Department of Statistics and Actuarial-Financial Mathematics
Master's Thesis. University of the Aegean

Sentiment Analysis combined with Time Series Analysis forecasting

Perperidou Ioanna
MSC in Statistics and Actuarial-Financial Mathematics

Supervisors:

dr. Xanthopoulos Stylianos	University of the Aegean.	*xanthos@aegean.gr*
dr. Bakeroudis Stavros	University of the Aegean	*svakeroudis@aegean.gr*
dr. Xalidias Nikolaos	University of the Aegean	*nick@aegean.gr*

Samos, September 2020

“I certify that the submitted work is my own work based on my personal study and research and that I have acknowledged all material and sources used in its preparation, whether they be books, articles, reports, lecture notes, and any other kind of document, electronic or personal communication. I also certify that this thesis has not previously been submitted for assessment in any other unit, or at any other time in this unit.” - Perperidou Ioanna

Abstract

One of the most difficult aspects regarding daily trading is knowing when the market will rise and fall. Search terms can help give us some insight into predicting those periods so you can profit in both bullish and bearish markets. The purpose of this thesis is to analyze if sentiment analysis, including Google Trends, can be used solely to predict a company's stock movement. The following problem to be investigated is whether sentiment analysis with a combination of classic methods forecasting, like Time Series Analysis, could have better results. We expect that google trends can be a good indicator of the global sentiment surrounding the stock market, so it has a strong impact on a company's stock price. However, we are unsure if the impact is strong enough to be used exclusively for stock market prediction.

Περίληψη

Ένα από τα πιο δύσκολα σημεία σχετικά με τις ημερήσιες συναλλαγές είναι να γνωρίζεις πότε θα αυξηθεί και πότε θα πέσει η αγορά. Οι όροι αναζήτησης μπορούν να μας δώσουν κάποια εικόνα για την πρόβλεψη αυτών των περιόδων, ώστε να μπορούμε να κερδίσουμε τόσο στις ανοδικές όσο και στις πτωτικές αγορές. Ο σκοπός αυτής της διπλωματικής είναι να αναλύσει εάν η ανάλυση συναισθημάτων, συμπεριλαμβανομένων των Google Trends, μπορεί να χρησιμοποιηθεί αποκλειστικά για την πρόβλεψη της κίνησης των μετοχών μιας εταιρείας. Το ακόλουθο πρόβλημα που πρέπει να διερευνηθεί είναι αν η ανάλυση συναισθημάτων με έναν συνδυασμό κλασικών μεθόδων πρόβλεψης, όπως η ανάλυση χρονοσειρών, θα μπορούσε να έχει καλύτερα αποτελέσματα. Αναμένουμε ότι τα Google Trends μπορούν να αποτελέσουν έναν καλό δείκτη του παγκόσμιου συναισθήματος που περιβάλλει το χρηματιστήριο, οπότε επηρεάζουν σημαντικά την τιμή της μετοχής μιας εταιρείας. Ωστόσο, δεν είμαστε σίγουροι εάν ο αντίκτυπος είναι αρκετά ισχυρός για να χρησιμοποιηθεί αποκλειστικά για προβλέψεις στο χρηματιστήριο.

Table of Contents

List of Figures

List of Tables

Background

Stock market

The stock market is where investors connect to buy and sell investments. When you need groceries, you go to the supermarket. When you're ready to buy stocks or mutual funds, you'll usually buy them online through the stock market, which anyone can access with a brokerage account, robo-advisor or employee retirement plan.

"A stock or share (also known as a company's "equity") is a financial instrument that represents ownership in a company or corporation and represents a proportionate claim on its assets (what it owns) and earnings (what it generates in profits)." — *Investopedia*

To be more specific, there are two types of stocks, common and preferred. The difference is while the holder of the former has voting rights that can be exercised in corporate decisions, the later doesn't. One way for companies to increase their income and invest in their business is when issuing stocks. In addition, for investors, stocks are a way to increase their money and outpace inflation over time. Public companies use the stock market exchange, like New York Stock or Nasdaq exchange, to sell their stocks. On the other, investors use stockbrokers to buy and sell these shares among themselves. Then, the stock exchanges track the supply and demand of each company's stock and affect the price of the stock. Moreover, Stocks can be categorized by the country where the company is domiciled. For example, Nestle and Novartis are domiciled in Switzerland and traded on the SIX Swiss Exchange, so they may be considered as part of the Swiss stock market, although the stocks may also be traded on exchanges in other countries, for example, as American depositary receipts (ADRs) on U.S. stock markets.

The concept behind how the stock market works is pretty simple. Operating much like an auction house, the stock market enables buyers and sellers to negotiate prices and make trades. The term "stock market" often refers to one of the major stock market indexes, such as the Dow Jones Industrial Average or the S&P 500. Because it's hard to track every single stock, these indexes include a section of the stock market and their performance is viewed as representative of the entire market.
You might see a news headline that says the stock market has moved lower, or that the stock market closed up or down for the day. Most often, this means stock market indexes have moved up or down,

meaning the stocks within the index have either gained or lost value as a whole. Investors who buy and sell stocks hope to turn a profit through this movement in stock prices.

Historically, stock trades likely took place in a physical marketplace. These days, the stock market works electronically, through the internet and online stockbrokers. Each trade happens on a stock-by-stock basis, but overall stock prices often move in tandem because of news, political events, economic reports and other factors.

Bull and bear market

Initially, the terms "bull" means that the market is going up aggressively over a period of time. As the market starts to rise, there becomes more and more greed in the stock market. You see more and more people thinking, *"Oh yeah it's time to put money into the market."* Opposite, the "bear" market it's a market where quarter after quarter the market is moving down about 20 percent. Then, people start to get really scared about putting money into the stock market.

Specifically, a bull market describes steady upward movement in the market and this situation inspires optimism, and a desire to invest heavily. Investors who exhibit similar behavior are known as bull investors. It's important to remember that a bear market is associated with a general sense of decline which tends to instill fear in the hearts of stockholders. This phenomenon usually occurs in periods of time where the economy is in recession, unemployment is high, inflation is growing rapidly, there are large reductions in stock values or stock markets are falling. Investors who exhibit similar behavior are known as bear investors.

Bull and Bear market

Interestingly, a bear market is named for the way that this particular animal attacks its victims. A bear swipes downward during an attack, thus becoming a metaphor for market activity under these conditions.

In some ways, bulls and bears are two sides of the same coin, as they tend to follow one another, each taking their turn. A bull market is often the result of economic expansion and optimism in the markets as a whole. While bears are part of the contraction that follows peaks and bubbles in the market. Both animals have their rightful place in the market. They are a reminder that investors are plagued with the inevitable highs and lows of the stock market. The stock market crash of 1929 was one of the worst in U.S. history. On October 29, 1929, now known as Black Tuesday, investors panicked and liquidated their holdings. The Dow Jones lost 89% of its industrial capital by 1932 and the Great Depression followed.

Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH) or the market theory is the basic theory regarding stock price forecasting. EMH asserts that the price of a stock reflects all information available and everyone has some degree of access to the information. The implication of EMH is one can outperform the market in the long run.

“In an efficient market at any point in time the actual price of a security will be a good estimate of its intrinsic value.” - *Eugene F. Fama*

In 1970, in “Efficient Capital Markets: a Review of Theory and Empirical Work, ”Eugene F. Fama defined a market to be “informationally efficient” if prices at each moment incorporate all available information about future values. The theory remains controversial but many believe that in a short period of time one can beat the market.

While a large amount of academics point support EMH, an equal amount of dissension also exists. For example, investors such as Warren Buffett have consistently beaten the market over long periods. According to the EMH that is impossible. Warren Buffet believes that, for less talented investors than

himself, EMH is a powerful enough concept that most investors need to be putting money into index funds.

“By periodically investing in an Index Fund, the know nothing investors can actually outperform most investment professionals” - *Warren Buffett*

The irony of that state is that Buffett's consistent investment success is the best example of why efficient market theory is preposterous. His views about the supposed efficiency of the market can be summed up by his repeated references over the years to the parable of Mr. Market, an allegory oft-repeated in Benjamin Graham’s renowned security analysis course at Columbia University where Buffett earned a master of economics in 1951.

Forecasting

“Forecasting: the attempt to predict the unknowable by measuring the irrelevant, this task employs most people on Wall Street.” - *Jason Zweig*

The words of Jason Zweig, author of the *Devil’s Financial Dictionary*, are particularly apt at this time of year. Because of the coronavirus we hear a lot from financial forecasters. Highly paid experts attempt to predict what the economy and the markets will do, even though decades of research confirms the prediction game is a pretty fruitless one.

Generally, forecasting is when someone predicts the future as accurately as possible, using all of the information available, such as historical data and knowledge of any future events that might impact the prediction. To be more specific is a planning tool, based on data from the past and present and analysis of trends, which attempts to cope with the uncertainty of the future. It is not unusual to hear a company's management speak about forecasts, "Our sales did not meet the forecasted numbers," or "we feel confident in our forecasted economic growth and expect to exceed our targets." In the end, all financial forecasts are informed guesses regardless of whether they reflect the specifics of a business, such as sales growth, or predictions for the economy as a whole.

Stock forecasting

Stock forecasting or stock market prediction is when we try to determine the future value of a company stock or other financial instrument traded on an exchange. If the prediction of a stock's future price is successful could yield significant profit. Moreover, stock prediction plays the most crucial role in determining where to put in the money or which stock to be acquired or sold. Admittedly, Market forecasts are particularly tricky. No one can see the future because the world is inherently uncertain and surprising things will happen. Even if you know what's going to happen, however, you might not know how markets will respond. If a forecasting model or technique can precisely predict the direction of the market, investment risk and uncertainty can be minimized.

Theoretical and empirical studies have shown that a positive relationship exists between financial markets and economic growth. Stock markets are characterized by high volatility, dynamism, and complexity. Movements in stock markets are influenced by several factors, such as macro-economic factors, international events, and human behavior.

Prediction methods

The stock market and its trends, in the finance field, are extremely volatile in nature. It attracts researchers to predict its next moves and capture the volatility. Market analysts and investors study the market behaviour and plan their sell or buy strategies accordingly. Everyday the stock market produces a large amount of data. Because that it is very difficult for an individual to consider all the past and current information for predicting future trends of a stock. For these reasons there are two methods for forecasting market trends. The first is Technical analysis and the other is Fundamental analysis. Technical analysis try to predict the future trend considering past price and volume . On the other hand, Fundamental analysis of a business try to get some insights involving analyzing its financial data . By the efficient-market hypothesis the efficacy of both technical and fundamental analysis is disputed because EMH states that stock market prices are essentially unpredictable.

Technical analysts, meanwhile, use historical securities data and predict future prices on the assumption that stock prices are determined by market forces and that history tends to repeat itself (*Levy,1967*). The Algorithm in place helps a trader to forecast the time at which the price would be the most favorable to either buy or sell a stock. A variety of technical approaches to market trend

prediction have been proposed in the research literature, ranging from AutoRegressive Integrated Moving Average (ARIMA) to ensemble methods. Huang et al. in their work demonstrated the superiority of Support Vector Machines (SVM) in forecasting weekly movement directions of the NIKKEI 225 index, and Lin et al. managed to achieve 70% accuracy by combining decision trees and neural networks.

Fundamentalists forecast stock prices on the basis of financial analyses of companies or industries. The method involves meticulous studying of a company's financial health, the value of assets, debts, cash, revenues, expenses, profitability and plans of development. Warren Buffett is perhaps the most famous of all Fundamental Analysts.

It is remarkable that the market is not only driven by fundamentals, but also short-term sentiments, which makes currencies volatile on a day-to-day basis. It is often seen that despite long-term fundamentals showing an uptrend, a currency remains down, due to an overall 'bad mood.' This bad mood means that the vast majority of traders are committed to a down position, due to some reason. Such sentiments often help traders assume a particular position. Sentiment analysis are often referred to as contrarians. These traders invest against the majority view of the market, since they believe that the markets always tend to move against the existing sentiment, sooner or later.
Sentiment trading by itself is quite risky, since it involves uninformed trades. Uninformed traders may be moving the market prices away from fundamental values. However, used in combination with other forms of analysis, it can help in getting a clearer picture.

Below we will combine technical analysis with opinion analysis

Methods

Literature Study

Sentiment Analysis

The study by Larose, D. T (2014) refers to John Naisbitt who said: “we are drowning in information but starved for knowledge.” The above well-known saying is more relevant than ever as we live in the "Information Age". The contribution of the internet to development is undoubtedly crucial as it has changed not only the way but also the mentality of billions of users around the world in accessing information sources in order to inform them. A more specialized field of natural language processing and word mining is Opinion Mining or Sentiment Analysis.

The term "sentiment" used in connection with the automatic analysis of the text to be evaluated and the prediction of opinion through it first appears in the years 2001 in the publications of Das, S. (2001) respectively, due to the interest of the latter in the analysis of sense at the market. Numerous studies have been carried out to understand the intricate relationship between sentiment and price on the financial market. Wang et al. investigated the correlation between stock performance and user sentiment extracted from StockTwits and Seeking Alpha 3 . Ding et al. proposed a deep learning method for event-driven stock market prediction and achieved nearly 6% improvements on S&P 500 index prediction. Arias et al. investigated whether information extracted from Twitter can improve time series prediction, and found that indeed it could help predict the trend of volatility indices (e.g., VXO, VIX) and historic volatilities of stocks. Bollen et al. in their research identified that some emotion dimensions, extracted from Twitter messages, can be good market trend predictors.

Google trends

Most new investors are attracted to the market when they hear about stock gains on the news and a rising market. Google search results tend to reflect this. Analyzing search phrases like “what stocks to buy,” “top stocks to buy,” and “hot stocks,” can paint a broad picture of whether investors are interested in buying or selling.

Google Trends is a tool that can be used to compare two search terms according to volume. This tool is useful in analyzing the stock market itself because it can be a good indicator of the global sentiment surrounding the stock market. The terms bull and bear market are used to describe how stock markets are doing in general—that is, whether they are appreciating or depreciating in value. At the same time, because the market is determined by investors' attitudes, these terms also denote how investors feel about the market and the ensuing trends.

Simply put, a bull market refers to a market that is on the rise. It is typified by a sustained increase in price, for example in equity markets in the prices of companies' shares. In such times, investors often have faith that the uptrend will continue over the long term. Typically, in this scenario, the country's economy is strong and employment levels are high. By contrast, a bear market is one that is in decline, typically having fallen 20% or more from recent highs. Share prices are continuously dropping, resulting in a downward trend that investors believe will continue, which, in turn, perpetuates the downward spiral. During a bear market, the economy will typically slow down and unemployment will rise as companies begin laying off workers.

Time Series Analysis

Analysis of time series is a statistical technique that deals with trend analysis or time series data. Time series data means that data is in a series of particular time periods or intervals. It is different from Time Series forecasting which is the use of a model to predict future values based on previously observed values. Specifically, analysis of time series includes methods for analyzing time series data in order to extract momentous statistics and other characteristics of the data. The most important use of studying time series is that based on past experience could predict the future behaviour of the variable.

While time series analysis is mostly statistics, with time series forecasting enters Machine Learning. Time series analysis is a preparatory step to time series forecasting. Various forecasting techniques are available for time series forecasting.

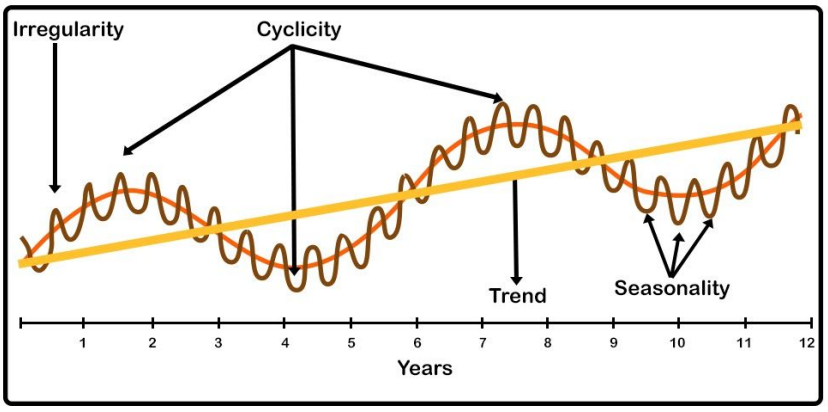
Stationarity

A stationary time series is if its statistical properties do not change over time. Moreover a stationary time series shows the mean value of the series that remains constant over a time period. In other words, it has constant mean and variance, and covariance is independent of time.

Usually stock prices are not a stationary process, often we see a growing trend, or the variance is changing over time. In particular its volatility increases over time. Ideally, we want to have a time series which is stationary for modelling. Of course that is difficult to find because not all of them are stationary but we can make some transformations to make them stationary.

Components of Time Series

The components of a time series affect the values of an observation in a time series. The four categories of the components are Trend, Seasonality, Cyclical and Noise.



Components of Time Series

Trend is the direction (increasing or decreasing) in which something is developing. A trend can be upward(uptrend) or downward(downtrend).

Seasonality refers to periodic fluctuations. In particular seasonality is the repeating short term cycle in the series.

"A repeating pattern within each year is known as seasonal variation, although the term is applied more generally to repeating patterns within any fixed period." - *page 6, Introductory time series with R*

There are many types of seasonality (daily, weekly, monthly, yearly). In addition a cyclic structure in a time series sometimes may not be seasonal.

If it constantly repeats at the same frequency, it is seasonal, otherwise it is called a cycle. Seasonal and Cyclic Variations are the periodic changes or short-term fluctuations.

One more factor which causes the variation in the variable is fluctuations which are not regular variations and are purely random or irregular. These fluctuations are unforeseen, unpredictable, uncontrollable, and are erratic. These forces are earthquakes, floods, wars, and any other disasters.

ARIMA models

The acronym ARIMA stands for Auto-Regressive Integrated Moving Average. Box and Jenkins (1970) proposed ARIMA models for time series analysis and forecasting. Some studies in order to forecast the returns of the stock market employing ARIMA models. Quite a few studies found that ARIMA models produced inferior forecasts for financial time series data. Moreover, ARIMA models are the most significant type of models for forecasting a time series. The ARIMA forecasting equation for a stationary time series is a linear equation in which the forecasting consists of lags of the dependent variable and/or of the forecast errors. Some significant cases of that models are the random-walk, autoregressive models and exponential smoothing models.

A nonseasonal ARIMA model is classified as ARIMA(p,d,q) model where p is the number of autoregressive terms, d is the number of nonseasonal differences needed for stationarity, and q is the number of lagged forecast errors in the prediction equation.

Jarque Bera test

The Jarque-Bera Test is a test for normality and is a type of Lagrange multiplier test. The test is named for Carlos Jarque and Anil K. Bera and in statistics is a goodness of fit test of whether sample data have the skewness and kurtosis matching a normal distribution. Normality is one of the assumptions for many statistical tests like the F test or t test and Jarque-Bera test usually run before one of these tests to confirm normality. Most of the time used for large data sets because other normality tests are not reliable when n is large (for example, Shapiro-Wilk isn't reliable with n more than 2.000).

The t-statistic JB is defined as $JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right)$ when n is the number of observations, and S, K the skewness and kurtosis respectively.

Typically, a normal distribution has a skew of zero because it's perfectly symmetrical around the mean and a kurtosis of three because kurtosis tells us how much data is in the tails and how "peaked" the distribution is. It is significant that it's not necessary to know the mean or the standard deviation for the data in order to run the test.

Ljung-Box test

The Ljung Box test or just the Box test is a way to test for the lack of serial autocorrelation, up to a specified lag *k*. The test named for Greta M. Ljung and George E. P. Box.

The test determines whether or not errors are independent and identically distributed (iid) which means white noise or whether there is something more behind them. Significantly whether any of a group of autocorrelations of a time series are different from zero. Essentially, it is a test of lack of fit because if the autocorrelations of the residuals are very small we say that the model doesn't show significant lack of fit. The Ljung–Box test is widely applied in econometrics.

The null hypothesis of the Box Test is that our model does not show lack of fit and the alternative hypothesis is just that the model does show a lack of fit.

The t-statistic is $Q = n(n+2)\sum_{k=1}^{h}(\hat{r}_k^2 / n-k\ )$ where *n* is the sample size, $\hat{r}_k$, h is the number of lags being tested and is the sample autocorrelation at lag k.

ARCH/ Garch models

Robert F. Engle (1982) introduced a model in which the variance at a time t is modeled as a linear combination of past squares residuals and called it an Autoregressive Conditional Heteroscedasticity (ARCH). Engle is the winner of the *Nobel Memorial Prize in Economic Sciences (2003)* for methods of analyzing economic time series with time-varying volatility (ARCH).

"As I look back over my career, this Prize is the high point" - *Robert F. Engle*

ARCH models relate to economic forecasting and measuring volatility. Some of the real-time examples where applied are stock prices, bond prices, inflation rates, oil prices, GDP and unemployment rates. In time series where the variance is increasing in a systematic way, such as an increasing trend, this property of the series is called heteroskedasticity.

Tim Peter Bolerslev (1986) introduced a more general structure in which the variance model looks more like an ARMA than an AR and called this a Generalized Autoregressive Conditional Heteroskedasticity GARCH (generalized ARCH) process. He earned his Ph.D. in 1986 written under the supervision of Robert F. Engle. In addition he is editor of the *Journal of Applied Econometrics.*

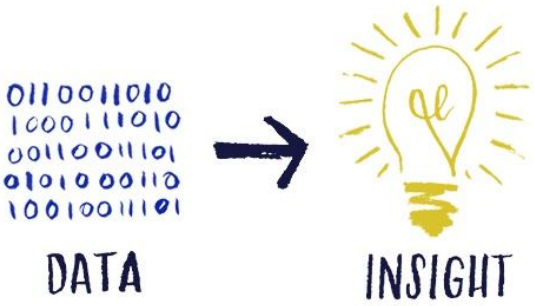
ACF / PACF

ACF is an (complete) auto-correlation function which gives us values of auto-correlation of any series with its lagged values. The ACF plot in simple terms, it describes how well the present value of the series is related with its past values. ACF considers all the components of an ARIMA model while finding correlations hence it's a 'complete auto-correlation plot'.

PACF is a partial auto-correlation function. Basically instead of finding correlations of present with lags like ACF, it finds correlation of the residuals with the next lag value hence 'partial' and not 'complete' as we remove already found variations before we find the next correlation. So if there is any hidden information in the residual which can be modeled by the next lag, we might get a good correlation and we will keep that next lag as a feature while modeling.

Data collection

We live in a world where vast amounts of data are collected daily. Analyzing such data is an important need. The abundance of data, coupled with the need for powerful data analysis tools, has been described as a ***data rich but information poor*** situation. The fast-growing, tremendous amount of data, collected and stored in large and numerous data repositories, has far exceeded our human ability for comprehension without powerful tools.



DATA INSIGHT: Don't be data rich but knowledge poor.

For the analysis of this work we used data from the ***Netflix* stock** both from ***Google Trends*** and ***Yahoo Finance***.

Netflix

In 2010 Netflix just started claiming its position as a competitor to traditional media with its share worth less than $ 8. As the coronavirus pandemic spread in early spring, Netflix benefited. Shelter in place orders meant that people were staying at home, with few options for entertainment other than their televisions and devices. But few predicted just how much of a lift Netflix would get. Before the company reported first quarter earnings in late April, Wall Street analysts scrambled to revise their subscriber growth estimates to reflect the pandemic's effect on streaming. Most suspected Netflix would add about 7.5 million new subscribers, in fact, the company more than doubled that with nearly 16 million new additions for a total of 183 million.

In addition, Netflix has said it expects 7.5 million global paid net additions this time. But the forecast came with a caveat by the company: "Given the uncertainty on home confinement timing, this is mostly guesswork." Thanks to the huge influx of subscribers, Netflix's stock has been one of the best performers in the S&P 500 this year. MoreoverThe pandemic hasn't been entirely beneficial to Netflix, however, as the virus led to a months-long pause in film and television production. Also, Netflix has said it has enough content in the well to last into 2021, but huge subscriber gains could lead to greater demand. Today Netflix is considered a technology and entertainment giant of the same caliber as Apple and Facebook. The video-streaming service added a record 15.77 million paid

subscribers during the first quarter. However, Netflix is also warning that some upcoming titles will be delayed due to the pandemic.

Stock data collection

Google Trends is a trends search feature that shows the popularity of a search term in Google. You can view whether a trend is on the rise or declining. You can also find demographic insights, related topics, and related queries to help you better understand the Google trends.

"Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means that there was not enough data for this term." - *Google Trends*

In Yahoo Finance we can view and download historical prices to forecast the future of a company or gain market insight. Historical data can be downloaded as a CSV file to be used offline, which you can open with Excel or a similar program. If the data requested is beyond the range of historical prices available through Yahoo Finance, all available data within the range is displayed. Historical prices usually don't go back earlier than 1970.

Data preprocessing

Time Series Analysis

Firstly, we downloaded the data from Yahoo Finance. We selected daily historical prices from Netflix, Inc. (NFLX) for time period 17/07/2017 - 17/04/2020. Then, we only dealt with the closing prices of the share and we calculated the Netflix stock returns.

A return, also known as a financial return, is the amount of net profit from a business or other investment compared with the total amount of capital. In particular, return on investment usually is expressed as a ratio, or percentage, which compares profit with capital. It is a commonly reported number and used for many types of financial analysis.

For the Analysis of Time Series we used the EViews 11 Student Version with the data from Yahoo Finance. EViews is a statistical package used mainly for time series oriented econometric analysis.

Sentiment Analysis

Regarding data from Google Trends, we find what is the interest of people over time, worldwide, to search for the terms “Netflix” and then for the search terms “netflix bull” and “netflix bear”, respectively.

The calculations for the Sentiment Analysis were done with the help of the data from the Google Trends in Microsoft Excel, a handy software that can be used to store and organize many data sets.

Information Criteria

An information criterion is a measure of the quality of a statistical model. It takes into account:

- how well the model fits the data
- the complexity of the model.

Information criteria are used to compare alternative models fitted to the same data set. All else being equal, a model with a lower information criterion is superior to a model with a higher value.

The most widely used information criteria for predictive modeling is Akaike's information criterion, which is usually referred to as the AIC and the most widely used information criteria for comparing Latent Class models is the Bayesian Information Criterion, which is usually referred to as the BIC.

Akaike information criterion — The AIC compares models from the perspective of information entropy, as measured by Kullback-Leibler divergence.

The AIC for a given model is $-2\,log\,L(\widehat{\theta})\,+\,2k$ **.**

ayesian (Schwarz) information criterion — The BIC compares models from the perspective of decision theory, as measured by expected loss.

The BIC for a given model is $-2\ log\ L(\widehat{\theta})\ +\ k\ log(T)$.

Where the sample is size greater than 7 the BIC is a more stringent criterion than the AIC. An intuitive way to think about the difference between the AIC and the BIC is that the AIC is closer to the idea of testing significance at the 0.05 level of significance whereas the BIC is more akin to testing at the 0.001 level.

Results

Sentiment Analysis with Google Trends

Initially we gather the data from Google Trends for the search terms “Netflix”, “netflix bull” and “netflix bear”. The time period is one month prior to the date for which we predict the movement of Netflix stock. Every number is used in order to predict the stock movement. Then we only kept the prices of the dates that the Netflix stock traded on the stock exchange.

The data are from 0 to 100.
“The numbers represent the search interest relative to the highest point on the chart for the selected region and time. A value of 100 is the peak popularity of the term, whilst a value of 50 means that the term is half as popular.” - *Google Trends*
In the sequel, we categorize the data according to their value.

- ➔ If the search term “**Netflix**” for a specific day is > 50 means that the Netflix stock return will probably rise. Otherwise, it means that the Netflix stock return will probably fall.

- ➔ If the search term “**netflix bull**” for a specific day is > 50 means that the Netflix stock return will probably rise. Otherwise, it means that the Netflix stock return will probably fall.

- ➔ If the search term “**netflix bear**” for a specific day is > 30 means that the Netflix stock return will probably fall. Otherwise, it means that the Netflix stock return will probably rise.

For the search term “netflix bear” we choose the limit of 30 because it is widespread that in economics people react more strongly to something negative and less to something positive.

At this point the Sentiment Analysis is divided into two versions. The first is less risky forecasting but not on a daily basis and the latter is everyday forecasting with an extra search term.

We consider that if both the search terms "netflix bull" and "netflix bear" are positive then the market is bullish. If both the search terms "netflix bull" and "netflix bear" are negative then the market is bearish. - 1st version

On the contrary, if the search term "netflix bull" is positive and the other is negative or if the search term "netflix bear" is positive and the other is negative with the 1st version we can not decide. At this time we use the search term "Netflix". If it is positive, the market is bullish. If it is negative, the market is bearish. - 2nd version

Finally, in order to complete the Sentiment Analysis, we compare the results with the real returns of Netflix stock. If Google Term has the same emotion with reality then forecasting is true. Otherwise, forecasting is false. The time period of forecasting is from 17/04/2020 to 13/08/2020.

The results of the analysis are summarized below.

1st version			
not on a daily basis	True	False	Sum
62,07% success	35	23	58

2nd version			
everyday forecasting	True	False	Sum
56,10% success	46	36	82

Time Series Analysis - ARIMA models

For the Time Series Analysis we use the stock returns which we have calculated from the closing price that were downloaded from Yahoo Finance. The time period is from 17/07/2017 to 13/08/2020. In addition, we use periodicity of 260 observations.

After calculating all the desired forecasts, using the appropriate ARIMA models, we categorized the returns as positive if the actual return from the previous day is less than the forecast day and as negative if the opposite is true.

At this point the Time Series Analysis is divided into two versions. The first is less risky forecasting but not on a daily basis and the latter is everyday forecasting with an extra search term.

We calculate the (absolute) difference between the actual return in a previous day and the predicted return and discard the times where their difference is less than 0,01. If the returns are positive, the market is bullish. Otherwise, the market is bearish. - 1st version

On the second version, we do not reject any day. Therefore, if the returns are positive, the market is bullish. Otherwise, the market is bearish. - 2nd version

Finally, in order to complete the Time Series Analysis, we compare the results with the real returns of Netflix stock. If time series results have the same emotion with reality then forecasting is true. Otherwise, forecasting is false.

The results of the analysis are summarized below.

1st version			
not on a daily basis	True	False	Sum
86,44% success	51	8	59

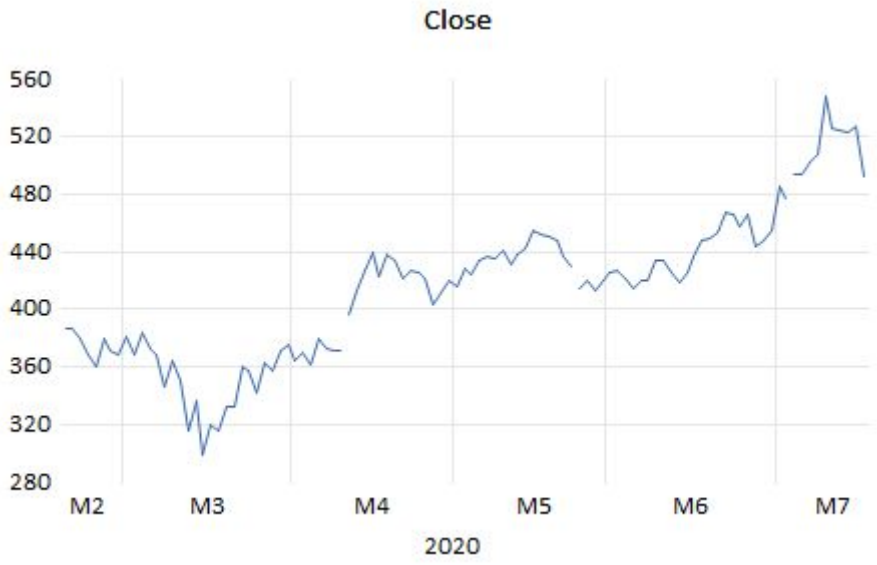
2nd version			
everyday forecasting	True	False	Sum
74,39% success	61	21	82

Forecasting with ARIMA models

The following procedure was done individually for each performance separately.

We mention the procedure for the date 20/07/2020.

The closing prices of the NFLX stock are shown in the chart below.

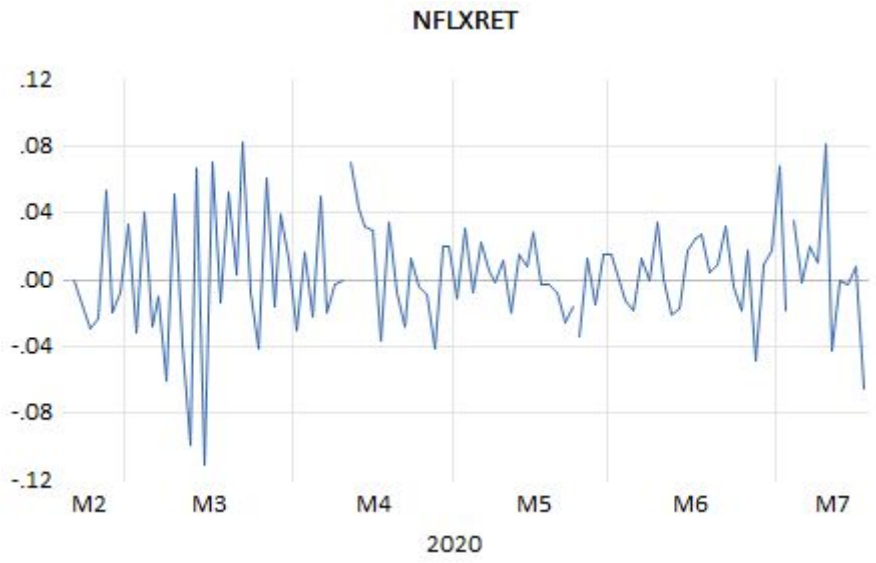


Closing prices of the Netflix stock

(M2 = February, M3 = March, etc.)

We observe a continuous increase in the stock price from 16/03/2020 onwards, which is probably due to covid-19 since people were staying at home for longer periods of time due to quarantine. Then we calculated the Netflix stock returns, which we use for analysis.

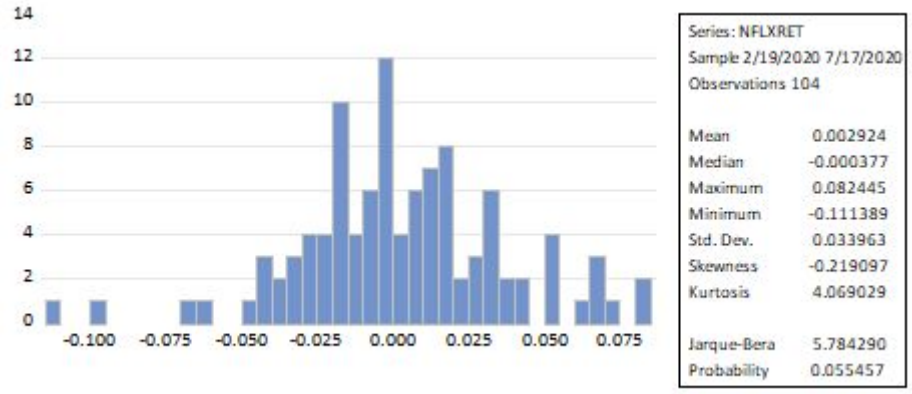
The time series of **Netflix stock returns** (NFLXRET) is shown below:



Netflix stock returns

From the graph we notice that the series shows intense heteroscedasticity in February and is stationary. However, in order to verify our observations, we will proceed with the appropriate statistical methods.

First, we calculated the basic performance characteristics of the Netflix stock. In particular, the sample mean is 0.0029 and the sample standard deviation is 0.033. The maximum value is 0.082 the minimum value is -0.111. In addition, the skewness coefficient is -0.291, close enough to zero. Negative asymmetry means that the mean value is less than the median. The coefficient of kyrtosis is 4,069. Due to the fact that the value of the factor is greater than 3, it is a subtle distribution, ie most values are concentrated close to the mean, as shown in the figure below. The curvature and asymmetry coefficients together with the sample number are variables of the Jarque-Bera statistical function. With the result function is performed in the critical area to determine if the population or sample follow the normal distribution. Therefore, their prices play a key role in controlling normal distribution. The table of results is given below:



Jarque-Bera test of probability

In order to check whether the NFLXRET time series follows the normal distribution, we carry out a Jarque-Bera regularity check.
Hypothesis H0: the time series follows a normal distribution,
versus case H1: the time series does not follow the normal distribution.

Under the null hypothesis of a normal distribution, the Jarque-Bera statistic is distributed as X^2 with 2 degrees of freedom.

The critical area is JB > X^2 (2.05) ⇒ 5.784 < 5.991.
Therefore, the H0 hypothesis is not rejected at a significance level of 5%.
So the NFLXRET time series follows the normal distribution.
Also, PValue = 0.0554 > 0.05.

AR(1)

The autoregressive model first class, AR (1), has the form:

$$Yt = \phi 0 + \phi 1 * X t\text{-}1 + \varepsilon t$$

where $\phi 0$ is the constant term and $\phi 1$ is the slope coefficient of the model.
Adjusting the NFLXRET values to the above AR (1) model with a fixed term yields the following results from Eviews:

Dependent Variable: NFLXRET
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 07/19/20 Time: 19:15
Sample: 2/20/2020 7/17/2020
Included observations: 104
Convergence achieved after 6 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003097	0.002395	1.293209	0.1989
AR(1)	-0.333531	0.089957	-3.707651	0.0003
SIGMASQ	0.001018	0.000109	9.349611	0.0000

R-squared	0.108981	Mean dependent var	0.002924
Adjusted R-squared	0.091337	S.D. dependent var	0.033963
S.E. of regression	0.032375	Akaike info criterion	-3.993305
Sum squared resid	0.105862	Schwarz criterion	-3.917025
Log likelihood	210.6519	Hannan-Quinn criter.	-3.962402
F-statistic	6.176698	Durbin-Watson stat	1.868289
Prob(F-statistic)	0.002946		

Inverted AR Roots	-.33

Regression Analysis for AR(1) with fixed term

From the regression results we see that the fixed term has p-value = 0.1989 > 0.05. So, at a significance level of 5% we do not reject zero assumption, i.e. the constant condition of the model is zero. Subsequently, re-evaluate model AR (1) without a fixed term.

Dependent Variable: NFLXRET
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 07/19/20 Time: 19:18
Sample: 2/20/2020 7/17/2020
Included observations: 104
Convergence achieved after 7 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.320327	0.088205	-3.631629	0.0004
SIGMASQ	0.001035	0.000112	9.233871	0.0000

R-squared	0.094211	Mean dependent var	0.002924
Adjusted R-squared	0.085330	S.D. dependent var	0.033963
S.E. of regression	0.032482	Akaike info criterion	-3.996188
Sum squared resid	0.107617	Schwarz criterion	-3.945334
Log likelihood	209.8018	Hannan-Quinn criter.	-3.975586
Durbin-Watson stat	1.866739		

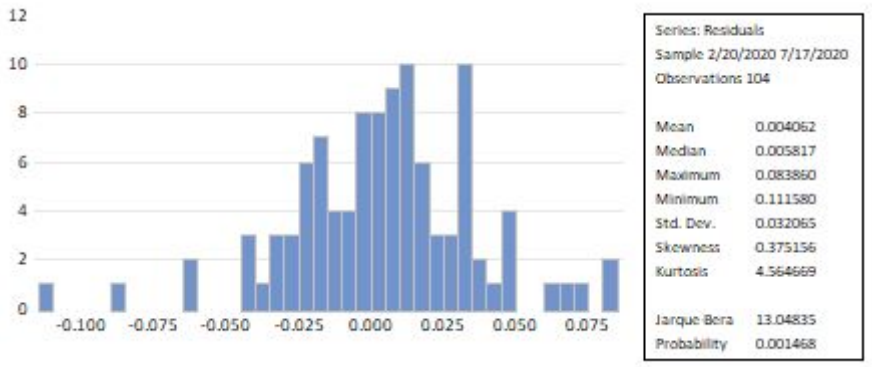
Inverted AR Roots	-.32

Regression Analysis for AR(1) without fixed term

For $\phi 1$ it holds that p-value = 0.0004 < 0.05, so we reject the zero hypothesis at a significance level of 5%. Therefore, the $\phi 1$ estimator is statistically significant. The estimated equation of the autoregressive model first class is:

Yt = -0.32 * Xt-1

If $|\phi 1| = 0.32 < 1$, we conclude that the NFLXRET time series is stationary. In order to investigate the distribution of regression residues, we first calculate the basic characteristics and the histogram of the errors. For $\phi 1$ it holds that p-value = 0.0004 < 0.05, so we reject the zero hypothesis at a significance level of 5%. Therefore, the $\phi 1$ estimator is statistically significant. The estimated equation of the autoregressive model first class is:



Jarque-Bera test of probability for residuals

We observe that the histogram of the residues is quite similar to the histogram of the time series. Next, we perform a Jarque-Bera regularity test.

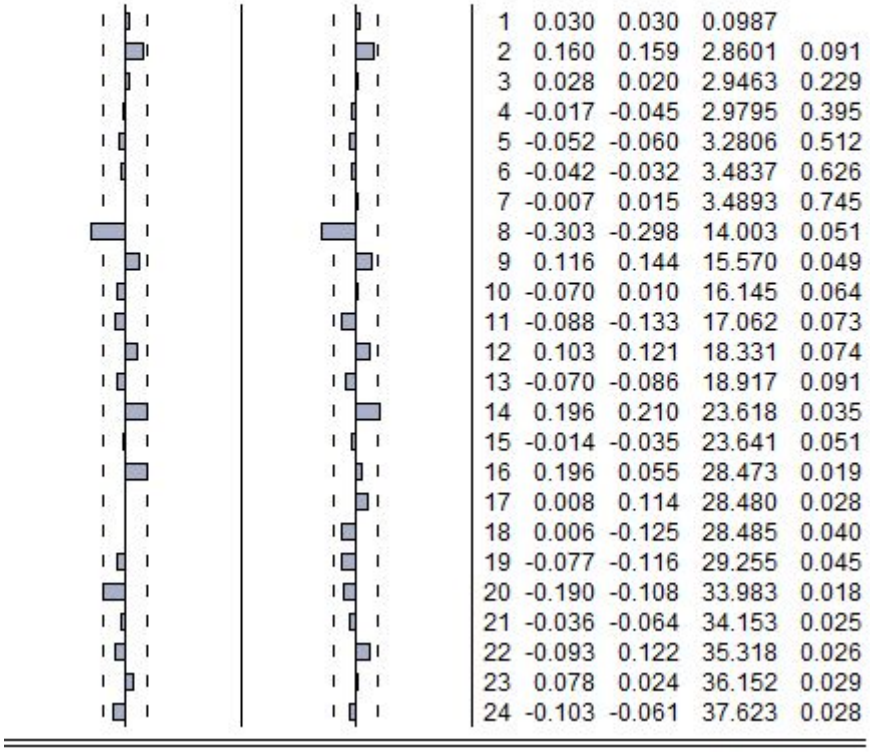
Case H0: the residues follow a normal distribution, compared to

Case H1: the residues do not follow the normal distribution.

Its critical area control: JB > X^2 (2.05) ⇒ 13.048 > 5.991. Therefore, the H0 hypothesis in significance level 5%. Therefore, the residues do not follow the normal allocation.

Next, we proceeded to construct diagrams with the autocorrelations and some partial autocorrelations of the AR(1) model residues (without the fixed term). The diagram are below:

Date: 07/19/20 Time: 19:32
Sample (adjusted): 2/20/2020 7/17/2020
Q-statistic probabilities adjusted for 1 ARMA term



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.030	0.030	0.0987	
		2	0.160	0.159	2.8601	0.091
		3	0.028	0.020	2.9463	0.229
		4	-0.017	-0.045	2.9795	0.395
		5	-0.052	-0.060	3.2806	0.512
		6	-0.042	-0.032	3.4837	0.626
		7	-0.007	0.015	3.4893	0.745
		8	-0.303	-0.298	14.003	0.051
		9	0.116	0.144	15.570	0.049
		10	-0.070	0.010	16.145	0.064
		11	-0.088	-0.133	17.062	0.073
		12	0.103	0.121	18.331	0.074
		13	-0.070	-0.086	18.917	0.091
		14	0.196	0.210	23.618	0.035
		15	-0.014	-0.035	23.641	0.051
		16	0.196	0.055	28.473	0.019
		17	0.008	0.114	28.480	0.028
		18	0.006	-0.125	28.485	0.040
		19	-0.077	-0.116	29.255	0.045
		20	-0.190	-0.108	33.983	0.018
		21	-0.036	-0.064	34.153	0.025
		22	-0.093	0.122	35.318	0.026
		23	0.078	0.024	36.152	0.029
		24	-0.103	-0.061	37.623	0.028

Ljung-Box test in the residuals of the AR(1)

We apply the Ljung-Box control for autocorrelation up to 6th order in the residuals of the AR(1) model.

Hypothesis H0: $\rho 1 = ... = \rho 6 = 0$, versus H1: at least one $\rho j \neq 0$, j = 1,2, ..., 6

The statistic is Q-Stat = 3.4837 < X^2 (6-1.0.05) = 11.070. Therefore, we do not reject the null hypothesis (H0) at a significance level of 5%. So, there is no autocorrelation up to 6th order in the residuals of the AR(1) model. Then we performed a statistical test to find out in which class it exists autocorrelation.

Statistical test for autocorrelation of order j, 1, = 1,2, ..., 24.

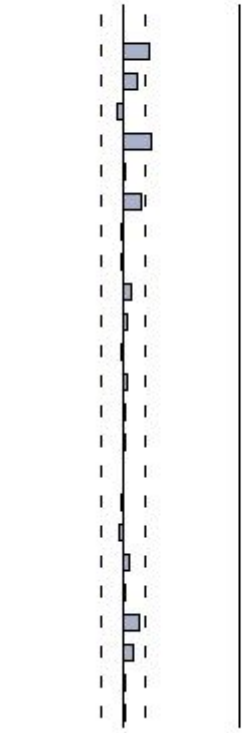
Hypothesis H0: pj = 0, versus H1: pj ≠ 0.

Critical range: $\hat{p}j$ [-Z0.05 * (1 / √N), Z0.05 * (1 / √N)] = [- 1,96√104, 1,96√104] = [-0,1921, 0.1921].

For j = 8,14,16 we reject the null hypothesis at a significance level of 5%, while for j ≠ 8,14,16 we do not reject Ho.

Therefore, there is 8th, 14th and 16th order autocorrelation in AR(1) residuals model. The existence of autocorrelation only for 8th, 14th, 16th grade is enough to find that the residuals have white noise behavior. Since the behavior of the residuals is white noise, we conclude that the model AR(1) is suitable for NFLXRET time series values. The graph of the autocorrelations and partial autocorrelations for the squares of the residuals is below.

Date: 07/19/20 Time: 19:48
Sample (adjusted): 2/20/2020 7/17/2020
Included observations: 104 after adjustments



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.006	0.006	0.0042	0.948
		2	0.240	0.240	6.2375	0.044
		3	0.116	0.120	7.6992	0.053
		4	-0.060	-0.124	8.0941	0.088
		5	0.252	0.211	15.169	0.010
		6	0.017	0.056	15.202	0.019
		7	0.172	0.088	18.564	0.010
		8	-0.022	-0.101	18.621	0.017
		9	-0.010	-0.037	18.634	0.028
		10	0.074	0.043	19.277	0.037
		11	0.026	0.068	19.357	0.055
		12	-0.015	-0.124	19.385	0.080
		13	0.028	0.010	19.481	0.109
		14	0.013	0.058	19.502	0.147
		15	0.013	0.027	19.523	0.191
		16	-0.001	-0.066	19.523	0.242
		17	-0.010	-0.015	19.536	0.299
		18	-0.041	-0.035	19.751	0.347
		19	0.048	0.097	20.050	0.392
		20	0.024	0.015	20.128	0.450
		21	0.133	0.118	22.491	0.372
		22	0.097	0.094	23.744	0.361
		23	0.019	0.008	23.794	0.415
		24	0.019	-0.091	23.846	0.470

Ljung-Box test in the squares of the residuals of the AR(1)

We apply the Ljung-Box test for autocorrelation up to 4th order in the squares of the residuals of the AR(1) model.

Hypothesis H0: $\rho 1 = ... = \rho 4 = 0$, versus H1: at least one $\rho j \neq 0$, j = 1,2,3,4

The statistic is Q-Stat = 8,0941 > X ^2 (4-1,0.05) = 7,815.

Therefore, we reject the H0 hypothesis at a significance level of 5%. So there is an autocorrelation up to 4th order in the squares of the residues of the AR(1) model. Therefore, we have an indication of ARCH / GARCH behavior in its dispersion values disturbing condition. We test for ARCH errors with the Lagrange Multiplier test (LM test).

Heteroskedasticity Test: ARCH

F-statistic	3.131486	Prob. F(2,99)	0.0480
Obs*R-squared	6.068830	Prob. Chi-Square(2)	0.0481

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 07/19/20 Time: 20:20
Sample (adjusted): 2/24/2020 7/17/2020
Included observations: 102 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000802	0.000233	3.444913	0.0008
RESID^2(-1)	0.000721	0.098519	0.007314	0.9942
RESID^2(-2)	0.246485	0.098506	2.502225	0.0140

R-squared	0.059498	Mean dependent var	0.001053
Adjusted R-squared	0.040498	S.D. dependent var	0.001912
S.E. of regression	0.001872	Akaike info criterion	-9.694162
Sum squared resid	0.000347	Schwarz criterion	-9.616957
Log likelihood	497.4023	Hannan-Quinn criter.	-9.662899
F-statistic	3.131486	Durbin-Watson stat	2.028830
Prob(F-statistic)	0.048006		

ARCH test of Heteroskedasticity, 2 lags

Assumptions: Ho: $\alpha 1 = \alpha 2 = 0$, i.e. there are no ARCH errors against H1: $\alpha 1 \neq 0$ and/or $\alpha 2 \neq 0$, there are ARCH errors. The statistical function LM = T * R ^ 2 = 6,068> X ^ 2 (2, 0.05) = 5,991. Therefore, we reject the null hypothesis (Ho), therefore we have ARCH behavior. Also, according to the p-value the second lag is not statistically significant, so we do not have ARCH (2) behavior. We assume autocorrelation in the squares of the residuals, up to three lags.

Heteroskedasticity Test: ARCH

F-statistic	2.580717	Prob. F(3,97)	0.0579
Obs*R-squared	7.465546	Prob. Chi-Square(3)	0.0585

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 07/19/20 Time: 20:26
Sample (adjusted): 2/25/2020 7/17/2020
Included observations: 101 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000702	0.000247	2.837164	0.0055
RESID^2(-1)	-0.029994	0.102222	-0.293422	0.7698
RESID^2(-2)	0.246598	0.098891	2.493635	0.0143
RESID^2(-3)	0.123920	0.102266	1.211739	0.2286

R-squared	0.073916	Mean dependent var	0.001051
Adjusted R-squared	0.045275	S.D. dependent var	0.001921
S.E. of regression	0.001877	Akaike info criterion	-9.679445
Sum squared resid	0.000342	Schwarz criterion	-9.575876
Log likelihood	492.8120	Hannan-Quinn criter.	-9.637517
F-statistic	2.580717	Durbin-Watson stat	1.938745
Prob(F-statistic)	0.057902		

ARCH test of Heteroskedasticity, up to 3 lags

From the LM test according to p-value = 0.0585 > 0.05. Therefore for a 5% significant level we do not have ARCH behavior. Also, the case for autocorrelation up to 3 lags shows that the first and third lags are not statistically significant, according to the p-value (0.7698 > 0.05 and 0.2286 > 0.05) at significance 5%. Therefore, we do not have ARCH (3) behavior.

MA(1)

Then we evaluate the MA(1) model.

The model of first class moving average, MA(1), has the form:

$$Yt = \theta 0 + \varepsilon t - \theta 1 * \varepsilon t\text{-}1$$

where $\theta 0$ is the constant term, εt is the white noise with mean value 0 and variance σ^2.

The MA(1) process is "reversed" in the sense that it is capable of an AR (∞) representation. The condition $| \theta 1 | < 1$ is known as the reversibility condition.

It is noted that for MA(1) but also for any other MA(q) process the issue of non-stationarity is not raised as they are all a linear combination of a finite number of white noise terms.

Adjusting the NFLXRET values to the above MA(1) model with a fixed term yields the following results:

Dependent Variable: NFLXRET
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 07/19/20 Time: 20:45
Sample: 2/20/2020 7/17/2020
Included observations: 104
Convergence achieved after 18 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003107	0.002454	1.265806	0.2085
MA(1)	-0.243872	0.100186	-2.434200	0.0167
SIGMASQ	0.001053	0.000116	9.053911	0.0000

R-squared	0.078638	Mean dependent var	0.002924
Adjusted R-squared	0.060393	S.D. dependent var	0.033963
S.E. of regression	0.032922	Akaike info criterion	-3.960362
Sum squared resid	0.109467	Schwarz criterion	-3.884081
Log likelihood	208.9388	Hannan-Quinn criter.	-3.929458
F-statistic	4.310144	Durbin-Watson stat	2.074340
Prob(F-statistic)	0.015986		

Inverted MA Roots	.24

Regression Analysis for MA(1) with fixed term

From the regression results we see that the fixed term has p-value = 0.2085 > 0.05. So, at a significance level of 5% we do not reject zero assumption, i.e. the constant condition of the model is zero. Subsequently, we re-evaluate model MA(1) without a fixed term.

Dependent Variable: NFLXRET
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 07/19/20 Time: 20:48
Sample: 2/20/2020 7/17/2020
Included observations: 104
Convergence achieved after 16 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.227839	0.094798	-2.403430	0.0180
SIGMASQ	0.001069	0.000118	9.031879	0.0000

R-squared	0.064249	Mean dependent var	0.002924
Adjusted R-squared	0.055075	S.D. dependent var	0.033963
S.E. of regression	0.033015	Akaike info criterion	-3.964174
Sum squared resid	0.111177	Schwarz criterion	-3.913320
Log likelihood	208.1370	Hannan-Quinn criter.	-3.943572
Durbin-Watson stat	2.080421		

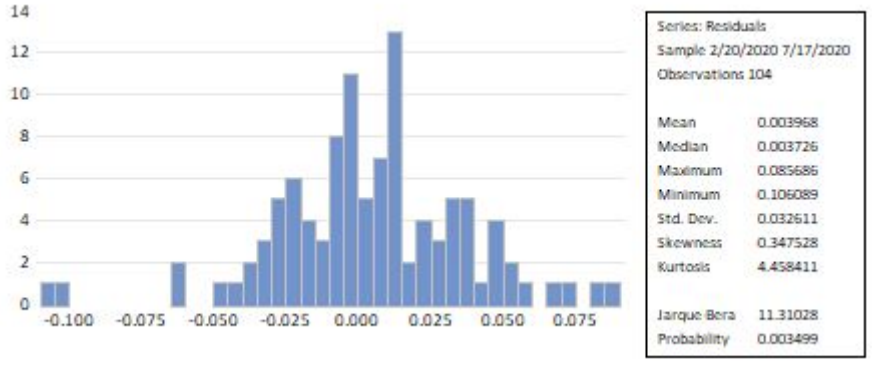
Inverted MA Roots	.23

Regression Analysis for MA(1) without fixed term

For θ 1 it holds that p-value = 0,0180 < 0,05, so we reject the zero case at a significance level of 5%. Therefore, the estimator θ 1 is statistically significant. The estimated equation of the moving average model first class is:

Yt = - 0.018 * ε t-1

In order to investigate the distribution of regression residuals, we calculate the basic characteristics and the histogram of the errors.



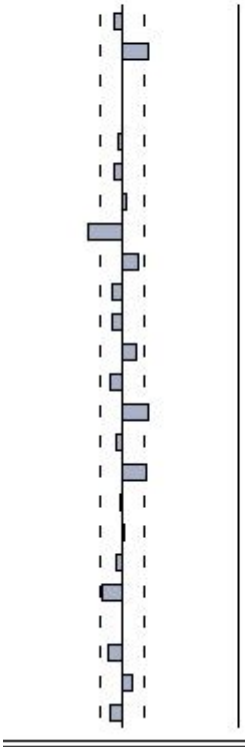
We observe that the histogram of the residuals is quite similar to the histogram

of the time series. Next, we perform a Jarque-Bera regularity test.

Case H0: the residues follow a normal distribution, compared to Case H1: the residues do not follow the normal distribution. Its critical area: JB > X^2 (2.05) ⇒ 11.310 > 5.991. Therefore, the H0 hypothesis at a significance level of 5%. Therefore, the residues do not follow the normal allocation.

Next, we proceeded to construct diagrams with the autocorrelations and partial autocorrelations of the MA(1) model residues (without the fixed term). The diagram are below:

Date: 07/20/20 Time: 13:12
Sample (adjusted): 2/20/2020 7/17/2020
Q-statistic probabilities adjusted for 1 ARMA term



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.077	-0.077	0.6301	
		2	0.228	0.223	6.2396	0.012
		3	-0.000	0.032	6.2396	0.044
		4	-0.001	-0.053	6.2396	0.101
		5	-0.041	-0.054	6.4245	0.170
		6	-0.068	-0.067	6.9519	0.224
		7	0.036	0.052	7.0988	0.312
		8	-0.315	-0.297	18.505	0.010
		9	0.142	0.102	20.841	0.008
		10	-0.087	0.062	21.730	0.010
		11	-0.083	-0.156	22.537	0.013
		12	0.120	0.128	24.253	0.012
		13	-0.103	-0.084	25.546	0.012
		14	0.226	0.196	31.778	0.003
		15	-0.056	0.001	32.174	0.004
		16	0.209	0.014	37.624	0.001
		17	-0.019	0.121	37.668	0.002
		18	0.014	-0.112	37.693	0.003
		19	-0.063	-0.132	38.206	0.004
		20	-0.181	-0.092	42.528	0.002
		21	-0.008	-0.077	42.535	0.002
		22	-0.118	0.115	44.400	0.002
		23	0.087	0.039	45.426	0.002
		24	-0.109	-0.070	47.077	0.002

Ljung-Box test in the residuals of the MA(1)

We apply the Ljung-Box control for autocorrelation up to 6th order in the residuals of the MA(1) model.

Hypothesis H0: $\rho 1 = ... = \rho 6 = 0$, versus H1: at least one $\rho j \neq 0$, j = 1,2, ..., 6

The statistic is Q-Stat = 6.9519 < X^2 (6-1.0.05) = 11.070. Therefore, we do not reject the null hypothesis (H0) at a significance level of 5%. So, there is no up to 6th order autocorrelation in the residuals of the MA(1) model. Then we performed a statistical test to find out in which class it exists

autocorrelation.

Statistical test for autocorrelation of order j, 1, = 1,2, ..., 24.

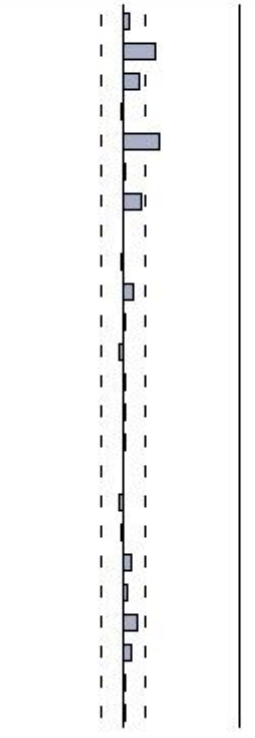
Hypotheses H0: pj = 0, versus H1: pj ≠ 0.

Critical range p^j [-Z0.05 * (1 / √N), Z0.05 * (1 / √N)] = [- 1,96√104, 1,96√104] = [-0,1921, 0.1921].

For j = 8,14,16 we reject the null hypothesis at a significance level of 5%, while for j ≠ 8,14,16 we do not reject Ho.

Therefore, there is 2nd, 8th, 14th and 16th order autocorrelation in MA(1) residuals. The existence of autocorrelation only for 2nd, 8th, 14th, 16th grade is enough to find that the residuals have white noise behavior. Since the behavior of the residuals is white noise, we conclude that the model MA(1) is suitable for NFLXRET time series values. The graph of the autocorrelations and partial autocorrelations for the squares of residuals are below:

Date: 07/20/20 Time: 13:18
Sample (adjusted): 2/20/2020 7/17/2020
Included observations: 104 after adjustments



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.050	0.050	0.2684	0.604
		2	0.284	0.282	8.9863	0.011
		3	0.155	0.142	11.602	0.009
		4	-0.023	-0.120	11.660	0.020
		5	0.332	0.282	23.942	0.000
		6	0.016	0.024	23.972	0.001
		7	0.169	0.022	27.216	0.000
		8	-0.002	-0.104	27.216	0.001
		9	-0.011	-0.017	27.231	0.001
		10	0.094	0.022	28.274	0.002
		11	0.011	0.041	28.289	0.003
		12	-0.026	-0.144	28.373	0.005
		13	0.017	0.031	28.407	0.008
		14	0.022	0.095	28.467	0.012
		15	0.016	0.006	28.498	0.019
		16	-0.002	-0.082	28.499	0.028
		17	-0.029	-0.003	28.608	0.038
		18	-0.023	0.001	28.675	0.052
		19	0.075	0.120	29.413	0.060
		20	0.033	0.005	29.554	0.077
		21	0.124	0.088	31.594	0.064
		22	0.067	0.075	32.191	0.074
		23	0.024	0.009	32.269	0.095
		24	0.022	-0.124	32.339	0.119

We apply the Ljung-Box test for autocorrelation up to 4th order in the squares of the residuals of the MA(1) model.

Hypothesis H0: $\rho 1 = ... = \rho 4 = 0$, versus H1: at least one $\rho j \neq 0$, j = 1,2,3,4

The statistic is Q-Stat = 11.66 > X^2 (4-1.0.05) = 7.815.

Therefore, we reject the H0 hypothesis at a significance level of 5%. Thus there is an autocorrelation up to 4th order in the squares of the residuals of the MA(1) model. Therefore, we have an indication of ARCH / GARCH behavior in its dispersion values disturbing condition.

We test for ARCH errors with the Lagrange Multiplier test (LM test).

Heteroskedasticity Test: ARCH

F-statistic	3.738166	Prob. F(3,97)	0.0137
Obs*R-squared	10.46685	Prob. Chi-Square(3)	0.0150

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 07/20/20 Time: 13:26
Sample (adjusted): 2/25/2020 7/17/2020
Included observations: 101 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000642	0.000246	2.610728	0.0105
RESID^2(-1)	-0.011034	0.102066	-0.108105	0.9141
RESID^2(-2)	0.284905	0.097549	2.920617	0.0043
RESID^2(-3)	0.146548	0.102105	1.435266	0.1544

R-squared	0.103632	Mean dependent var	0.001087
Adjusted R-squared	0.075909	S.D. dependent var	0.001961
S.E. of regression	0.001885	Akaike info criterion	-9.671185
Sum squared resid	0.000345	Schwarz criterion	-9.567616
Log likelihood	492.3949	Hannan-Quinn criter.	-9.629258
F-statistic	3.738166	Durbin-Watson stat	1.930189
Prob(F-statistic)	0.013675		

ARCH test of Heteroskedasticity, 2 lags

From the LM test according to p-value = 0.015 < 0.05. Therefore for 5% level of importance we have ARCH behavior. Also, the case of autocorrelation up to 3 lags indicates that the first and third lags are not statistically significant, according to the p-value (0.914 > 0.05 and 0.154 > 0.05) at a significance 5%. Therefore, we do not have ARCH (3) behavior.

ARMA(1,2)

Next we evaluate the ARMA (1,1) model. It is a model of particular practical importance as with this we can describe thoughtful processes that one would otherwise need
large number of parameters if we used only AR, or only MA
procedures
The self-reciprocating model of ARMA mobile means (1,1), has the form:

$$Y_t = \phi_1 * Y_{t-1} - \theta_1 * \varepsilon_{t-1} + \varepsilon_t$$

where ε_t is the white noise with mean value 0 and scatter σ^2.
The stagnation condition is $|\phi_1| < 1$ (as in AR (1) model)

Adjusting the NFLXRET values to the above ARMA (1,2) model with a fixed term yields the following results:

Dependent Variable: NFLXRET
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 07/20/20 Time: 13:46
Sample: 2/20/2020 7/17/2020
Included observations: 104
Convergence achieved after 13 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003057	0.002683	1.139577	0.2572
AR(1)	-0.569778	0.193452	-2.945327	0.0040
MA(1)	0.261772	0.237106	1.104027	0.2722
SIGMASQ	0.001004	0.000108	9.285836	0.0000

R-squared	0.121129	Mean dependent var	0.002924
Adjusted R-squared	0.094763	S.D. dependent var	0.033963
S.E. of regression	0.032314	Akaike info criterion	-3.987586
Sum squared resid	0.104419	Schwarz criterion	-3.885878
Log likelihood	211.3545	Hannan-Quinn criter.	-3.946381
F-statistic	4.594132	Durbin-Watson stat	1.929575
Prob(F-statistic)	0.004690		

Inverted AR Roots	-.57
Inverted MA Roots	-.26

Regression Analysis for ARMA(1,2) with fixed term

From the regression results we see that the fixed term has p-value = 0.257 > 0.05. So, at a significance level of 5% we do not reject zero assumption, i.e. the constant condition of the model is zero. It also seems that the term MA(1) is statistically insignificant since p-value = 0.272 > 0.05. Subsequently, re-evaluate the ARMA model (1,2) without a fixed term.

Dependent Variable: NFLXRET
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 07/20/20 Time: 13:50
Sample: 2/20/2020 7/17/2020
Included observations: 104
Convergence achieved after 20 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.279771	0.102065	-2.741111	0.0072
MA(2)	0.199000	0.085136	2.337441	0.0214
SIGMASQ	0.000999	0.000108	9.211073	0.0000

R-squared	0.125908	Mean dependent var	0.002924
Adjusted R-squared	0.108599	S.D. dependent var	0.033963
S.E. of regression	0.032066	Akaike info criterion	-4.011761
Sum squared resid	0.103851	Schwarz criterion	-3.935481
Log likelihood	211.6116	Hannan-Quinn criter.	-3.980858
Durbin-Watson stat	1.975356		

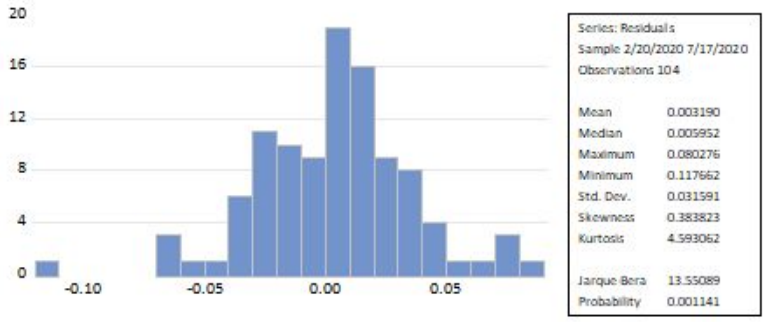
Inverted AR Roots	-.28	
Inverted MA Roots	-.00+.45i	-.00-.45i

Regression Analysis for ARMA(1,2) without fixed term

For $\phi 1$ it holds that p-value = 0,007 < 0,05, so we reject the zero case at a significance level of 5%. Therefore, the $\phi 1$ estimator is statistically significant. The estimated equation of the autoregressive moving average model ARMA(1,2) is:

$$Y_t = -0{,}27 * Y_{t-1} - 0{,}19 * \varepsilon_{t-1} + \varepsilon_t$$

If $|\phi 1| = -0.27 < 1$, we conclude that the NFLXRET time series is stationary.
In order to investigate the distribution of residuals, we calculate the basic characteristics and the histogram of the errors.



Jarque-Bera test of probability for ARMA(1,2)

Next, we perform a Jarque-Bera normality test.

Case H0: the residuals follow a normal distribution, compared to Case H1: the residuals do not follow the normal distribution. Its critical area: JB > X^2 (2.05) ⇒ 13.55 > 5.991. Therefore, the H0 hypothesis is rejected at a significance level of 5%. Therefore, the residuals do not follow the normal distribution.

Next, we proceeded to construct diagrams with the autocorrelations and partial autocorrelations of the ARMA (1,2) model residuals (without the fixed term). The diagram are below:

Date: 07/20/20 Time: 13:56
Sample (adjusted): 2/20/2020 7/17/2020
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.019	-0.019	0.0402	
		2	-0.012	-0.012	0.0559	
		3	0.031	0.031	0.1606	0.689
		4	-0.014	-0.013	0.1820	0.913
		5	-0.048	-0.048	0.4411	0.932
		6	0.019	0.016	0.4815	0.975
		7	-0.021	-0.021	0.5308	0.991
		8	-0.305	-0.304	11.184	0.083
		9	0.152	0.150	13.859	0.054
		10	-0.032	-0.042	13.981	0.082
		11	-0.108	-0.102	15.353	0.082
		12	0.087	0.086	16.255	0.093
		13	-0.067	-0.107	16.791	0.114
		14	0.150	0.205	19.530	0.077
		15	-0.020	-0.070	19.580	0.106
		16	0.172	0.106	23.291	0.056
		17	0.013	0.131	23.314	0.078
		18	0.015	-0.080	23.342	0.105
		19	-0.066	-0.080	23.910	0.122
		20	-0.182	-0.154	28.250	0.058
		21	-0.031	-0.099	28.378	0.076
		22	-0.046	0.098	28.662	0.095
		23	0.105	0.032	30.153	0.089
		24	-0.089	-0.007	31.237	0.091

Ljung-Box test in the residuals of the ARMA(1,2)

We apply the Ljung-Box test for autocorrelation up to 6th order in the residuals of the ARMA(1,2) model.

Hypothesis H0: $\rho 1 = ... = \rho 6 = 0$, versus H1: at least one $\rho j \neq 0$, j = 1,2, ..., 6

The statistic is Q-Stat = 0.4815 < X^2 (6-1.0.05) = 11.070. Therefore, we do not reject the null hypothesis (H0) at a significance level of 5%. So, there is no up to 6th order autocorrelation in the ARMA(1,2) residuals. Then we performed a statistical test to find out in which class it exists autocorrelation.

Statistical test for autocorrelation of order j, 1, = 1,2, ..., 24.

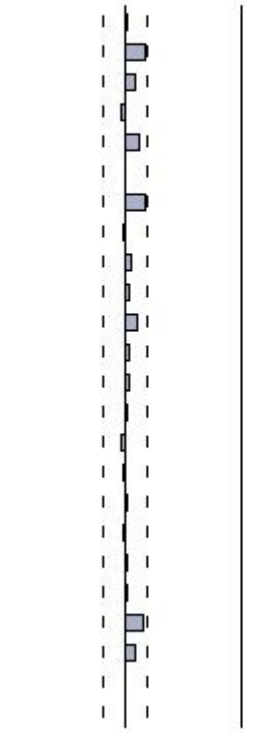
Hypothesis H0: pj = 0, versus H1: pj ≠ 0.

Critical range $\hat{p}j$ [-Z0.05 * (1 / √N), Z0.05 * (1 / √N)] = [- 1,96√104, 1,96√104] = [-0,1921, 0.1921].

For j = 8 we reject the null hypothesis at a significance level of 5%, while for j ≠ 8 we do not reject Ho. Therefore, there is 8th order autocorrelation in the residuals of the ARMA(1,2) model. The existence of autocorrelation only for 8th grade is sufficient to find that the residuals have white noise behavior. Since the behavior of the residuals is white noise, we conclude that the ARMA model (1,2) is suitable for NFLXRET time series values.

The graph of the autocorrelations and partial autocorrelations for the squares of the residuals is below:

Date: 07/20/20 Time: 14:00
Sample (adjusted): 2/20/2020 7/17/2020
Included observations: 104 after adjustments



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.012	0.012	0.0148	0.903
		2	0.175	0.175	3.3238	0.190
		3	0.095	0.094	4.3020	0.231
		4	-0.030	-0.063	4.4019	0.354
		5	0.130	0.101	6.2846	0.280
		6	0.007	0.015	6.2903	0.391
		7	0.187	0.163	10.258	0.174
		8	-0.022	-0.052	10.313	0.244
		9	0.060	0.013	10.730	0.295
		10	0.037	0.009	10.890	0.366
		11	0.106	0.128	12.220	0.347
		12	0.030	-0.028	12.324	0.420
		13	0.038	0.008	12.498	0.487
		14	0.009	-0.050	12.508	0.566
		15	-0.036	-0.018	12.666	0.628
		16	-0.019	-0.057	12.710	0.694
		17	0.009	0.023	12.720	0.755
		18	-0.024	-0.060	12.792	0.804
		19	0.013	0.028	12.814	0.848
		20	0.015	0.008	12.842	0.884
		21	0.158	0.198	16.178	0.760
		22	0.088	0.070	17.209	0.752
		23	-0.002	-0.037	17.209	0.799
		24	-0.001	-0.086	17.209	0.840

Ljung-Box test in the squares of the residuals of the ARMA(1,2)

We apply the Ljung-Box test for autocorrelation up to 4th order in the squares of the residuals in the ARMA(1,2) model.

Hypothesis H0: $\rho 1 = ... = \rho 4 = 0$, versus H1: at least one $\rho j \neq 0$, j = 1,2,3,4

The statistic is Q-Stat = 4,401 < X^2 (4-1,0.05) = 7,815. Therefore, we do not reject the H0 hypothesis at a significance level of 5%. Therefore, there is no autocorrelation up to 4th order in the squares of the residuals in the ARMA(1,2) model.

Information Criteria

In summary, we list the models we evaluated based on the Akaike (AIC) and Schwarz (BIC) information criteria.

	AIC	BIC

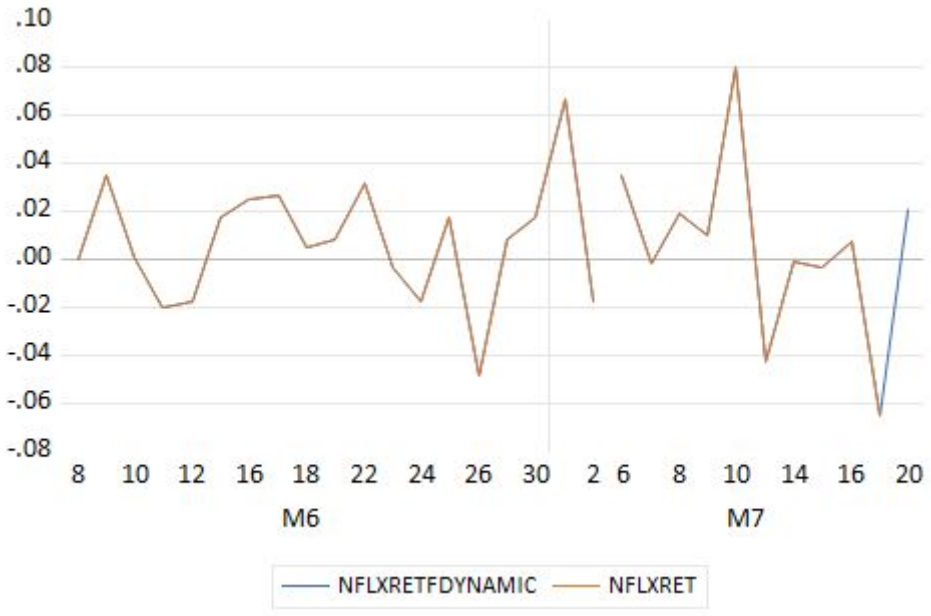
AR(1) without fixed term	-3,9961	**-3,9453**
MA(1) without fixed term	-3,9641	-3,9133
ARMA(1,2) without fixed term	**-4,0117**	-3,9354

According to these criteria the best model is the ARMA (1,2), because it has the lowest value of the information criteria, AIC = -4,0117. The next most suitable model according to the information criteria is AR(1) which has the lowest value with the Schwarz criterion, BIC = -3,9453. The Schwarz criterion is more parsimony.

Prediction

Then, after evaluating the model with ARMA(1,2) we proceed to its prediction.
The basic choice (among other things) is whether EViews will give us static forecasts (static Forecast, a series of rolling single-step-ahead) or dynamic forecasts (dynamic Forecast, multiple-step-ahead).

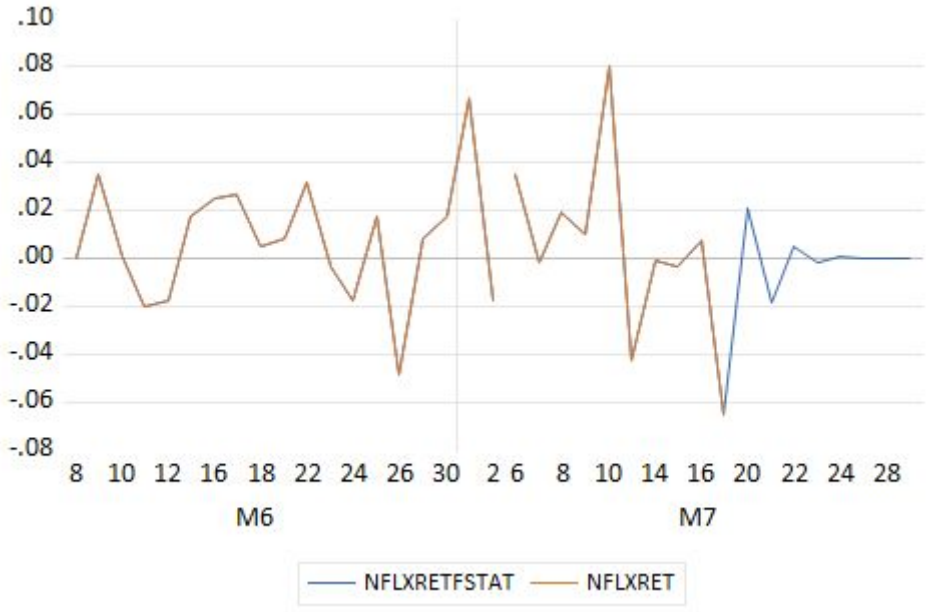
The dynamic forecast is:



Dynamic forecast

where the actual values of returns of Netflix stock are in orange and the predicted returns for the next time are in blue.

The static forecasts is:



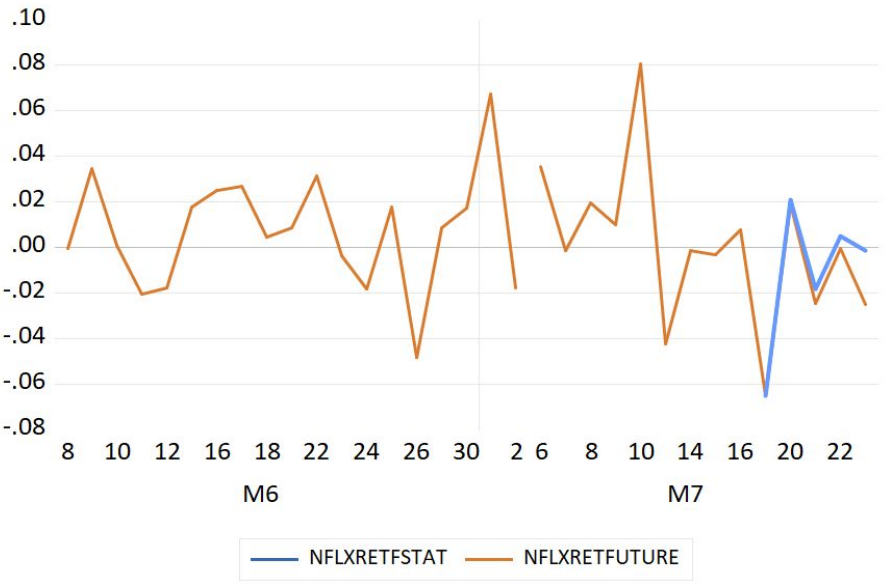
Static forecast

where the actual returns of the Netflix stock are in orange and the predicted returns are in blue.

Remarks:

1. If we are interested in a one-step ahead forecast (forecast for the next time period), both methodologies yield the same result.
2. If we want to forecast for more periods, then in "dynamic forecast" the previous forecasts are used (those that we have already calculated, i.e. the values of the previous forecasts for the dependent variable are used) while in "static forecast" the real values are used for the forecast of the next periods.

In the graph below the real returns of the Netflix share for the period 19/02/2020 to 23/07/2020 are shown in orange and the forecasts with the autocorrelation moving average model ARMA(1,2) for the period 18/07/2020 to 23/07/2020 are indicated in blue.



Forecasting results with real returns of the Netflix stock

We observe that for the next period of time, i.e. for 20/07/2020 the forecast is very good since the real value of the return is 0.020795 which is very close to the expected value of the return which is 0.019108.

The following table shows the other Netflix stock returns:

Netflix returns	**20/07/2020**	**21/07/2020**	**22/07/2020**	**23/07/2020**
Predicted	0,020796	-0,018400	0,005148	-0,001440
Real	0,019108	-0,024502	-0,000571	-0,024989

The last date from our data is on 17/07/2020. The return of the Netflix stock on 17/07/2020 is -0.065227. Therefore, given the time series forecast, the investor will have to buy the stock because he expects its return to increase. Finally the real price shows us that indeed the return on the stock increased on 17/07/2020 so the model seems to be appropriate.

Google Trends with ARIMA models

Google Trends and ARIMA methods are two different ways to predict the return of a stock. Believing that we would have a better performance in forecasting we combined these two methods into one. We used the returns from the first version of the prediction with ARIMA models, due to the fact that it had better performance and on the dates on which we did not have a prediction we replaced the returns by the search term “netflix bear”.

The results of the analysis are summarized below.

Google Trends and ARIMA models			
everyday forecasting	True	False	Sum
80,49% success	66	16	82

Discussion

After predicting the movement of Netflix return stock with three different ways we came to the following results:

	Google Trends	ARIMA models	Google Trends	ARIMA models	Google Trends + ARIMA models
Forecasting	everyday	everyday	not everyday	not everyday	**everyday**
Performance	56,1%	74,39%	62,07%	86,44%	**80,49%**
Sum	82	82	58	59	82

We notice that the method which has better performance is the ARIMA models but not for all the days --> (86,44%)

For daily forecasting, the better model is that which combines Google Trends with ARIMA models → (80,49%)

Conclusions

The most important statement is that we live in a world with plenty of data. Initially, to predict the movement of a share is both difficult and significant. It is widely known that the Time Series Analysis, being one of the classic method forecasting, has satisfactory results. However, the combination of a technical analysis with the Sentiment analysis seems to have better results. Moreover, analysis of sentiment steams from opinion mining and that it is easy to analyze through Google Trends. Every search term is an extra data in order to achieve a better forecast. Analyzing gradually in this work we come to the fact that Sentiment Analysis is not sufficient to predict the movement of netflix stock but the combination of this analysis with the classic method of time series, using ARIMA models, ends up in very good results, on a daily basis. To conclude, Google Trends with ARIMA models predict a respectable percentage of 80.49% of the Netflix share movement on the stock market.

References

- Alsing, O., & Bahceci, O. (2015). *Stock Market Prediction using Social Media Analysis*. http://www.diva-portal.se/smash/get/diva2:811087/FULLTEXT01.pdf
- Cochrane, J. H. (2014). Eugene F. Fama, efficient markets, and the Nobel Prize. *Finance*. https://review.chicagobooth.edu/magazine/winter-2013/eugene-fama-efficient-markets-and-the-nobel-prize
- DataVedas. (2018). *Trend, Seasonality, Cyclicity and Irregularity*. INTRODUCTION TO TIME SERIES DATA. https://www.datavedas.com/introduction-to-time-series-data/
- Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the variance of United Kingdom Inflation. In *Econometrica* (50th ed., Vol. 4, pp. 987-1008). http://www.econ.uiuc.edu/~econ536/Papers/engle82.pdf
- Engle III, R. F. (2003). Biographical. https://www.nobelprize.org/prizes/economic-sciences/2003/engle/biographical/
- Fothergill, D. (2016). *Three golden rules for forecasting*. https://marketingland.com/three-golden-rules-forecasting-176220
- Glen, S. (2016). *Jarque - Bera Test*. Statistics How To. https://www.statisticshowto.com/jarque-bera-test/
- Glen, S. (2018). *Ljung Box Test*. Statistics How To. https://www.statisticshowto.com/ljung-box-test/
- Google Trends. (n.d.). *Search term "Netflix", "netflix bull", "netflix bear"*. https://trends.google.com/trends/?geo=US
- Han, J., Kamber, M., & Pei, J. (2000). *Data Mining Concepts and Techniques* (3rd ed.). Morgan Kaupmann. http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf
- Hayes, A. (2020). *Stock*. Investopedia. Retrieved Σεπτέμβριος 26/9/2020, 2020, from https://www.investopedia.com/terms/s/stock.asp
- Hayes, A. (2020). *Stock*. Investopedia. https://www.investopedia.com/terms/s/stock.asp
- Huang, W. (2003). Forecasting NIKKEI 225 index with support vector machine. *https://www.researchgate.net/publication/268658458_Forecasting_NIKKEI_225_index_with_support_vector_machine*.

- *Information is data with context. Context sparks insight. Insight creates opportunity.* (n.d.). Business Intelligence and Data Analytics. http://www.winter-park.com/data
- Jenkins, A. (n.d.). *Netflix Q2 earnings preview: Another huge quarter fueled by the coronavirus?* https://fortune.com/2020/07/15/netflix-q2-earnings-stock-nflx-2020-coronavirus-pandemic/
- Joshi, K., Bharathi, P. H. N., & Rao, P. J. (2013). Stock trend prediction using news Sentiment analysis. https://arxiv.org/ftp/arxiv/papers/1607/1607.01958.pdf
- Kotow, E. (2019). *Bear Market Vs. Bull Market*. Hedge Trade. https://hedgetrade.com/bear-market-vs-bull-market/
- Larose, D. T. (2014). *Discovering knowledge in data: an introduction to data mining (Vol. 4). John Wiley & Sons.* file:///C:/Users/dimto/Downloads/EBook%20Data%20Mining.pdf
- Levy, R. A. (1967). Relative Strength as a Criterion for Investment Selection. *22*(4), 595-610. http://www.technicalanalysis.org.uk/history/Levy67.pdf
- Mallikarjuna M., & Rao, P. R. (2019). Evaluation of forecasting methods from selected stock market returns. https://jfin-swufe.springeropen.com/articles/10.1186/s40854-019-0157-x
- Mudinas, A., Zhang, D., & Levene, M. (2019). *Market Trend Prediction using Sentiment Analysis: Lessons Learned and Paths Forward*. https://arxiv.org/abs/1903.05440
- Naftemporiki. (2019). *Η πιο πετυχημένη μετοχή της δεκαετίας*. Naftemporiki. https://www.naftemporiki.gr/finance/story/1546296/netflix-i-pio-petuximeni-metoxi-tis-dekaetias
- Preneur, S. (2019). *The curious case of Warren Buffet and the Efficient Market Hypothesis*. https://steemit.com/steemleo/@shanghaipreneur/the-curious-case-of-warren-buffett-and-the-efficient-market-hypothesis
- Schwartz, H. (2018). *Using google trends to predict stocks*. Seeking Alpha. https://seekingalpha.com/article/4191521-using-google-trends-to-predict-stocks
- Xu, S. Y. (2012). Stock Price Forecasting Using Information from Yahoo Finance and Google Trend. *https://www.econ.berkeley.edu/sites/default/files/Selene%20Yue%20Xu.pdf*.
- Yadav, R., Kumar, V. A., & Kumar, A. (2019). News-based supervised sentiment analysis for prediction of futures buying behaviour. 157-166. https://www.sciencedirect.com/science/article/pii/S0970389619301569
- Yahoo Finance. (n.d.). Netflix, Inc, (NFLX). https://finance.yahoo.com/quote/NFLX/
- Zweig, J. (2011). Making Sense of Market Forecasts. https://www.wsj.com/articles/SB10001424052748703675904576064320900295678

Appendices

Ημερομηνίες	Πραγματικές αποδόσεις		Προβλεπόμενες αποδόσεις με ARIMA models		Προβλεπόμενες αποδόσεις με Google Trends
4/17/2020	-0,036911		**-0,036911**		
4/20/2020	0,034353	+	0,001151	+	-
4/21/2020	-0,008366	-	-0,003477	-	-
4/22/2020	-0,028606	-	0,00181	+	+
4/23/2020	0,012529	+	0,003821	+	+
4/24/2020	-0,004008	-	-0,003775	-	-
4/27/2020	-0,008494	-	-0,001804	+	-
4/28/2020	-0,041649	-	-0,000129	+	-
4/29/2020	0,019959	+	0,001862	+	-
4/30/2020	0,019326	-	-0,004486	-	-
05/01/2020	-0,010909	-	-0,000031	+	-
05/04/2020	0,031016	+	0,002476	+	+
05/05/2020	-0,008105	-	-0,002493	-	-
05/06/2020	0,022558	+	0,002519	+	-
05/07/2020	0,005227	-	-0,001661	-	-
05/08/2020	-0,002245	-	0,001565	+	-
5/11/2020	0,011411	+	0,000203	-	+
5/12/2020	-0,019749	-	-0,001021	-	+
5/13/2020	0,014937	+	0,002084	+	+
5/14/2020	0,008397	-	-0,002308	-	-
5/15/2020	0,027695	+	0,001004	+	+
5/18/2020	-0,003545	-	-0,00074	-	-

5/19/2020	-0,003403	+	0,002188	+	-
5/20/2020	-0,007472	-	-0,017887	-	-
5/21/2020	-0,02551	-	-0,001099	+	+
5/22/2020	-0,015885	+	0,000795	+	-
5/26/2020	-0,033891	-	0,000186	-	+
5/27/2020	0,012344	+	0,000775	+	+
5/28/2020	-0,015361	-	-0,003231	-	-
5/29/2020	0,015214	+	0,001178	+	-
06/01/2020	0,014748	-	0,000104	-	+
06/02/2020	0,003263	-	-0,000769	-	-
06/03/2020	-0,012497	-	0,000658	+	+
06/04/2020	-0,018106	-	0,001655	+	+
06/05/2020	0,012719	+	0,000092	-	-
06/08/2020	-0,000262	-	-0,003211	-	-
06/09/2020	0,034709	+	0,000587	+	+
6/10/2020	0,000991	-	-0,001916	-	-
6/11/2020	-0,02053	-	0,001893	+	-
6/12/2020	-0,0176	+	0,00176	-	-
6/15/2020	0,017772	+	-0,001343	-	+
6/16/2020	0,024982	+	-0,003634	-	-
6/17/2020	0,026689	+	0,000192	+	-
6/18/2020	0,00469	-	0,002354	+	+
6/19/2020	0,008558	+	0,001435	-	+
6/22/2020	0,031561	+	0,00032	-	+
6/23/2020	-0,003803	-	-0,002562	-	-
6/24/2020	-0,018037	-	0,002265	+	+
6/25/2020	0,017604	+	0,00286	+	+
6/26/2020	-0,048314	-	-0,004355	-	-
6/29/2020	0,00866	+	0,002114	+	+

6/30/2020	0,01744	+	-0,000702	-	-
07/01/2020	0,067247	+	-0,002014	-	-
07/02/2020	-0,018017	-	-0,002854	-	-
07/06/2020	0,03548	+	0,006741	+	+
07/07/2020	-0,001316	-	-0,00084	-	-
07/08/2020	0,019507	+	0,002285	+	+
07/09/2020	0,009905	-	-0,000762	-	-
7/10/2020	0,080688	+	0,002238	+	-
7/13/2020	-0,042334	-	-0,003831	-	-
7/14/2020	-0,00118	+	0,006864	+	-
7/15/2020	-0,003086	-	-0,003136	-	+
7/16/2020	0,007893	+	0,00018	+	+
7/17/2020	-0,065227	-	-0,00124	-	-
7/20/2020	0,019108	+	0,020796	+	-
7/21/2020	-0,024502	-	-0,004183	-	-
7/22/2020	-0,000571	+	-0,003061	+	-
7/23/2020	-0,024989	-	-0,001892	+	+
7/24/2020	0,00601	+	0,003981	+	-
7/27/2020	0,031637	+	-0,003139	-	-
7/28/2020	-0,014405	-	-0,004825	-	-
7/29/2020	-0,00825	+	0,005087	+	-
7/30/2020	0,002725	+	0,003312	-	-
7/31/2020	0,00634	+	-0,0043	-	+
08/03/2020	0,019923	+	0,000292	+	-
08/04/2020	0,022101	+	0,001956	+	-
08/05/2020	-0,014775	-	0,001446	-	-
08/06/2020	0,013881	+	0,000898	-	-
08/07/2020	-0,028188	-	-0,002831	-	-
8/10/2020	-0,022942	+	0,002207	+	-

8/11/2020	-0,034031	-	0,002079	-	-
8/12/2020	0,01829	+	-0,00043	-	-
8/13/2020	0,012325	-	-0,004698	-	-

➔ "**+**" means that the market is **bullish** for the Netflix stock

➔ "**-**" means that the market is **bearish** for the Netflix stock