Timeseries Analysis on Automobile Stocks (DAX)

GAURAV PANDEY
Department of Computer Science
Ramakrishna Mission Vivekananda Educational and Research Institute
Belur Math, Howrah
Pin – 711202, West Bengal

A report submitted to
Ramakrishna Mission Vivekananda Educational and Research Institute
in partial fulfilment of the requirements for the degree of
MSc in Big Data Analytics
2021

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my faculties for providing their invaluable guidance, comments and suggestions throughout the course of the project. I would specially thank Dr. Sudipta Das for helping me and constantly motivating me to work harder.

Also, I would like to thank my Head of Department Swathy Dhyanagamyananda for his assistance and support.

Ramakrishna Mission Vivekananda Educational and
Research Institute, Belur Math, Wet Bengal

Date: Jan 29, 2021

Gaurav Pandey

CERTIFICATE

This is to certify that the report entitled 'Timeseries analysis on Automobile Stocks in German stock market' submitted by Mr Gaurav Pandey, who has been registered for the award of MSc in Big Data Analytics degree of Ramakrishna Mission Vivekananda Educational and Research Institute, Belur Math, Howrah, West Bengal is absolutely based upon his/her own work under the supervision of Dr. Sudipta Das of Assistant Professor, RKMVERI and that neither his/her report nor any part of the report has been submitted for any degree/diploma or any other academic award anywhere before.

Date: Jan 29, 2021

Dr. Sudipta Das
Assistant Professor
RKMVERI

TABLE OF CONTENTS

ABSTRACT

This report is about using the data available on DAX index fetched from the finance platform named yahoo finance. From the DAX we are specifically focused on the automobile industry where my major focus on the stocks of three companies specifically, i.e., BMW, Volkswagen and Porsche. As later on Volkswagen was involved in the buyout of Porsche, this is what we are going to keep our major focus in the project and how this deal affected the entire German automobile industry. We have taken BMW stocks as keep track that how a major deal affected the entire market. It simply means that BMW is our market benchmark here, something we can compare our stocks as we make further progress in our project.

LIST OF FIGURES

INTRODUCTION

In this project, we are going to do a brief timeseries analysis on automobile industry of German stock exchange particularly our focus is specific to the following three companies:

1. Volkswagen
2. BMW
3. Porsche

As this is a timeseries project, time-interval plays important role here and we are observing and analysing four time periods specifically in this this entire project. Those time periods would be as followed:
Start Date- It is period from when we start observing the and analysing the stock prices of the above-mentioned companies.
1st Announcement Date- This is a very important date registered in history of DAX because it is a date when Volkswagen announced that it is going to acquire 49.9% of holdings of the Porsche.
2nd Announcement Date- This is a date when Porsche completely becomes the part of Volkswagen because on this date Volkswagen acquired the remaining 50.1% of holdings of Porsche.
End Date- It is a period of time till where we analyse the market and its affect.

Time series forecasting can be done on the closing price index of the above-mentioned companies, we can try different approaches in forecasting the future possible closing price that the company can incur in the immediate future. Some of the models we can use are as follows:

1. ARCH(q)
2. GARCH(p,q)
3. ARIMA or SARIMAX models

PROBLEM DEFINITION

We try to analyse the data and gather whatever relevant information we can get out of the dataset. We try to look at the data that leading up to the Volkswagen buyout of Porsche. Compare what models describe the stocks best (before and after each acquisition announcement).

DESCRIPTION OF METHOD

METHODOLOGY:

Volatility – When we talk about the volatility, we generally talk about the magnitude of residuals. We just don't care if it is positive or negative, we just care that how far off our predictions are from the reality.

When making long-term huge purchases, investors often seek stability in their investment that is why as long as portfolio rarely experience big shocks they are willing to consider it. And here comes the need of special models that measures volatility.

A positive shock or a negative shock is equally unpleasant since we(investors) are seeking stability.

That is why in finance, volatility is synonymous with the variability/variance.

Thus, measuring volatility is the key when we want safety in our investment.

ARCH– It refers to "Auto-regressive conditional heteroskedasticity". The term heteroskedasticity means "different dispersion" and conditional means "a value that depends on others". The name suggests that we will be dealing with the variance that is depend on the other variance of past periods. This model consists of several equations including μ equation and σ^2- equation. We need several equations in this model to measure unexpected shocks we needed some sort of norm which we can set with the μ-equation without such a base model we don't really have anything to deviate from.

It can be modelled as followed:

$r_t = c + \phi_1 \mu_{t-1} + \varepsilon_t$

where,

$\varepsilon_t =$ The residual values left after estimating the coeff.

$\mu_t =$ A function of past values and past errors which is modeled using ARMAX model

It first fit mean equation to data and then estimate the residuals.
And,

$$\sigma_t^{2} = \alpha_0 + \alpha_1 \varepsilon_{t-1}^{2}$$

where,

$\alpha_0 =$ constant factor

$\alpha_1 =$ coefficient associate with the first term

$\sigma_t^{2} =$ conditional variance

$\varepsilon_{t-1}^{2} =$ squared value of the residual epsilon for the previous period

Based on the epsilon we extract as it measures conditional variance.

GARCH – It refers to "Generalised Auto-regressive conditional heteroskedasticity". It is very similar to ARCH in explanation we can say that adding past values of conditional variances which would serve as a great benchmark for estimations while mathematically it can be represented as follows:

$$V\left(y_t \mid y_{t-1}\right) = \Omega + \alpha_1 \varepsilon_{t-1}^{2} + \beta_1 \sigma_{t-1}^{2}$$

where,

$V\left(y_t \mid y_{t-1}\right) =$ The variance today is conditional on the values of variables yesterday

$\Omega =$ constant

$\alpha_1 =$ Numeric coeff. for the square residual for past period

$\beta_1 =$ Numeric coeff. for the conditional variance from last period

$\varepsilon_{t-1}^{2} =$ Square residual for past period

$\sigma_{t-1}^{2} =$ Conditional variance from last period

No higher-order GARCH models outperform GARCH (1,1) when it comes to variance of market returns.
All the effects of the conditional variance 2 days ago will be contained in the conditional variance of yesterday. We do not need to include more than 1 GARCH component.

SARIMA – Seasonal Autoregressive Integrated Moving Average, SARIMA or Seasonal ARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. It adds three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality. The model equation is as follows:

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D X_t = \Theta(B^s)\theta(B)Z_t$$

Where,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$
$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_P B^P$$
$$\theta(B) = 1 - \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$$
$$\Theta(B) = 1 - \Theta_1 B + \Theta_2 B^2 + \cdots + \Theta_Q B^Q$$
$$\{Z_t\} \sim WN(0, \sigma^2)$$

's' is the period of seasonality, *'d'* is the normal differencing and *'D'* is the seasonal differencing.

Endogenous Variable – The variable we are estimating.

EXPERIMENTAL EVALUATION

DATA COLLECTION:

The data is recorded from the German stock market (There are other markets as well where these companies participate in stock exchange but choosing German stock exchange makes sense because all three companies home country is Germany.). To make sure all the values and all the changes in the values are stock specific we needed a market benchmark that is why we have picked up Volkswagen's greatest competitor in market i.e., BMW (this could be any automobile maker). Data is imported from the yahoo finance.

1. **Start Date**
   The starting date we have selected from which we are going to analyse the data would be **2009-04-05.**

2. **1st Announcement Date**
   The date when Volkswagen made its first announcement to acquire Porsche is **2009-12-09.**

3. **2nd Announcement Date**
   The date when Volkswagen made its second announcement to acquire remaining stocks Porsche is **2012-07-05.**

4. **End Date**
   The end date we have selected from which we are going to analyse the data would be **2014-01-01.**

5. **Returns**
   It is customary in finance to consider the logarithmic return because of the nature of dataset. It is defined as followed:

$$r_t = \log \frac{P_t}{P_{t-1}} = \log P_t - \log P_{t-1}$$

   We will create squared returns as well to examine the volatility in the series.

DATA DESCRIPTION:

Table 2

VARIABLES	DESCRIPTION
date	Column of dates
open	Opening price of stock
high	Highest price of the day
low	Lowest price of the day
close	Closing price of the stock
ret_vol	Calculated return from closing price (Volkswagen)
ret_por	Calculated return from closing price (Porsche)
ret_bmw	Calculated return from closing price (BMW)
sq_vol	Squared return of Volkswagen
sq_por	Squared return of Porsche
sq_bmw	Squared return of BMW
q_vol	Extracting no. of purchases and sales each day of Volkswagen
q_por	Extracting no. of purchases and sales each day of Porsche
q_bmw	Extracting no. of purchases and sales each day of BMW

DATA ANALYSIS:

Figure 1

We see some similarity in the way they move. This indicates the trends of entire automobile industry market. However, Volkswagen prices shift a lot more in magnitude after the third quarter of 2009 to get a better idea. We should separate the different intervals by colour if every period of timeseries is random colour as in the following figure:

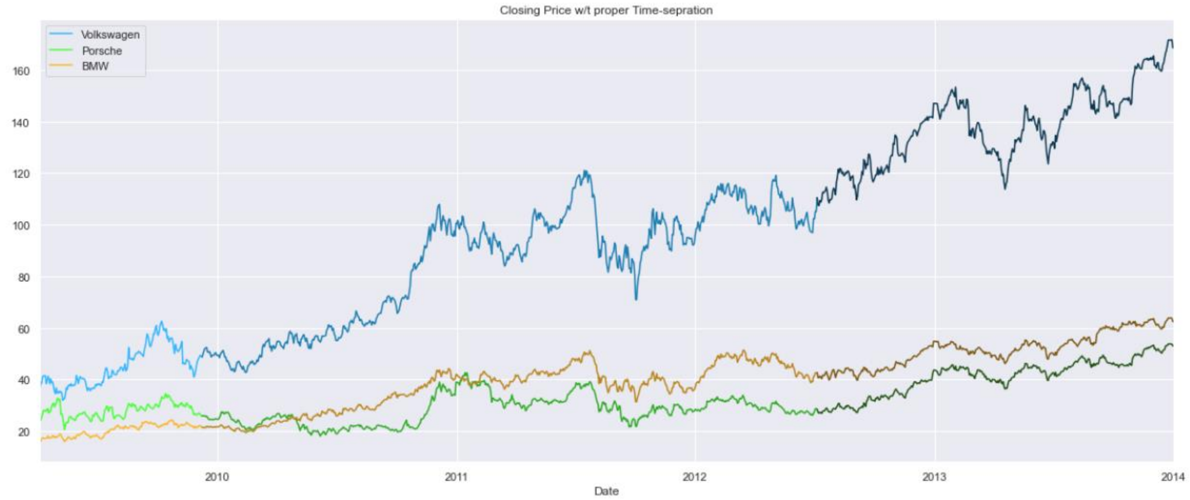

Figure 2

As we can see prior to the first announcement, we see that the two stocks i.e., Volkswagen and Porsche move in a similar fashion. However, Volkswagen number seems to be roughly twice as high afterwards.

Afterwards the gap between the two seems to grow bigger and bigger. If we look at BMW's numbers, we will see that they resemble the Porsche's numbers much closer than the Volkswagen's or at least that is what it looks like.
To get an idea of similarity we are going to perform correlation to get an exactr idea about the growth as followed:
Correlation output (Start date to End Date):

```
Correlation among manufacturers from 2009-04-05 to 2014-01-01

Volkswagen and Porsche correlation:     0.8348050084649201
Volkswagen and BMW correlation:         0.9799384525546676
Porsche and BMW correlation:            0.8039398538655349
```

Figure 3

As the output shows us, we get a higher correlation value between Volkswagen and BMW than the Volkswagen and Porsche. This suggests Volkswagen moves in similar way to the market benchmark.

Correlation output (Start date to 1st Announcement Date):

```
Correlation among manufacturers from 2009-04-05 to 2009-12-09

Volkswagen and Porsche correlation:     0.6633400945227693
Volkswagen and BMW correlation:         0.8428353828601808
Porsche and BMW correlation:            0.6095046037466961
```

Figure 4

From the above output, we see much lower correlation all around with the Porsche and BMW is very low. This suggests the stock prices for these brands weren't too similar before the start of buyout. The possible explanation could be Porsche's evaluation becomes heavily tied to that of Volkswagen becomes market trendsetter so all other companies have to adjust to it.
In that way Volkswagen affect both the Porsche and BMW simultaneously. One directly and other is indirectly. Then the pull of Volkswagen in times of boom makes it so that Porsche and BMW appear to change in a similar manner without a direct causal link between them.
This is the case because Porsche and BMW correlation keeps increasing as Volkswagen grows as a market leader. Interdependence of Volkswagen and Porsche may increase over time as Volkswagen continue to buy their shares.

Correlation output (1st Announcement Date to 2nd Announcement Date):

```
Correlation among manufacturers from 2009-12-09 to 2012-07-05

Volkswagen and Porsche correlation:     0.7422114342529975
Volkswagen and BMW correlation:         0.9795943008992956
Porsche and BMW correlation:            0.7035985533162737
```

Figure 5

The new result shows increase among all correlations but interestingly we can see very high correlation between Volkswagen and BMW.
If the we look at the Figure 1 we can see that over that time period Volkswagen stocks grow in value several time faster than the market benchmark. However, since the market moves basically the same way Volkswagen becomes the sort of trend-setter after the partial takeover of the Porsche.

Correlation output (2nd Announcement Date to End Date):

```
Correlation among manufacturers from 2012-07-05 to 2014-01-01

Volkswagen and Porsche correlation:     0.9405237308477348
Volkswagen and BMW correlation:         0.9284447459439166
Porsche and BMW correlation:            0.9494111766213098
```

Figure 6

Now, this time we see higher correlation between Volkswagen and Porsche than Volkswagen and BMW. This shows that the two being recognised as a single entity and the prices of one directly affect the other. We can see higher correlation between BMW and Porsche interestingly they do not seem to have any obvious common link between them. Volkswagen continues to be a market trend-setter these correlations should all remain high.

Correlation output (End Date to current date):

```
Correlation among manufacturers from 2014-01-01 to 2020-12-23 00:00:00

Volkswagen and Porsche correlation:     0.9392810285264768
Volkswagen and BMW correlation:         0.5244440547982339
Porsche and BMW correlation:            0.47523916002767586
```

Figure 7

If there is another boom in the Volkswagen's market then it will again become the trend-setter.

As we can see Volkswagen's trend change after the 2nd announcement due to change in trend there is a need of shift in new model is required.

Now, we will try to fit the best model via auto-ARIMA method where exogenous variable taken as Porsche and BMW.

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	178
Model:	SARIMAX(1, 0, 0)	**Log Likelihood**	-323.195
Date:	Sat, 26 Dec 2020	**AIC**	656.390
Time:	01:22:36	**BIC**	672.299
Sample:	04-06-2009	**HQIC**	662.841
	- 12-09-2009		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.7236	0.447	1.618	0.106	-0.153	1.600
por	0.4179	0.100	4.169	0.000	0.221	0.614
bmw	0.6615	0.314	2.108	0.035	0.046	1.277
ar.L1	0.9634	0.020	48.675	0.000	0.925	1.002
sigma2	2.1787	0.148	14.718	0.000	1.889	2.469

Ljung-Box (L1) (Q):	0.77	**Jarque-Bera (JB):**	166.09
Prob(Q):	0.38	**Prob(JB):**	0.00
Heteroskedasticity (H):	1.61	**Skew:**	-0.52
Prob(H) (two-sided):	0.07	**Kurtosis:**	7.62

Figure 8

So, the period prior to 49.9% announcement. We can see that an AR(1) with the two exogenous variables is the best fit.

All coefficients apart from the intercept are significant. So, this is an good fit.

We see the estimate to the AR(1) value is very close to 1 this means we are sticking really close to the value of the last period with very little deviation. This is not much surprising since we are using prices that do not oscillates completely randomly.

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	672
Model:	SARIMAX(1, 1, 1)	**Log Likelihood**	-550.899
Date:	Wed, 30 Dec 2020	**AIC**	1111.797
Time:	08:07:04	**BIC**	1134.341
Sample:	12-09-2009	**HQIC**	1120.529
	- 07-05-2012		
Covariance Type:	opg		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
vol	0.1704	0.015	11.128	0.000	0.140	0.200
bmw	0.3030	0.042	7.206	0.000	0.221	0.385
ar.L1	0.6598	0.229	2.883	0.004	0.211	1.108
ma.L1	-0.5910	0.245	-2.412	0.016	-1.071	-0.111
sigma2	0.3024	0.009	32.415	0.000	0.284	0.321

Figure 9

After 1st announcement, ARIMAX (1,1,1) is the optimal choice. This is the really big shift.

Possible reasons could be:

1. The optimal model is integrated one. Since, the auto-arima enforces stationarity we know that the new dataset is certainly non-stationary.
2. The model finds past residuals to have significant explanatory power. This indicates that the real-life announcement actually changes the trend for Volkswagen.

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	657
Model:	SARIMAX(0, 1, 0)	**Log Likelihood**	-1064.147
Date:	Wed, 30 Dec 2020	**AIC**	2134.294
Time:	08:38:28	**BIC**	2147.752
Sample:	0	**HQIC**	2139.512
	- 657		
Covariance Type:	opg		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
por	0.8277	0.066	12.546	0.000	0.698	0.957
bmw	1.4806	0.071	20.791	0.000	1.341	1.620
sigma2	1.5015	0.063	23.904	0.000	1.378	1.625

Figure 10

Moving onto the last interval, we see that the model has changed completely on more time, ARIMAX (0,1,0) once again recognised the non-stationary nature of stock prices. But we are no longer relying on past values. Instead of that we can see a huge shift towards the coefficient for Porsche prices.

I can presume this shift because of the two automobile manufacturers are now a single entity. Due to this reason price of Porsche today will be more accurate estimator than the prices of Volkswagen yesterday.

Hence, I can say that new information seems to have a higher impact on trends compare to past patterns.

<u>Fitting model for Porsche:</u>

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	178
Model:	SARIMAX(2, 0, 0)	**Log Likelihood**	-218.214
Date:	Sat, 26 Dec 2020	**AIC**	448.427
Time:	01:23:05	**BIC**	467.518
Sample:	04-06-2009	**HQIC**	456.169
	- 12-09-2009		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.5516	0.360	1.533	0.125	-0.154	1.257
vol	0.1101	0.051	2.157	0.031	0.010	0.210
bmw	0.8068	0.145	5.557	0.000	0.522	1.091
ar.L1	1.0747	0.059	18.194	0.000	0.959	1.191
ar.L2	-0.1589	0.059	-2.716	0.007	-0.274	-0.044
sigma2	0.6721	0.048	14.098	0.000	0.579	0.766

Ljung-Box (L1) (Q):	0.00	**Jarque-Bera (JB):**	559.43
Prob(Q):	0.98	**Prob(JB):**	0.00
Heteroskedasticity (H):	0.21	**Skew:**	-0.96
Prob(H) (two-sided):	0.00	**Kurtosis:**	11.47

Figure 11

Before 1st announcement,
AR(2) with two exogenous is best model.
It is similar to Volkswagen over the same period.
After 1st announcement,
We see a shift towards integrated model ARIMA (1,1,0). This directly translates into a roughly one point drop in the AR coefficient from 1.08 to 0.07.
This is not too odd, since we are using integrated value here however an additional explanation comes from the rise of the Volkswagen coefficient which now has a bigger impact. The direct jump of 0.12 to 0.19. This is actually more than 50% growth.

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	390
Model:	SARIMAX(0, 1, 0)x(0, 0, [1], 5)	**Log Likelihood**	-192.826
Date:	Wed, 30 Dec 2020	**AIC**	393.653
Time:	08:07:05	**BIC**	409.507
Sample:	07-05-2012	**HQIC**	399.938
	- 01-01-2014		
Covariance Type:	opg		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
vol	0.2119	0.013	16.876	0.000	0.187	0.237
bmw	0.2286	0.038	6.036	0.000	0.154	0.303
ma.S.L5	-0.1249	0.057	-2.210	0.027	-0.236	-0.014
sigma2	0.1578	0.006	27.851	0.000	0.147	0.169

Figure 12

This time model discovers seasonal pattern. These trends are affected more by the current events rather than the pre-existing patterns.

NOTE- I have not used ACF & PACF plot to identify the parameters of the functions required to compute because I have used auto_arima command of python which automatically finds the best fit amongst all parameters.

VOLATILITY

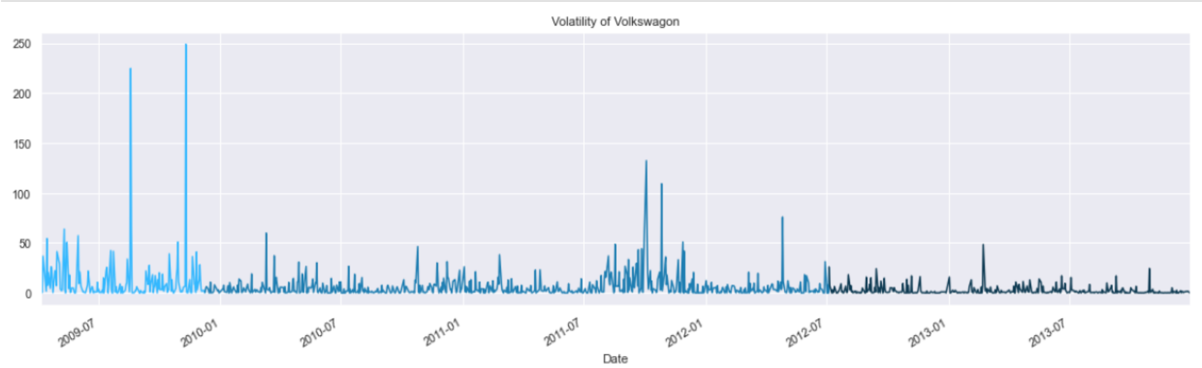

Figure 13

We expect some major shift surrounding such events. We see VW (volkswagen) has the highest volatility before any of the announcements.
The stocks become more stable following each of the announcements. While it exhibits instability in the time leading upto each purchase such behaviour can most certainly be attributed to the rumours spread across the market about the deal on the horizon.
However, Porsche had tried to purchase VW a few years back many people weren't certain how it will all play out.
Now, to measure volatility we need to fit GARCH model and ARCH model.

Figure 14

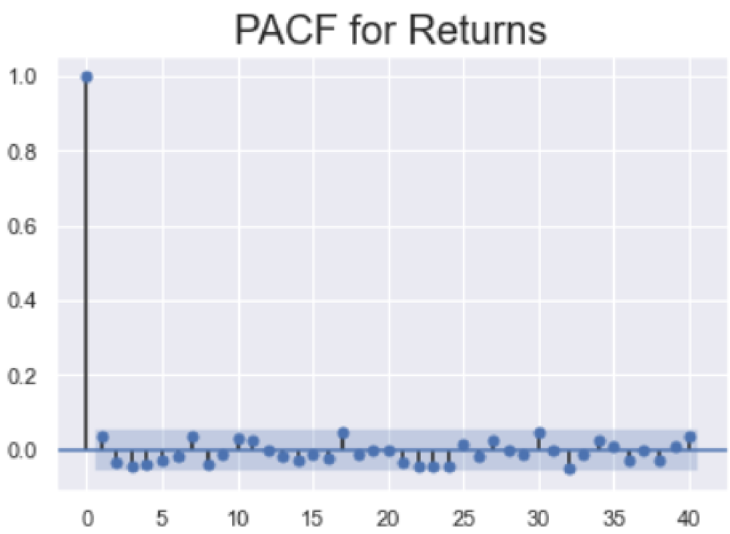

Figure 15

Simple GARCH (1,1) should be the best fit for the returns.
<u>GARCH (1,1) for constant mean. {Start date to ann_1}:</u>

Constant Mean - GARCH Model Results

Dep. Variable:	ret_vol	**R-squared:**	-0.000
Mean Model:	Constant Mean	**Adj. R-squared:**	-0.000
Vol Model:	GARCH	**Log-Likelihood:**	-474.991
Distribution:	Normal	**AIC:**	957.983
Method:	Maximum Likelihood	**BIC:**	970.710
		No. Observations:	178
Date:	Wed, Dec 30 2020	**Df Residuals:**	174
Time:	08:07:06	**Df Model:**	4

Mean Model

	coef	**std err**	**t**	**P>|t|**	**95.0% Conf. Int.**
mu	0.1571	0.266	0.590	0.556	[-0.365, 0.679]

Volatility Model

	coef	**std err**	**t**	**P>|t|**	**95.0% Conf. Int.**
omega	10.6194	4.227	2.512	1.200e-02	[2.334, 18.905]
alpha[1]	0.1540	9.197e-02	1.675	9.400e-02	[-2.624e-02, 0.334]
beta[1]	4.7917e-13	0.349	1.372e-12	1.000	[-0.684, 0.684]

Figure 16

p-value=1 for β shows that the trends in variance aren't as persistent as we would expect. So, there is no need for a GARCH component. We can try ARCH model for this one.

<u>GARCH (1,1) for constant mean. {ann_1 to ann_2}:</u>

Constant Mean - GARCH Model Results

Dep. Variable:	ret_vol	**R-squared:**	-0.000
Mean Model:	Constant Mean	**Adj. R-squared:**	-0.000
Vol Model:	GARCH	**Log-Likelihood:**	-1526.73
Distribution:	Normal	**AIC:**	3061.46
Method:	Maximum Likelihood	**BIC:**	3079.50
		No. Observations:	672
Date:	Wed, Dec 30 2020	**Df Residuals:**	668
Time:	08:07:06	**Df Model:**	4

Mean Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
mu	0.1892	8.634e-02	2.191	2.843e-02	[1.998e-02, 0.358]

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	0.1679	7.914e-02	2.122	3.388e-02	[1.279e-02, 0.323]
alpha[1]	0.0688	1.682e-02	4.091	4.302e-05	[3.585e-02, 0.102]
beta[1]	0.9040	2.108e-02	42.883	0.000	[0.863, 0.945]

Figure 17

p-value=0 for β shows that the autocorrelation in the conditional variance is significant and dies off based on the coefficient value. The significant constant μ indicates that there is some constant trend in the return values. Additionally, the significant coefficient for omega suggests the exact same feature can be found in we never expect returns or their volatility to ever be perfectly stable which is normal for a market lacking efficiency.

<u>GARCH (1,1) for constant mean. {ann_2 to end date}:</u>

Constant Mean - GARCH Model Results

Dep. Variable:	ret_vol	**R-squared:**	-0.001
Mean Model:	Constant Mean	**Adj. R-squared:**	-0.001
Vol Model:	GARCH	**Log-Likelihood:**	-724.558
Distribution:	Normal	**AIC:**	1457.12
Method:	Maximum Likelihood	**BIC:**	1472.98
		No. Observations:	390
Date:	Wed, Dec 30 2020	**Df Residuals:**	386
Time:	08:07:06	**Df Model:**	4

Mean Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
mu	0.2298	9.845e-02	2.334	1.958e-02	[3.685e-02, 0.423]

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	0.7719	0.677	1.141	0.254	[-0.554, 2.098]
alpha[1]	0.1853	0.145	1.273	0.203	[-9.991e-02, 0.470]
beta[1]	0.5136	0.331	1.550	0.121	[-0.136, 1.163]

Figure 18

Non-significant coefficient for constant omega and alpha 1 shows that there is some autocorrelation in volatility which fits the volatility clustering feature. Since, the entire interval is much more stable even the beta coefficient is dropped by roughly 1/3rd.

CONCLUSION

- After all the models fitting, we can say at some place GARCH is better where as at some places it is not, we can go for improvement in model always.
- Overall, I see really different volatility trends in all three periods which is quite intriguing.
- The purchase of Porsche has destroyed many rumours in the market and provided Volkswagen with a larger share of market.
- In return of larger share of market it has resulted in much lower volatility after each announcement making Volkswagen stocks a more appealing prospect.