Overall, according to the results given by the implementation of these models we could say that for the banking industry GARCH models perform much better than Data Science Models.

Models	RMSE DBK	RMSE BAC	RMSE BMW	RMSE FORD
ARCH	0.2504	0.2273	0.2005	0.2219
GARCH	0.2076	0.1982	0.1828	0.2174
GARCH-t	0.2074	0.2018	0.186	0.2236
EGARCH	0.2161	0.1997	0.1836	0.2243
GJR-GARCH	0.2072	0.1787	0.1722	0.2156
LSTM	2.6075	1.8953	2.5942	3.2206
RANDOM FOREST	2.0383	1.5683	1.8348	1.6323

Table 1 RMSE results

In the specific case of the designing and implementation of the GARCH models in the banking industry we could see by the results showed above that the worst input was from ARCH model with 0.2504 for DBK and 0.2273.

At the same time, even if we didn't take into consideration BIC as a performance criteria, we could refer that the worst result was from ARCH model as well with 2.3283.9 for DBK and 22449.5 for BAC. In contrast, the best model performance in the banking industry was from GJR GARCH.

Although GARCH models had acceptable results in the banking industry, this was not the case for data science models, specifically LTSM and Random Forest. In this regard, LSTM had the worst performance in both companies (DBK and BAC).

We can see a similar behavior in the automobile industry. In terms of performance, the ARCH model and the Data Science models showed the worst results. According to the ARCH model, BMW obtained a performance of 0.2005 while FORD achieved a result of 0.2219. Meanwhile, the worst performance was recorded in the application of the LTSM model. A BMW with a 2.594 and a FORD with a 3.2206. In the automobile industry, however, the GJR-GARCH model produced the most accurate results. BMW had 0.1722 and FORD had 0.2156.

V. CONCLUSION

The purpose of this article is to examine whether 7 models (GARCH and Data Science) can be applied to financial timeseries forecasting for four international stock markets, including two in Germany and two in US.

Our work proved that ARCH extensions perform much better than this model and Data Science models (LSTM and Random Forest) given the chosen measurement method. Previous studies say that LSTM performs very well for this type of projects that require breaking the traditional statistical framework, however, since we are looking to predict the volatility of the markets, GARCH and its extensions allow to predict the volatility based mainly on historical closed values.

The statistical methods of GARCH and Data Science Models require a large amount of data, reducing the original sample when the model fits. Perhaps if we design a LSTM with a different training method and compare it again with the results yielded by GARCH and its extensions, the input may be more favorable.

VI. REFERENCES

Altaf, H. (December of 2008). *ResearchGate*. Obtained from Comparison of GARCH and Neural Network Methods in Financial Time Series Prediction: https://www.researchgate.net/publication/251873538_Comparison_of_GARCH_and_Neural_Network_Methods_in_Financial_Time_Series_Prediction

Bee Guan , T. (3 de May of 2021). *Medium*. Obtained from How to Predict Stock Volatility with Python: https://python.plainenglish.io/how-to-predict-stock-volatility-with-python-46ae341ce804

Espinosa Acuña, O. A. (2016). Ajuste de modelos garch clásico y bayesiano con innovaciones T- student para el índice COLAP. *Revista Economía del Caribe*, 1-32.

Guagliano, C. (2018). Monitoring volatility in financial markets. *ESMA Report on Trends, Risks and Vulnerabilities*, 76-83.

IBM Cloud Education. (7 de December de 2020). *IBM*. Obtenido de Random Forest: https://www.ibm.com/cloud/learn/random-forest#toc-what-is-ra-DaEaNVdG

JJ. (23 de March de 2016). *Medium.com*. Obtenido de MAE and RMSE — Which Metric is Better?: https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d#:~:text=RMSE%20has%20the%20benefit%20of,then%20MAE%20is%20more%20appropriate.

Napoles, G., Van Houdt, G., & Mosquera, C. (2020). A Review on Long- Short Term Memory Model. *Artificial Intelligence Review*, 1-34.

O´Reilly. (s.f.). *Chapter 4. Machine Learning-Based Volatility Prediction*. Obtained from https://www.oreilly.com/library/view/machine-learning-for/9781492085249/ch04.html

Park, H., Youngjun, K., & Ha Young, K. (June de 2022). *Science Direct*. Obtained from Stock market forecasting using a multi-task approach integrating long short-term memory and the random forest framework: https://www.sciencedirect.com/science/article/abs/pii/S1568494621009947

Rios, G., & Hurtado, C. (2008). *Series de Tiempo*. Chile: Universidad de Chile.

San Juan, F. J. (s.f). *Economipedia*. Obtained from Curtosis: https://economipedia.com/definiciones/curtosis.html