**Selecting Good Volatility Models:**

**Forecasting Volatility of S & P 500**

Eric Liu

Zhili.liu@baruchmail.cuny.edu

Baruch College

May 12, 2021

**Abstract**

Can volatility be predictable? The simple answer is yes. While there are numerous factors (uncertainties) that influence volatility, volatility can be expected to behave in some certain ways over an extended period of time and we can make well-informed predictions. Volatility clustering can be captured by different time series models and these models enable us to make forecasts. However, particular models have advantages and limitations. In this paper, we outline some certain stylized behaviors of volatility. We further explore the characteristics of each model as applied to S&P 500 and select good models to forecast the volatility of the market.

## Part 1: Introduction

Stock market volatility is an essential issue in the financial theory for the industry practitioners and investors to price and allocate assets and manage risks. Volatility is defined as the standard deviation (or square root of variance) of an asset price's returns over a certain period of time. A high price volatility often indicates higher risk, and a low price volatility usually means the risk is relatively low. Understanding the behavior of volatility helps industry practitioners take advantage of investment opportunities. A portfolio manager, for example, may want to sell or rebalance a portfolio if he/she predicts the market will become highly volatile. A good volatility model is a critical tool for this portfolio manager to make educated guesses on the future volatility of the market.

There is extensive research to examine and inspect different volatility models. In this paper, we will test and discuss the advantages and disadvantages of several time series volatility models, namely: moving average, exponential moving average, GARCH and GJR-GARCH, and will select well-fitted models to make predictions.

A good volatility model must be able to capture volatility clustering and accurately reflect its conventionalized behaviors. Before going further, we will sketch five well-established Stylized facts (Engle and Patton, 2000). Volatility:

- Has heavy tail probabilities
- Is persistent
- Is serially correlated (mean reverting)
- May have an asymmetric effect
- May be influenced by exogenous variables.

Different from the classic assumptions of the Ordinary Least Squares (OLS) Linear Regression, such that the error term of the observations has homoscedastic (unconditional) variance and mean of zero, one of the well-researched stylized facts of volatility is that it has heavy tails. In this scenario, a normal distribution is not a good assumption for the distribution of an asset returns. Typically, the distribution has excess kurtosis, the mean of the error term is not zero and its variance is heteroscedastic (conditional). Later in this paper, we will give an empirical example to examine the normality of the distribution of S&P 500.

Volatility is expected to be persistent. Oppose to unconditional variance which tends to be infinite, conditional variance is not consistent and finite. That being said, it is legitimate to assume that tomorrow's volatility will be impacted by the volatility today. If volatility today is high, we can reasonably expect volatility tomorrow will continue to be high, vice versa. A good volatility model must reflect this characteristic of persistence.

It is reasonable to expect volatility will be mean-reverting. Asset returns are anticipated to have a normal level of volatility and it should be mean reverting. Although today's volatility will heavily influence volatility tomorrow, eventually, volatility should go back to its normal level. Typically, a time period of high volatility will be followed with a downward trend and a time of low volatility will rise to its mean level.

Volatility clustering may have an asymmetric effect, positive and negative shocks are unlikely to impact volatility symmetrically. It is also widely accepted that negative shocks generally have heavier effect compared to positive shocks. For example, if an asset price falls by the amount of *k*, it is expected that volatility will increase more compared to a scenario where the asset price increases by a similar amount (Kenton, 2022). Asymmetric affect is also called leverage effect.

Finally, we cannot ignore the influence asserted by exogenous variables. An exogenous variable is a relevant external factor that is not directly incorporated into a volatility model but still are used as weighted input that would affect volatility. Generally, there are three types of external drives (Fidelity, n.d.) that are highly influential to volatility clustering:

- Political and economic factors
- Industry and sector factors
- Company performance.

The role of government has a wide range of influence on stock prices. It asserts influences from regulations, legislation, to trade agreements, and so forth. Similarly, economic data, such as GDP, inflation data, unemployment rate, that also could drive the changes of stock prices.

Certain events are likely to increase or decrease volatility in the relevant industries. For example, if a country where is a major wheat producer is (unfortunately) at war and that causes shortage of wheat, the wheat related stocks will naturally become more volatile.

Company performance, for instance, earnings announcements, can also push its volatility up and down. Negative news tends to lead to negative reactions from investors, vice versa. Amazon stocks dropped 14% after it announced its first quarter earnings on April 29th that fell short of the Wall Street estimates (Semenova, 2022).

A good volatility model should reflect all above characteristics of volatility. In the paper, we will present the daily returns of S&P 500 to illustrate the stated points and analyze a few volatility models.

## Part 2. An Empirical Example of S&P 500

### 2.1. Test Normality

The example we provide consist of 8148 observations of the daily close prices from Yahoo Finance of S&P 500 for the period of January 1st, 1990, to May 4th, 2022. We used this to calculate daily returns. The following table (Figure 1) gives a quick statistical overview of the of the log difference of the market returns.

```
                     RET
nobs         8148.000000
NAs             0.000000
Minimum       -11.984055
Maximum        11.580037
1. Quartile    -0.441485
3. Quartile     0.567819
Mean            0.036993
Median          0.058821
Sum           301.416241
SE Mean         0.012655
LCL Mean        0.012185
UCL Mean        0.061801
```

```
Variance       1.304984
Stdev          1.142359
Skewness      -0.188236
Kurtosis      10.875554
```

Figure 1. Statistical Summary of S&P 500 daily returns

Is the distribution normal? We should be able to receive a quick conclusion of normality from the above statistical summary. As mentioned earlier in the paper, the mean of the error terms of a normal distribution is zero. In our example, the mean is 0.0370, and the median is 0.0588. The mean is not zero and is smaller than the median. It shows that distribution is not symmetric.

We can further look at the skewness and kurtosis. The normal distribution has skewness of zero and kurtosis of three. By contrast, the data set has a negative skewness (-0.188236) and excess kurtosis (7.88). These indicated that the distribution has longer left tail. This point can be echoed by the minimum and maximum returns. The largest daily loss is 11.9841% and the largest daily gain is 11.5800%. The largest loss is greater than the largest gain, it proves the distribution is asymmetric and its left tail is longer. The following graph (Figure 2) illustrates the shape of the best fit normal distribution (the red line) and the actual market returns distribution (the histogram). Visual inspection confirms that the distribution is not normal.

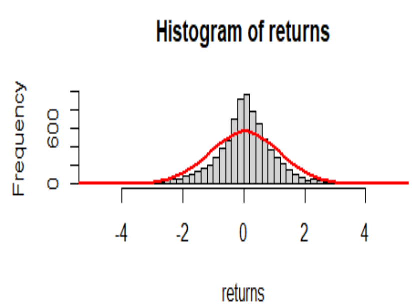


Figure 2. Distribution of S&P 500 daily returns

To be more systematic, we can use statistical test. In our examination, we used R-Studio to perform a *Kolmogorov-Smirnov test* (see below table). The P-value we received is 2.2e-16 which is less than 0.025 (two-sided test) at 5% significance level. The result shows the distribution is not normal. (Alternatively, a *Shapiro-Wilk test* can be performed if the number of observations is between three to five thousand. Our data set contained 8148 observations.)

```
data:  returns
D = 0.99949, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Figure 3. Kolmogorov-Smirnov test on S&P 500 Daily Returns

## 2.2. Test Autocorrelation and Stationarity

Autocorrelation can be visually detected by making an Autocorrelation Function (ACF) plot, a method to plot the correlation coefficient against the time lags. In Figure 4, we plotted the log returns and absolute returns of the market. Both ACF plots show a strong indication of autocorrelation of the returns. The bars reach both blue dash lines on a regular basis even up one hundred time lags in the log return ACF plot; the bars keep exceeding the positive blue dash line in the absolute return ACF plot. We can also see a pattern of trend as the bars in the right graph graduate deteriorates towards to one direction.

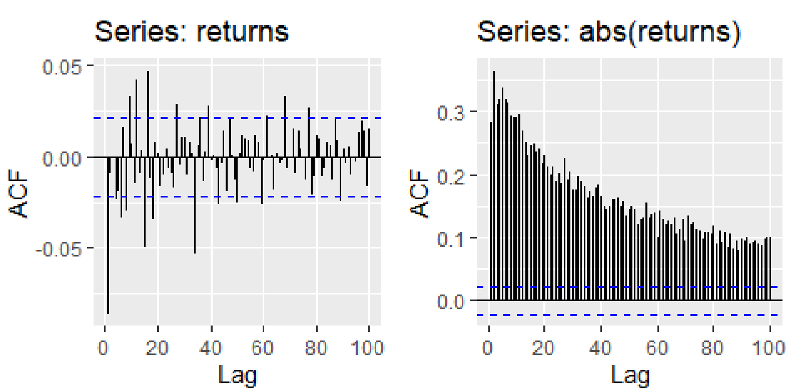


Figure 4. ACF Plot Log-Difference (Left) and Absolute Returns (Right)

Although volatility has a conditional variance due to the exogenous variables, it is expected to be stationary process (mean-reverting). We can perform an Augmented Dickey-Fuller (ADF) test with the percentage returns to confirm this hypothesis of stationarity:

$$H_O: \emptyset = 1 \ \rightarrow Non - Stationary$$

$$H_1: \emptyset < 1 \ \rightarrow Stationary$$

We let BIC to select the optimal lags and received the following results:

| | t value | Pr(>\|t\|) |
|---|---|---|
| Intercept | 1.412 | 0.158 |
| trend | 0.244 | 0.807 |
| lag(RET) | -67.609 | <2e-16 *** |
| lag(diff(RET), lag.max = 25, Criterion = "BIC") | -1.525 | 0.127 |

At 1% significance level, the rejection value with a trend is -3.96, the ADF test value (-67.609) which is significantly smaller than the rejection value. Therefore, we reject the $H_O$ and conclude that the percentage of returns of S&P 500 is stationary, as expected.

**Part 3. Select Good Volatility Models**

Having illustrated the stylized facts of volatility, we now examine some volatility models and select models that are able to reflect these factors.

3.1. Moving Average and Exponential Moving Average

Volatility models, such as moving average (MA), exponential moving average (EMA) or even plotting the absolute returns, are able to capture the clustering and help industry practitioners to price their assets (Figure 5 and 6).

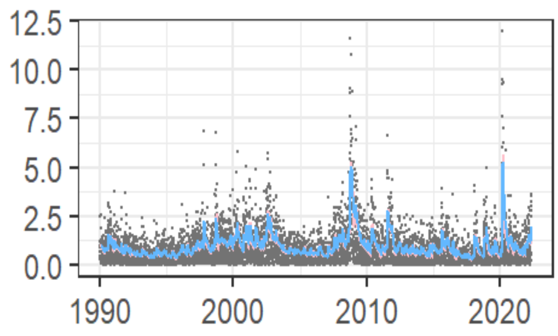


Figure 5. Capture volatility with Abs Returns, MA and EMA

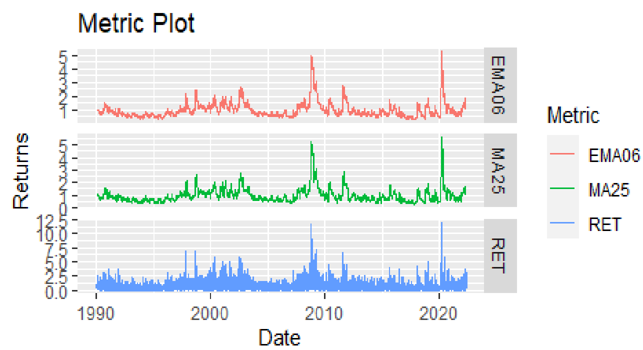


Figure 6. Metric Plot with different Y-Axis

In Figure 5, the grey dots are the absolute returns. It shows volatility clustering spikes up then goes down on a roughly regular basis. Observing the absolute returns is a simple approach to measure volatility as a proxy. In Figure 5, the absolute returns, compared to MA (pink line) and EMA (blue line), show larger clustering. The same conclusion can be observed in figure 6, where we plot the three methods with different y-axis. The spikes of the absolute return (blue) go up to 12.5%. MA and EMA have sigma spikes peak around 5%. The MA and EMA approaches can be modeled as follow:

$$\boldsymbol{MA} \rightarrow \sigma_{t+1}^2 = \frac{1}{M} * \sum_{M=1}^{M} R_{t-M+1}^2 \tag{1}$$

Where $R = asset\ returns$ , $M = obeservation\ periods$

$$\boldsymbol{EMA} \rightarrow \sigma_{t+1}^2 = \lambda * \sum_{M=0}^{\infty} (1-\lambda)^M * R_{t-M}^2 \tag{2}$$

Where $\lambda = smoothing\ prameter$

Moving Average is a simple and straightforward estimation of the variance taking the squared return over a certain period of time of M. Typically, industry practitioners observe the last 25 days as the time window ($M = 25$). In model (1), each observation has same amount of impact to predict the volatility tomorrow because observation period is modeled as $\frac{1}{M}$. In other words, the observation in the last 25th day influences the volatility tomorrow as much as today. However, one of the stylized facts of volatility is persistence. Each day of observation is unlikely to have the same weight of impact. The MA model may not be accurate because it gives too much weight to the old observation (Maverick, 2021).

An improved model is to called Exponential Moving Average. EMA introduces a smoothing parameter $\lambda$ to smooth out the input of each observation's weight. Similar to MA, EMA uses the squared returns as input. If the number of observations is infinite, the sum of each observation's weight should totals one. However, if the sum of weight becomes one, then volatility is non-stationary. This assumption violates the stylized fact that volatility should be meaning reverting. Industry practitioners commonly take the value $M = 25$ and $\lambda = 0.06$, the total weight of the 25-day time periods is about 74%. The weight of each data point gradually decays as the observation is further apart from today's observed value as indicated in below table (Figure 7).

| Day (m) | λ (a) | (1-λ)^m (b) | Weight (=a*b) | Day (m) | λ (a) | (1-λ)^m (b) | Weight (=a*b) |
|---|---|---|---|---|---|---|---|
| **1** | 0.06 | 0.9400 | 0.0564 | **14** | 0.06 | 0.4205 | 0.0252 |
| **2** | 0.06 | 0.8836 | 0.0530 | **15** | 0.06 | 0.3953 | 0.0237 |
| **3** | 0.06 | 0.8306 | 0.0498 | **16** | 0.06 | 0.3716 | 0.0223 |
| **4** | 0.06 | 0.7807 | 0.0468 | **17** | 0.06 | 0.3493 | 0.0210 |
| **5** | 0.06 | 0.7339 | 0.0440 | **18** | 0.06 | 0.3283 | 0.0197 |
| **6** | 0.06 | 0.6899 | 0.0414 | **19** | 0.06 | 0.3086 | 0.0185 |
| **7** | 0.06 | 0.6485 | 0.0389 | **20** | 0.06 | 0.2901 | 0.0174 |
| **8** | 0.06 | 0.6096 | 0.0366 | **21** | 0.06 | 0.2727 | 0.0164 |
| **9** | 0.06 | 0.5730 | 0.0344 | **24** | 0.06 | 0.2265 | 0.0136 |
| **10** | 0.06 | 0.5386 | 0.0323 | **23** | 0.06 | 0.2410 | 0.0145 |
| **11** | 0.06 | 0.5063 | 0.0304 | **24** | 0.06 | 0.2265 | 0.0136 |
| **12** | 0.06 | 0.4759 | 0.0286 | **25** | 0.06 | 0.2129 | 0.0128 |

| Day (m) | λ (a) | (1-λ)^m (b) | Weight (=a*b) | Day (m) | λ (a) | (1-λ)^m (b) | Weight (=a*b) |
|---|---|---|---|---|---|---|---|
| **13** | 0.06 | 0.4474 | 0.0268 | **Total Weight** | | | **0.7381** |

Figure 7. Weight of **Each Observation in a 25**-Day Window

Compared to the MA model, the EMA should increase the accuracy of predictions because it allows us to distribute the weight of each data point more precisely over a selected time frame (Treloar, n.d.). Does it mean EMA is a good volatility model? Since volatility is the squared root of the conditional variance, the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is deemed to be more superior and preferred when concerning the stylized fact of uneven variance.

3.2. GARCH and GJR-GARCH

GARCH model is a measure that takes the past time-varying volatility (standard deviation) of an asset return into account. It is able to echo more of the stylized facts, mean-reverting and heteroskedasticity variance in particular. GARCH presumes that the variance is conditional, the assumption is also called ARCH effect. GARCH (1,1), one lag of the error term and one lag of the conditional variance, is relatively easy to use and delivers reliable estimates. The model is given by (3) and the mean conditional variance is given by (3.1).

$$\boldsymbol{GARCH\ (1,1)} \rightarrow \sigma_{t+1}^2 = \omega + \alpha * R_t^2 + \beta * \sigma_t^2 \qquad (3)$$

Where $\omega$, $\alpha$ and $\beta$ are three parameters that need to be estimated.

$$\boldsymbol{Mean\ Variance} \rightarrow \ \sigma_t^2 = \frac{\omega}{1-(\alpha+\beta)} \qquad (3.1)$$

ARCH effect can be tested by observing the sum of $\alpha$ and $\beta$. GARCH assumes the value of the sum is between zero to 1. When it is closer to zero, it is indicative of no ARCH effect. In which the variance is more finite and homoscedastic. When the summed value is closer to one, it is an evidence of ARCH effect, in which the variance is conditional. In this study, we used the *rugarch package* in R-Studio, and the estimates of the S&P 500 returns came back as followed:

```
   mu    ar1  omega alpha1  beta1
 0.063 -0.026  0.019  0.108  0.877
```

From the estimates, we can conclude that there is an ARCH effect in the time series data. $\alpha + \beta = 0.108 + 0.877 \approx 0.99$. The summed value is very close to one, which shows the characteristic of heteroscedasticity of the variance of the market returns.

One shortcoming of the GARCH model is that the model disregards the stylized fact of the asymmetric effect as we discussed earlier in this study. Positive news and negative news influence asset returns asymmetrically, this fact is missed in the GARCH estimation. Hence, being able to recognize positive or negative shocks is an essential tool to better perform forecasts on volatility. Glosten-Jagannathan-Runkle (GJR)-GARCH, is an extended form of GARCH by adding an interactive term to take positive and negative shocks into consideration. The model is given by:

$$\boldsymbol{GJR - GARCH} \rightarrow \sigma_{t+1}^2 = \omega + \alpha * R_t^2 + \gamma_1 * R_t^2 * I(I = 1 | R_t^2 < 0) + \beta * \sigma_t^2 \quad (4)$$

$$Where\ I = Interactive\ term. When \begin{cases} I = 1 | R_t^2 < 0 \rightarrow negative\ shocks; \\ I = 0 | R_t^2 > 0 \rightarrow positive\ shocks. \end{cases}$$

When there is a negative shock, the interactive term $I = 1$. The interactive term is included in the model (4), and predicted volatility will increase. When there is a positive shock, $I = 0$. The estimation of volatility is the same as model (3). We can perform a hypothesis test on $\gamma_1$ to detect if there is an asymmetric affect.

$$H_O: \gamma_1 = 0 \ \rightarrow no\ asymmetric\ effect$$

$$H_1: \gamma_1 \neq 0 \ \rightarrow \ asymmetric\ effect$$

We used the S&P 500 returns with the ugarchspec function, and the results came back as below:

| | Estimate | T-value |
|---|---|---|
| Mu | 0.0509788 | 4.013847 |
| ar1 | -0.0358972 | -1.826201 |
| Omega | 0.0384646 | 7.514080 |
| alpha1 | 0.0355420 | 2.912133 |
| beta1 | 0.8101167 | 48.702062 |
| gamma1 | 0.2348100 | 8.394250 |

Figure 8. Estimates of GJR-GARCH

The T-value of $\gamma_1$ is 8.39. Whether at 1% significance level ($T_\alpha = 2.33$) or 0.1% significance level ($T_\alpha = 3.09$), we reject the null hypothesis. This indicates that there is an asymmetric effect on the S&P 500 returns.

Both GARCH or GJR-GARCH are viewed more sophisticated volatility models compared to the MA and EMA. They are not only able to capture the clustering but also deliver reliable predictions. The following graphs (figure 9 and 10) are the S&P 500 volatility from 1990 to present that captured by GARCH and GJR-GARCH. During the more volatile times such as the times around 2008, 2013, 2016 and 2020, the $\sigma$ captured by GJR-GARCH is slightly higher compared to GARCH because of the asymmetric effect.

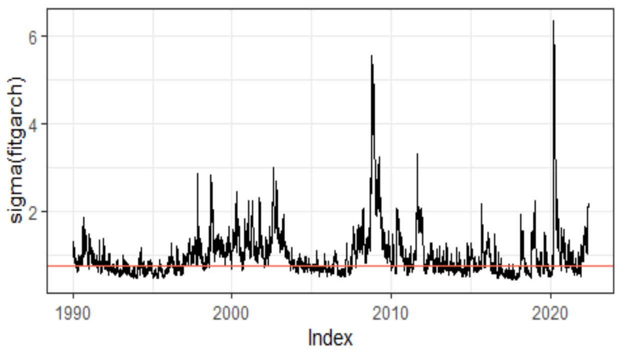


Figure 9. S&P volatility 500 GARCH

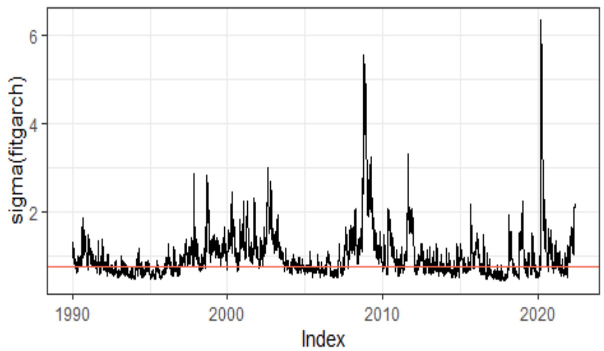


Figure 10. S&P 500 volatility GJR-GARCH

**Part 4. Predict Volatility of S&P 500**

In this section, we make forecast the S&P 500 volatility for the next 250 days (roughly one trading year) with both GARCH and GJR-GARCH model and compared the difference. The red line is the prediction of GARCH, and the blue line represents GJR-GARCH prediction (see figure 11)

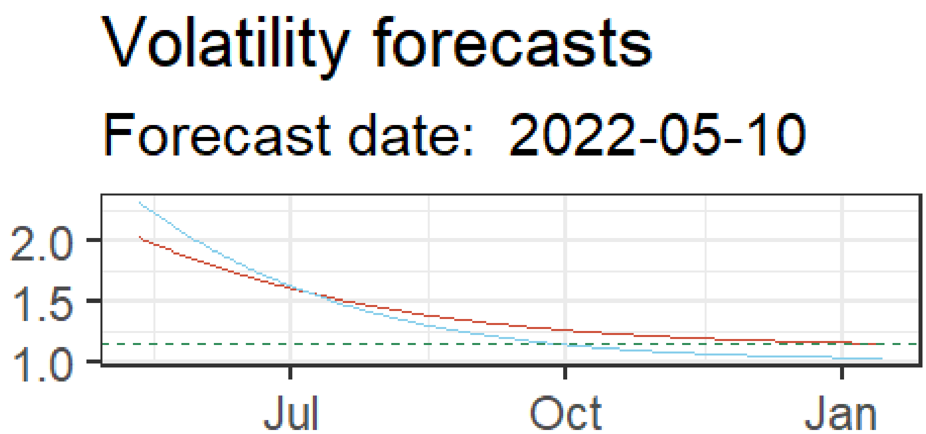


Figure 11: Forecast of S&P 500 Volatility

Both methods forecast volatility in condition of the information that is available today. Based on that, the forecast made by GJR-GARCH predicts the volatility stay higher than the GARCH estimation but will go down more over the course of 250 days. As discussed above, GJR-GARCH includes an interactive term that is able to recognize the asymmetric effect. When the market price drops, return is negative, the model will returns larger volatility to the estimation.

**Part 5. Conclusion.**

In this paper, we summarized the dynamic of volatility, including introduced its stylized facts and the possible external drivers. We captured the clustering with absolute returns, MA, EMA, and two GARCH family methods. We discussed compared each model and explored the limitations. We deemed that GARCH and GJR-GARCH are better volatility models and made forecasts of S&P 500 with these two models.

The Moving Average model is simple to use, however, it cannot make accurate predictions since it gives too much weight to the older data. Although the Exponential Moving Average added a smoothing parameter to better allocate the input of weights, GARCH is generally a preferred approach to use. While GARCH better models the characteristics of volatility, it misses the recognition of asymmetric effect. GJR-GARCH added an interactive term to fix that. As a result, the GJR-GARCH forecast captures higher volatility when the market is more volatile.

Reference

Engle, R., and Patton, A. (2000, October). What good is a volatility model. *NYU Stern School of Business.* 1 (2001) 237–245

Kenton, W. (2022, January 05). Asymmetric Volatility Phenomenon (AVP). *Investopedia*.

“Understanding Stock Market Volatility”. (n.d.). Understanding Stock Market Volatility and How It Could Help you? *Fidelity, Fidelity International*. No.: 199006300E

Semenova, A. (2022, April 29th). Amazon Stock Plunges 14% in Biggest One-Day Drop Since 2006. *Yahoo Finance*.

Maverick, J., B. (2021, October 08). Advantages and Disadvantages of the Simple Moving Average (SMA). *Investopedia.*

Treloar, T. (n.d.). A closer Look At A Classic - Examining The EMA. *Fidelity* V. 26:6 (42-46)