Informatica & Economie

The correlation between market and public sentiment; the intercommunication of the Russo-Ukrainian war

Yasin Koçak

Supervisors: Suzan Verberne & Arno Knobbe

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl 26/05/2023

Contents

Abstract

The rise of social media has much influence on the communication in the world. The financial world also participates in the communication through social media and therefore it is interesting to look at the relationship between the financial world and social media. In this research the relationship between market and public sentiment is analyzed. Specifically, the correlation between stock market changes and two public sentiments, the economic public sentiment and the Russo-Ukrainian war public sentiment, is analyzed. In our aim to answering our research questions we construct two datasets: one for the market sentiment through stock market changes and one for the public sentiment(s) through sentiment analysis on tweets related to the economy and the Russo-Ukrainian war. This research resulted in several (significant) correlations between market and public sentiment, which may indicate for a relationship between the financial world and Twitter.

1 Introduction

The impact and influence of social media in the modern world is both evident and significant. This influence has its positive and negative sides. People can easily share information, worldwide communication has never been as easy as today. But on the other hand, people can mislead others by spreading misinformation and creating an erroneous perspective on certain issues that are going on in the world. It is certainly good for people in general to be aware of the influence of social media on theirs and others perspective on events happening in the world. To what extent people react to the information shared on social media platforms and vice versa is an interesting and thoroughly investigated topic [Ame15]. In this research we are going to zoom in to exactly such phenomenons. In addition, we will zoom in on the Ukrainian-Russo war and its intercommunication with the public and market sentiment.
Important to clarify is that this research solely focuses on correlation and not causation. Thus, this research does not aim to explore a causal relationship. Our aim is to explore the correlation between public and market sentiment by measuring the statistical relationship.

1.1 Market sentiment

Since the beginning of the stock market, as in many financially driven sectors, the question was raised which factors and in which manner the prices are affected. There have been a lot of changes, which resulted in the stock market being way more complicated than it ever was before. A stock, in its simplest explanation, is nothing more than a (very small) ‘share’ of a publicly traded company. Like in many financial markets, the price of stocks is driven by supply and demand. However, the stock market is complicated and the supply and demand mechanism goes along with it. Therefore it is very difficult, if not impossible, to predict stock prices. Nevertheless, there are certain trends which can be recognised in the stock market. A market can be ‘bull’ (upward trend) or ‘bear’ (downward trend) and this varies from time to time [ADA22]. Therefore, the sentiment of the market can be evaluated for a certain time period. In a year, a month, or even in an hour there happens a lot to the price of a stock. In this thesis we define this ‘market sentiment’ as the average (daily) change of three stock market indices, namely:

1. S&P500
2. NASDAQ
3. DJIA

These indices are the three major indices of the stock market and are a good representation of the change of the stock market in general [Tre21]. Besides the stock price changes, we will also look at the trading volumes (activity on the stock market) on each day.
Which factors influence the stock market cannot be summarized or explained in one sentence. In a certain way you might say that every financial transaction, or for that matter, action, influences the stock market. Buying the newest IPhone, eating at the MacDonald’s or buying a Dove shampoo, with these three transactions someone already influenced three stock prices. Of course, this is a big exaggeration and the influence is insignificant, but while talking about the stock market it is important to know that the mechanism behind it is indeed very complicated. The performance of a company is important for the trend of the stock price of that specific company, although, how

important is a matter of dispute [BGM90]. There are many other important factors which affect the stock prices. Macroeconomic factors like the state of the economy, but also the activity/reaction of central banks on certain events. For example, political events both national and international [CPS88]. Also, trading volumes, by definition, affect the stock market prices. Further, as a sort of 'multiplier effect', the market sentiment can influence itself.

1.2 Public sentiment

Since the rise of social media there is another factor that has a huge impact on the financial world. Social media brings a whole new scope of influence with it. There are a lot of different platforms, accessible for everyone, to discuss and interact with each other. People can advise each other and share their views and opinions on certain economic or general issues (which may have an underlying economic impact). Therefore social media creates a whole new domain of communication where certain sentimental trends can be developed or current opinions properly can be debunked [May08]. Obviously, these interactions influence both people and economy. Public sentiment is defined as the view, opinion or feeling of the average person of a certain country, group or society [Dic]. Although public sentiment can be measured in a variety of ways, in this thesis we will solely focus on the public sentiment on Twitter i.e. on the views or emotions people express on Twitter. We define the overall interactions and communication on Twitter as the public sentiment. To measure the public sentiment on Twitter we will use user-tweets. Subsequently, to measure the sentiment in these tweets we will use sentiment analysis. Sentiment analysis is a natural language processing task in which through text analysis the sentiment of certain information can be extracted [Mej09]. Sentiment analysis is used in various fields and a known phenomena in the finance world. In this research we measure two different public sentiments using a dataset we crawled from Twitter:

1. The public sentiment about the economy
2. The public sentiment about the Russo-Ukrainian war

1.3 Russo-Ukrainian war

The Russo-Ukrainian war started in 2014. It is an ongoing conflict which started with the annexation of the Crimea by Russia to protect the Russian population from the new government formed in Ukraine. Since then tension went up and down, there have been several ceasefire agreements, but also thousands of deaths on both sides. The major escalation of this war happened on the 24$^{\text{th}}$ of February 2022, when Russia invaded Ukraine. The economic consequences of this invasion of Ukraine continue to be significant. Many (western) countries reacted with sanctions on Russia. Because Russia is one of the world's largest and most powerful countries, the impact of those sanctions were enormous [Ser22]. There are of course many other wars going on in the world, but the invasion of a European country, that was on the edge of entering the NATO, has (logically) caused for a more severe reaction from the western countries. Because of the huge economic impact from this invasion, one might wonder how a war like this affects the stock market in general. For this and many other reasons, it is also interesting to look at what the public sentiment is on the Russo-Ukrainian war in general. Therefore we will also look at the intercommunication of Russo-Ukrainian war with the market and public sentiment.

1.4 Research questions

The usage of social media is expected to grow in the future, which means interesting questions could be asked: How broad is its impact on the world? What changed with respect to earlier time periods? How much influence is there exactly by social media? Also, because the world is becoming more united as 'one big country' in several ways, through for example social media, the impact of a war can be of an order that a whole economy and the underlying markets can collapse or recess. As mentioned, our focus in this research is on the market sentiment based on the change of three different stock market indices and the public sentiment based on tweets through sentiment analysis. In this thesis we will look at the correlation between these two sentiments and the intercommunication of the Russo-Ukrainian war, with both the public and market sentiment. Our collected data consists of the sentiment values, both market and public. The data covers the period from 2021-12-24 until 2022-06-23.
In the end, our aim is to answer the following research questions:

- What is the correlation between the economic public sentiment and market sentiment?
- What is the correlation between the Russo-Ukrainian war public sentiment and market sentiment?
- What is the correlation between the Russo-Ukrainian war public sentiment and the economic public sentiment?

1.5 Thesis overview

In this thesis we will answer and evaluate our research questions in the following way. In Section 2 we will discuss, evaluate and look into related work about the impact of social media in general and its impact on the stock market, in particular through sentiment analysis. Also, we will look at the impact of the Russo-Ukrainian war and its coherence with the rest of the world's economies. In Section 3 we clarify our methods and the choices made in our research. Section 4 is dedicated to our results, which conclusions can be drawn and how to interpret certain results. In Section 5 we investigate points of improvement and shortcomings in our research. Lastly, in Section 6, we state a general conclusion and further research that might be done.

2 Related work

The rise of social media and its global impact is, in a superficial way, clear and perspicuous. Research shows that 59% of all adults use social media, with an average of 2 hours and 29 minutes on a daily basis [Cha22]. Analysis of social media and its impact are broadly done in the past decade. In many fields and in variety of precision: from the effect of social media on the psychological well-being [CL16] to the impact of social media on the academic performance of students [MRQ+15]. For good reasons these researches are taken very seriously, because social media and its impact are a relatively new phenomenon and therefore, interesting. As with many new 'phenomena' the question might arise in which manner a certain money flow can be optimized, or, maybe even better, invent a new one. In the context within this research this could be the knowledge of how tweets and market prices are correlated, from which you can optimize a trading strategy.

2.1 Market sentiment

Before we dive in the impact of Twitter, it is useful to have a basic understanding on how the stock market operates and what can be said about future prices according to experts. There are two main theories about prediction of the prices in the stock market. The chartist theorists claim that the past behaviour of prices are a reliable source to predict future prices. On the other hand, the theory of random walk suggest that future prices are independent of past prices [Fam65]. The efficient market hypothesis confirms the theory of random walks and states that stock prices contain all the information i.e. predicting trends through technical analysis is of no fundamental worth. Of course, prices cannot be solely independent from the past neither can they be completely depend on it. It is a point of dispute and one might ask which role public sentiment plays in the whole discussion. Regarding to the market sentiment it is important to concertize which theories there are about the trends of the market and how they are analysed. Market trends can be divided in three categories. The first one is the commonly known trend of the market, known as bull or bear, which is a trend for a period of one year or more. A bull market is a rising market where people intend to buy, whereas a bear market is the opposite, a falling market where people intend to sell. Within the bull market a pullback is the second category. Such pullbacks are a tendency to sell and are a short period in which the market is falling, lasting from a couple of weeks to a couple of months. Within the bear market this phenomenon is called a 'rally' and this is a tendency to buy, resulting in a short rise of the market, again from a couple of weeks to a couple of months. At last, the third category are trends lasting for at most three weeks, these can be classified as 'noise' in the market [ADA22].

2.2 The relation between public sentiment and the stock market

The question arises what can be said about how different human discourses could affect or influence financial behaviours, or at a larger scale, financial markets. Stone et al. described that words and sentences are substantial human artifacts and are good indicators of financial activities and behaviour [SDS66]. Researchers try to understand how sentiments impact individual decision making and subsequently the financial markets in general. You could speak about two kinds of sentiments: investor sentiment and textual sentiment. Investor sentiment is the perception and view of investors on certain investments, the risks within them and the potential outcomes, but also on the market in general. By analysing the raw data of texts we could obtain the textual sentiment. Textual sentiment is in some sense more objective than investor sentiment, due the biases a particular investor might have. Textual sentiment is directly measurable and therefore it gives a more realistic view on the conditions of the market. For this research we only look into the textual sentiment. Sentiment analysis and the impact of certain texts on the market have been done in various manners. Antweiler and Frank analysed millions of posts on internet stock messages boards about the companies represented in the DJIA. They conclude that there is financially relevant information in these posts, although their research was not solely focused on the sentiment of the messages, but also on for example, the quantity of posts on a day and the effects thereof [AF04]. One of their conclusions was that activity on the message boards and trading volume are positively related. Analyses on media (channels) and the sentiment within their news is also a broad field of research. Garcia studied the effects of financial news from the New York Times and its relationship on asset pricing and stock returns [Gar13].

The popularity of Twitter has increased over the past 10 years. Because of this the amount of text data has become larger, which resulted in a more interesting dataset to use. Sentiment analysis were already done on blogs or product reviews, but with the rise of Twitter, there came another domain to analyse the text sentiment [VC12].

In recent years there has been done a variety of research on micro blogs like Twitter and its impact on the stock market. Research investigating the correlation between Twitter mood and the stock market, specifically on the predictability of one on the other, show different results through different methods. Bollen et al. used two mood tracking tools, OpinionFinder that gives a positive and negative mood, and GPOMS that measures mood in terms of 6 dimensions. Results show the 'calm' dimension of the GPOMS is indeed predictive, and thus correlated for the stock market data [BMZ11]. In a similar research Mittal and Goel observed that the calmness and happiness have a predictive relationship with the DJIA values. With these moods as feature set, a Self Organizing FNN performs good in predicting the stock market values of the DJIA [MG12]. In our research we will evaluate the sentiment of tweets in terms of a negative and positive score i.e. negative and positive sentiment values. Afterwards we will look at the relationship with the market sentiment.

2.3 The economic impact of the Russo-Ukrainian war

The impact of the Russo-Ukrainian war is felt worldwide. Both economically and politically. Schneider et al. stated that international markets react negatively to war, though not all conflicts affect the stock market in the same way. This may be due the difference in interest of one conflict compared to another. Schneider et al. stated that the stock market reaction is dependent on the anticipated damage and expected substantial consequences an international conflict may have. Also, international conflicts influence the volatility of the stock market in general [ST06]. The Russo-Ukrainian war is quite recent and therefore there has not been done much research on its effects on the stock market. Also, because the war is still going on, concrete conclusions on the ultimate effects cannot be drawn. However, there is some research on the impact of the Russo-Ukrainian war and the expected course the economy tends to go. In April 2022 the world bank forecasted that the economy of the region of Europe and Central Asia will shrink by 4.1 percent in 2022, while the pre-war forecast was a growth of 3 percent [Gro22]. Another study on the Russo-Ukrainian war confirms the conclusions of Schneider et al. Lo et al. stated that the financial markets increased in volatility and lowered asset returns, mostly due to the dependence on Russian commodities [LMBS22].

3 Methods

3.1 Data

3.1.1 Datasets

In our research we collected three (main) datasets, all covering the period from 2021-12-24 until 2022-06-23. Also, each dataset has been divided into three seperate datasets: one dataset of two months just before the war started, one dataset of two months right after the invasion and one dataset covering the 3rd and 4th month after the invasion. These smaller datasets will be used to evaluate the sentiments of each different period. A good overview of our research methodology is given in figure 1. Our main datasets are:

- The average percentage change of the stock market on each day.
- The negative and positive sentiment values of tweets about the economy, i.e. economic sentiment.
- The negative and positive sentiment values of tweets about the Russo-Ukrainian war, i.e. Russo-Ukrainian war sentiment.

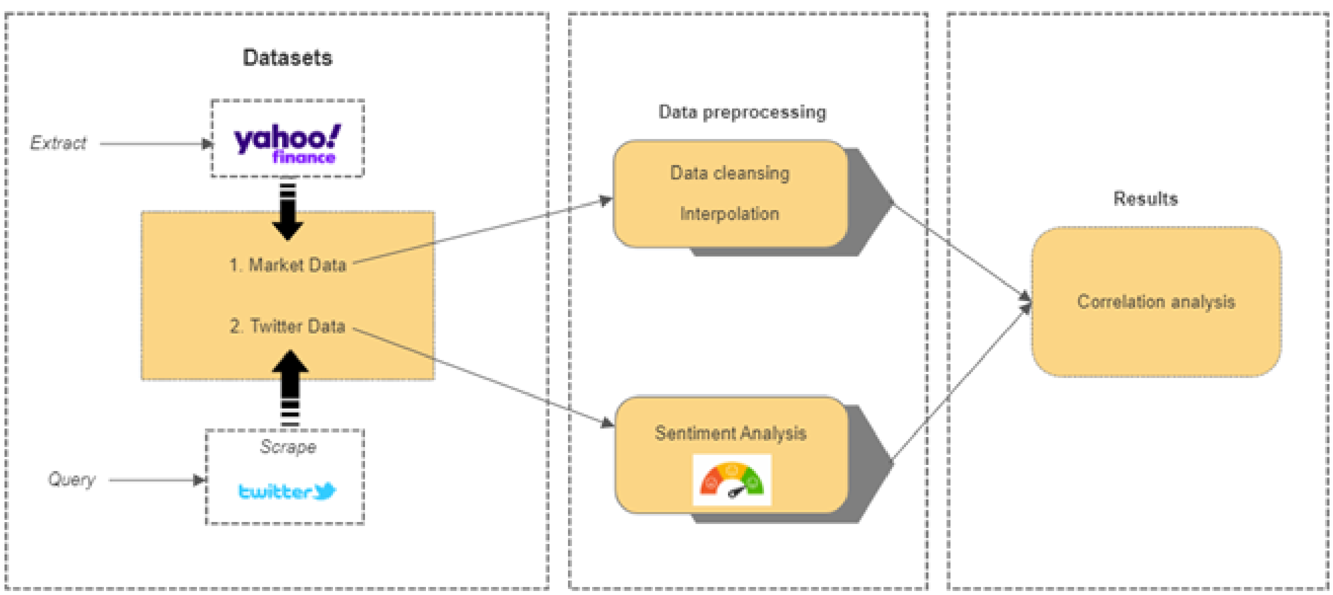

Figure 1: Research methodology

3.1.2 Stock market data pre-processing

To obtain the average percentage change of the stock market of each day we used, as said, three very important stocks: S&P500, NASDAQ and DJIA. First we obtained, of each stock seperately, the open, close, adjusted close, high and low values. We extracted these values from Yahoo!Finance. Because the market is closed in the weekend and on holidays, this data is not available. In order

to make the dataset complete, we approximated the values using the quadratic interpolation[1]. How the missed values should be approximated, is a point of dispute. How the stock market behaves, or would have behaved if it was open in the weekend, is impossible to predict with 100% accuracy. As discussed in Section 2, the stock market is not characterized with perfectly dependence or independence on its past values [Fam65]. Mittel and Goel used the (predominantly) concave character of the stock market to estimate the missing values [MG12]. We, however, think an quadratic approximation is more justifiable. The stock market certainly is not linear and most of the time it is not following a (very)high-order polynomial function as well. Also it is not concave all the time and convexity occurs significantly. It is somewhere in between and therefore we argue the quadratic approximation is the best interpolation method to use for the missing values.
Because of the (large) difference in the prices of each stock, we could not use the nominal price change of each stock, as it would be disproportional. Therefore, we calculated the percentage price change of each stock on each day, using the open and adjusted close values. The preference for the adjusted close values is because of its better representation of the real intrinsic value of the stock. Although it slightly differs, its important to make this distinction. Subsequently, we took the average of the three percentage changes on each day for the three stocks and interpreted this average as the average percentage change of the stock market as a whole. For the market volume we took the average of the three stocks. This market volume is not the change compared to the previous day, but just the volume that has been traded on a single day. This volume is the average quantity of trades of the three stocks.

3.2 Twitter data and sentiment analysis

To obtain the sentiment on Twitter for both the economy and the Russo-Ukrainian war, first tweets had to be obtained. To obtain the tweets we used the snscrape module[2] to scrape them from Twitter. For the economic sentiment as well as the Russo-Ukrainian war sentiment we used seedterms to select the right tweets. This way we will only scrape the tweets relevant to our research.
To get the tweets about the economy we compiled a set of 20 seedterms, which are all economic related. To get a general view of the economy and not solely the opinion of traders we choosed a variety of seedterms. Some are mainstream economic terms like 'income' or 'bank', which can be used by any person to express their feeling i.e. sentiment about the economy, or at least, an economic domain. On the other hand, some terms are more specifically stock market related, like 'stockbroker' or 'DJIA'. For each seedterm we retrieved 25 tweets per day to get a total of 500 tweets per day. These 500 tweets will give us a broad and variated perspective on how people (on Twitter) write about the economy. In table 1 we list the economic seedterms.

[1]We used the interpolate() function of `https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.interpolate.html`

[2]`https://github.com/JustAnotherArchivist/snscrape/blob/master/snscrape/modules/twitter.py`

Economic seedterms			
S&P500	dividend	debt	bank
DJIA	stockbroker	finance	dividend
NASDAQ	inflation	financial	liquidity
forex	economy	dividends	recession
stock	crypto	interest	income

Table 1

For the tweets about the Russo-Ukrainian war we used the same strategy. This time we used 8 seedterms, all Russian, Ukrainian or Russo-Ukrainian related. For each seedterms we scraped 50 tweets per day, which gives a total of 400 tweets. Obviously in times of war and tension between two countries, most of the tweets about these countries will go about the war and the (possible) effects on people their lives due to this war. Therefore the seedterms in table 2 will give us a good perspective of people their feelings about the Russo-Ukrainian war.

Russo-Ukrainian War seedterms	
Russia	Ukrainian
Ukraine	Russian
Russo-Ukrainian War	Zelensky
Crimea	Putin

Table 2

After we scraped one dataset with 500 tweets per day and one dataset with 400 tweets per day, both for the period from 2021-12-24 until 2022-06-23, we had to obtain a sentiment score of each day. We used the roBERTa-base model[3] for our sentiment analysis, together with two classes from the transformers package. From the transformers package we used the Autotokenizer and AutoModelForSequenceClassification classes. The AutoTokenizer class is used to prepare the inputs, i.e. tweets, for the model. The AutoModelForSequenceClassification class is used for the sequence classification of each tweet.
The roBERTa-base model was pretrained on around 58 million tweets and finetuned for sentiment analysis on the TweetEval benchmark and gives, together with the softmax function, the probability which sentiment a tweet has. The generated probabilities are a categorical distribution with three categories: negative, neutral and positive. The probabilities of negative and positive will function as our 'sentiment values' and form the foundation of our public sentiment. The neutral probabilities

[3] `https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment`

will not be taken into account. We first calculated the average sentiment value of all the individual sentiments values of each tweet for each day. We call this the negative/positive economic/Russia-Ukrainian war sentiments. Also, we took the ratios of the negative and positive sentiment values. This way we get values which represents both the negative and positive sentiment of a tweet. We also took the average of the sentiment values of the past 3 days. However, we chose to remove these values from our dataset. We will discuss this in section 3.3. In the end, we get a total of 8 public sentiment values:

- Positive economic sentiment (Pos-eco)
- Negative economic sentiment (Neg-eco)
- Positive Russo-Ukrainian war sentiment (Pos-war)
- Negative Russo-Ukrainian war sentiment (Neg-war)
- Positive economic sentiment/negative economic sentiment (P/N-eco)
- Negative economic sentiment/positive economic sentiment (N/P-eco)
- Positive Russo-Ukrainian war sentiment/negative economic sentiment (P/N-war)
- Negative Russo-Ukrainian war sentiment/positive economic sentiment (N/P-war)

A P/N ratio will increase if the positive sentiment increases or the negative sentiment decreases. On the contrary, the N/P ratio increases if the negative sentiment increases or the positive sentiment decreases. Therefore these ratios are dependent both on the negative and positive sentiments. Now we generated two datasets we can use to calculate correlations: The economic sentiment scores and the Russo-Ukranian war sentiment scores, both for the same period and with the same strategy.

3.3 Correlation Analysis

In our attempt to answer our research questions, we used correlation analysis. We tried to find the best correlation test and looked if our data is fit to do any correlation test in the first place. Due to investigation, we selected the Spearman correlation test. First of all, the Spearman correlation test tests for the monotonicity between two variables. This is in contrast with the Pearson correlation test, which assumes linearity between the variables. Also, Spearman can handle potential outliers. Furthermore, Spearman does not require to have a normal distributed dataset. In fact, Spearman is a nonparametric statistic, which does not assume anything about the underlying data distributions [RLPM15]. With our datasets, to use the Spearman correlation coefficient, we must check two criteria of our data. First we had to check for the monotonicity of all the used variables. A relationship is monotonic when there is a consistent increase or decrease of the datapoints. In other words, a 'bell shaped' or 'U-shaped' scatterplot will show a non-montonic relationship. To check for monotonicity in our data we plotted all the variables against eachother. Due to logical compactness reasons we just show four graphs and their monotonicity, to give an idea of the monotonicity of our data, shown in figure 2. Our data is quite noisy, but clearly not U-shaped or bell-shaped. Thus for monotonicity, we conclude our data is monotone enough to use for the Spearman correlation coefficient.

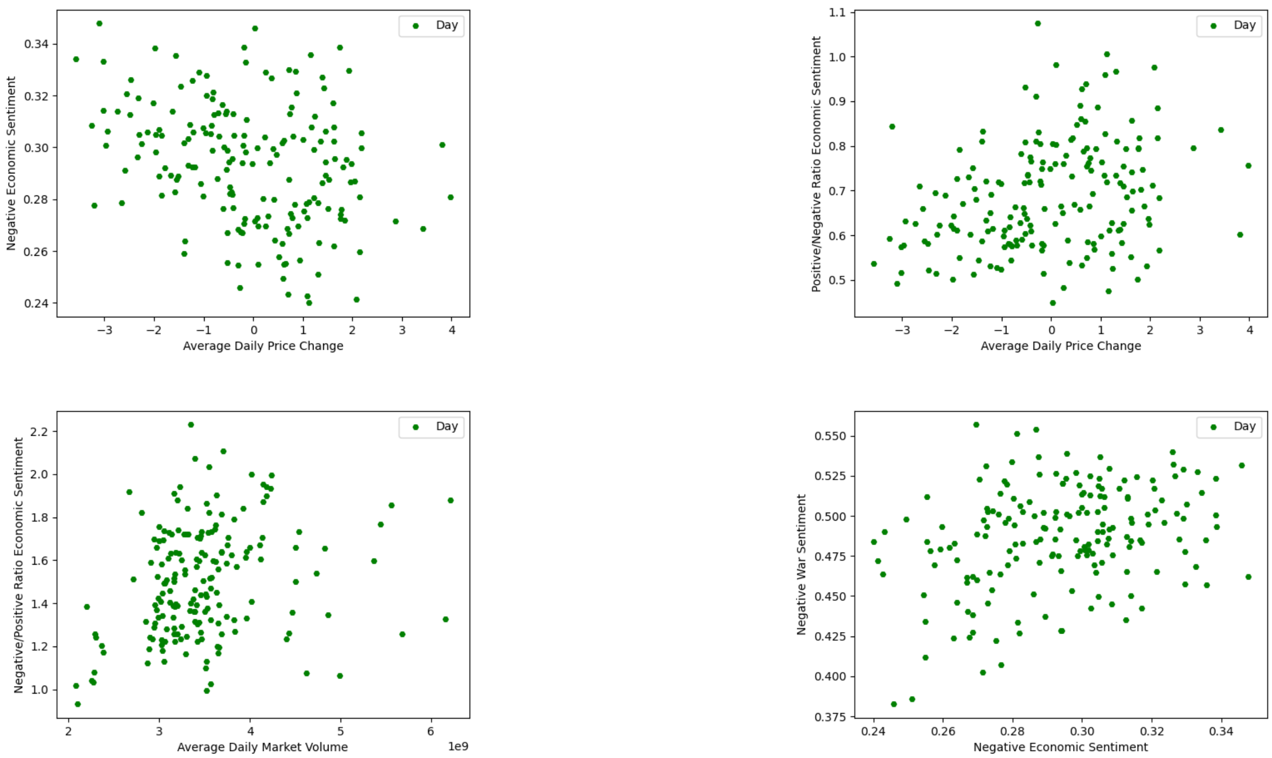

Figure 2

Because our data are time series, we also had to check for the within-dependency of each variable. The within-dependency of time series can be seen as the correlation of a variable between one of its datapoints with previous datapoints. To measure this, the autocorrelation function can be used, which measures the correlation between the time series and its own lagged values. It represents the similarness of one value with its previous values, the k^{th} lagged values. A time series k^{th} lagged value is the same time series, but then k periods (in our case: days) in the past. To give an idea of how this function operates, it is shown in figure 3 [Wil14].

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T} (r_t - \overline{r})(r_{t-k} - \overline{r})}{\sum_{t=1}^{T} (r_t - \overline{r})^2}$$

Figure 3: The autocorrelation function

The function has the following variables:

- $r(t)$ = The 'original' time series
- $r(t-k)$ = Time series shifted by k units
- $\bar{r}$ = The average of the (original) times series

We calculated the autocorrelation with lags k=1, k=2 and k=3. For example if k=3, we compute the correlation between $r(t)$ and $r(t-3)$. In other words, the time series is compared to itself but then three days behind in time. In many researches, correlations are translated in 'weak', 'moderate' or 'high'. Such allocations could make use of certain cutoff points, which one must be careful with. While coefficients less then 0.1 indicates no correlations and coefficients greater then 0.9 indicates a highly correlated relationship, everything in-between could be interpreted differently and according to the context [SBS18]. For the autocorrelation, the coefficient with lag k=1 of most variables where around 0.4 and decreased for higher k. However, the coefficients for the average sentiment of the past 3 days were mostly around the 0.8 for the k=1 lag and decreased a little for higher k. In summary, most of our data is moderately autocorrelated and the four variables which represents sentiments of the average of three days are highly autocorrelated. This was to be expected, because the averages of the past three days are partly computed with the same values and thus within-dependency is an inevitable consequence. Therefore, we decided to remove these variables from our dataset. This way the remaining variables of our dataset are at least not highly correlated. The autocorrelation coefficients are summarized in table 3.

Autocorrelations				
k^{th} lag	Average Daily Price Change	Average Daily Volume	Negative Economic Sentiment	Positive Economic Sentiment
k=1	0.42	0.79	0.42	0.44
k=2	-0.01	0.52	0.43	0.49
k=3	-0.06	0.36	0.33	0.48
Negative War Sentiment	Positive War Sentiment	N/P Ratio Economic Sentiment	P/N Ratio Economic Sentiment	N/P Ratio War Sentiment
0.53	0.32	0.44	0.47	0.36
0.48	0.22	0.49	0.5	0.2
0.44	0.25	0.44	0.44	0.27
P/N Ratio War Sentiment	Neg-Eco Sent Past 3 Days	Pos-Eco Sent Past 3 Days	Neg-War Sent Past 3 Days	Pos-War Sent Past 3 Days
0.35	0.88	0.91	0.91	0.84
0.28	0.75	0.82	0.8	0.64
0.28	0.58	0.7	0.69	0.42

Table 3

4 Results

The results of our research consist of a variety of correlation matrices and the time series graphs of the different types of sentiments.

4.1 Public-Market correlation

For our main research question we tested two market sentiment values against eight public sentiment values. Because of this multiple testing the likelihood of a type 1 error increases. Hence, we have to correct and do this by applying the Bonferroni correction [Arm14]. Therefore we multiple our obtained p-values of each individual test with 16 and set our α level at 0.05. If a correlation coefficient has a p-value < 0.05, we call this coefficient 'significant'. If not, we ignore the coefficient in our results and conclusion. Figure 4 shows the obtained significant correlations coefficients of the period from 2021-12-24 until 2022-06-23. We found 9 significant coefficients out of the potential 16 coefficients, with all the coefficients around the $|\rho| = 0.3$.

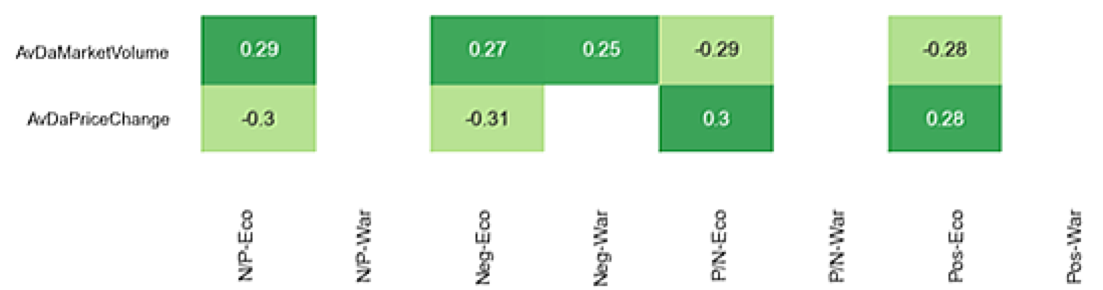

Figure 4: Public-Market correlation, from 2021-12-24 until 2022-06-23

The average daily market volume correlates positively only with the negative sentiments and negatively only with the positive sentiments. For the average daily price change this is exactly the opposite. For The Russo-Ukrainian war sentiment only the negative sentiment correlates (although it is the lowest of all) with the average daily market volume. Furthermore, the average daily price change correlates with all the economic sentiments and none of the public sentiments.
Now we analyze the two month time periods. For the period from 2021-12-24 until 2022-02-23, the pre-war period, we found 7 significant correlations with the volume and none with price change, shown in figure 5. Also here volume correlates positively only with the negative sentiments and negatively only with the positive sentiments. The correlations are much higher compared to the 6 months period, with two peak correlation of $|\rho| = 0.6$. For the two post-invasion periods we did not plot the figures, because we only found significant correlation between the average daily price change and negative economic sentiment for both periods. For the post-invasion period 1, from 2022-02-24 until 2022-04-23, we found a coefficient of -0.39 and for the post-invasion period 2, from 2022-04-24 until 2022-06-23, we found a coefficient of -0.38. It is noteworthy that the lower found coefficients can be explained because the data sets of these periods individually are smaller than the total data set. Therefore, the Spearman test will give higher p-values and consequential fewer significant coefficients.

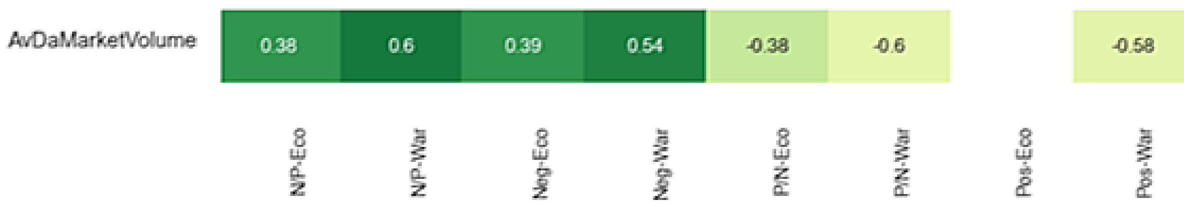

Figure 5: Public-Market pre-war correlation, from 2021-12-24 until 2022-02-23

4.2 Public-Public correlation

We also calculated the correlation between the public sentiments. We looked at the correlations between the economic sentiments and the Russo-Ukrainian sentiments. We tested 4 economic sentiments against 4 Russo-Ukrainian sentiments and again take in to consideration the Bonferroni correction. We did not look at the inner correlations in the public sentiments separately, since these sentiments are highly dependent on each other. For example, the negative economic sentiment will be highly correlated with the positive economic sentiment, but such correlation is meaningless.

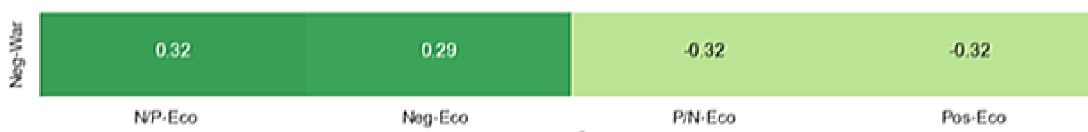

Figure 6

We found that only the negative Russo-Ukrainian war sentiment correlates with economic sentiments, and with all of them. The coefficient were around $|\rho| = 0.3$, as shown in figure 6. The negative economic sentiment correlates positively with negative Russo-Ukrainian war sentiment and conversely the positive economic sentiments correlates negatively with negative Russo-Ukrainian sentiment. When we look at the different periods, once again we see more correlations in the pre-war period and less in the post-invasion periods. In fact, we found that everything is correlated in the pre-war period and nothing is correlated in the post-invasion periods. The coefficient are around 0.4, as shown in figure 7. These results together with the results found earlier, states that correlation between sentiments is higher and more significant pre-war then in the overall 6 month period.

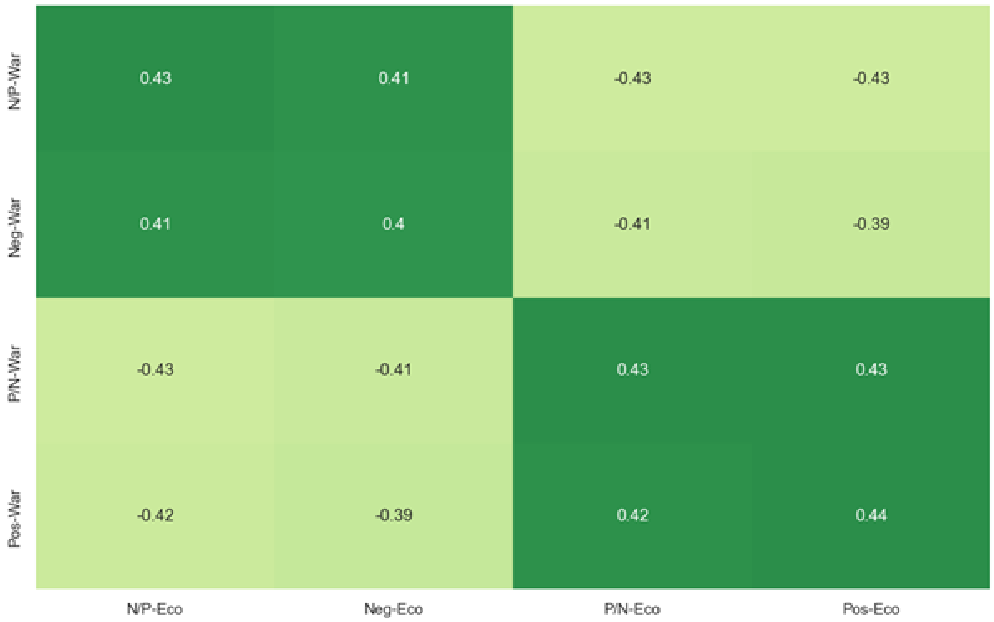

Figure 7

4.3 War effects

We visualized some of the variables from our composed datasets. In figure 8 we see that the percentage price changes highly fluctuate, which indicates that from this figure the market cannot be classified bull or bear in this period, as there is not a certain trend to be recognized. To be noticed is that percentage changes analyzed on each day separately is not a good classifier for a market trend, often there is being looked at the changes over a longer period. In figure 8 we also see that the market volume grew from $2.2 \cdot 10^9$ to $3.5 \cdot 10^9$ from December 2021 to June 2022, while the P/N ratio of economic sentiment fell from 0.93 to 0.66 this period. This indicates that a more negative sentiment effectuates a higher trading volume. In the figures you can also see that different peaks come around the same period. In the period of mid March for example, both sentiments peaked, with $6.1 \cdot 10^9$ for the volume and 1.01 for the P/N ratio of economic sentiment. The P/N ratio of the Russo-Ukrainian war sentiment is, besides some peaks in the pre-war periods, relatively stable. Certain events that happened during the war do not seem to play a significant role for the Russo-Ukrainian sentiment. People post, logically, far more negative tweets about the Russo-Ukrainian war then about the economy. The (relatively) stable character of the P/N ratio of the Russo-Ukrainian war sentiment compared to the fluctuations in the P/N ratio of economic sentiment suggests that people's perspectives on the economy do not depend on the Russo-Ukrainian war.

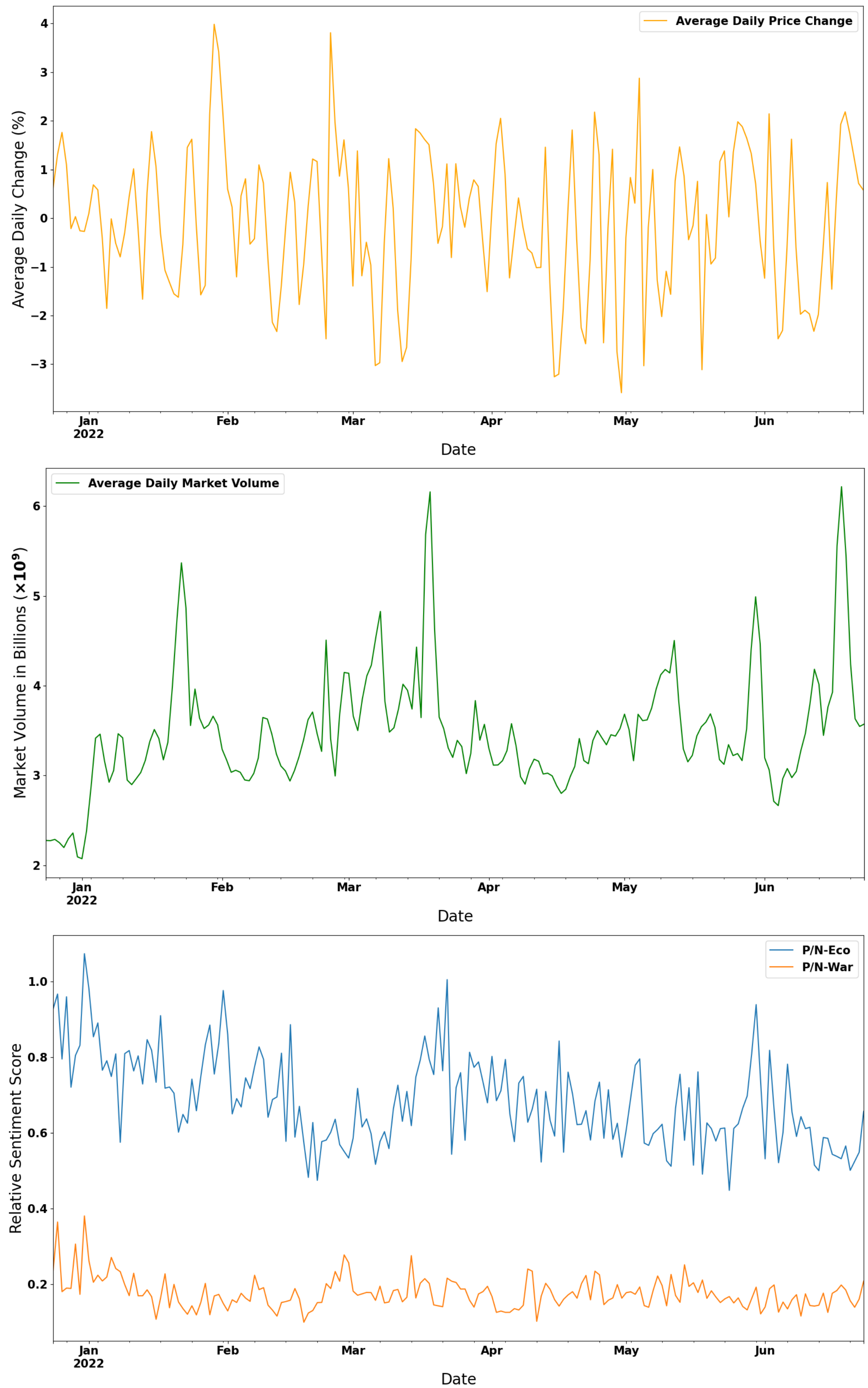

Figure 8: Various values from our dataset visualized

5 Discussion

5.1 Explanations of correlation

5.1.1 Public-Market correlation

For the public-market correlation we found that the average daily market volume correlates positively only with the negative sentiments and negatively only with the positive sentiments, for the average daily price change this is exactly the opposite. All these coefficients are explainable. When a sentiment is negative, people may tend to trade more than normally on the market. This is because traders might lose confidence in the market and sell their shares, anticipating on future market trends. Also, traders who see possibly profitable entries (for the long term) might buy more shares then normal, because they think shares are underpriced due the negative sentiment. All these actions lead to a high market volume, because traders tend to buy and sell more than they do on average. On the other hand, if there is a positive sentiment the market volume decreases. This might be because the volatility of the market is not high when there is a positive public sentiment. It is interesting that for the Russo-Ukrainian war sentiment only the negative sentiment correlates (although it is the lowest of all) with the average daily market volume.
Also we found that the average daily price change correlates with all the economic sentiments and none of the public sentiments. The negative coefficients are self-explanatory, the price drops when there is a negative economic sentiment. When the negative economic sentiment is dominant, this may lead to a pullback, since the market was in a bull trend in 2021. For the positive coefficients this is obviously the opposite, the price may rise when there is a positive economic sentiment.
The average daily market volume correlates positively with the negative sentiment of the Russo-Ukrainian war. Again, this is due the fact that people tend to trade more when there is negative sentiment. This single correlation of 0.25 found between the the average daily market volume and the Russo-Ukrainian war sentiment indicates a weaker relationship between the market sentiment and the Russo-Ukrainian war sentiment than between the market sentiment and the economic sentiment. This is not a very surprising result since tweets about the economy are (in general) more closely related to the market than tweets about a war.

5.1.2 Segmented periods

Now we analyze the two month time periods. The coefficients are much higher for the pre-war period in comparison to the total period. We see that all public sentiments correlate with the volume in the pre-war period and none of the public sentiments correlate with the volume in the post-invasion periods. This indicates that in a turbulent period, there is more activity on the market. On the other hand, we see none of the public sentiments correlates with the price change in the pre-war period and only the negative economic sentiment correlates with the price in both post-invasion periods. Hence, the price is largely independent on the public sentiment, if analyzed in short periods.

5.1.3 Public-Public correlation

For the correlation between public sentiments we found that only the negative Russo-Ukrainian war sentiment correlates with economic sentiments. That only the negative Russo-Ukrainian war sentiment correlates with economic sentiment may indicate that negative sentiment has a higher impact on the economic sentiment. However, both of the ratios of the Russo-Ukrainian war did not correlate, while they are dependent on the negative sentiment. Because we did not find any other significant coefficients and the found coefficients can be considered as moderate, you might say in general the economic sentiment and the Russo-Ukrainian war sentiment are not correlated. This indicates that the peoples perspective on the economy is not dependent on the Russo-Ukrainian war. However, when we look at the different periods we found that everything is correlated in the pre-war period and nothing is correlated in the post-invasion periods. This indicates that people's perspectives on the economy did depend on the Russo-Ukrainian war in the pre-war period, but did (definitely) not in the post-invasion periods. Therefore, it seems that people are more sensitive to a war that has yet to come than to a war that has already started.

5.2 Limitations

We used three stock market indices, the S&P500, NASDAQ and the DJIA, to measure the market sentiment. Despite the fact that these indices are, as the name suggests, good indicators for the sentiment of the represented companies within them, the overall market sentiment depends on much more variables. In sections 1.1 and 2.1 we discussed the complexity of the market. For example, banking decisions can also be used as indicators. Lately, the crypto market can also be a measure for overall market sentiment. Although this is a very new market and has relatively low market capitalization and is therefore less reliable. Also, the used market indices are (mainly) traded on the American stock market, taking into account other stock markets (indices) could shift the market sentiment and therefore our results.
The interpolate function used to approximate the missing values of the market sentiment gives a biased view of which direction the market would have gone in the weekends/holidays. These values are not available and therefore an approximation is always needed. Accurate calculations, which also take into account a variety of market performance indicators, may give a more realistic view of the tendency a market might have gone. But such calculations are beyond the scope of this thesis.
Furthermore, the tweets we crawled from Twitter are filtered for any other language than English. Of course, English is a worldwide spoken language and the United States plays the most important role in the overall economic playing field, however the expressions of any language could give a more objective view of the public sentiment. The fact that this research also focused on the Russo-Ukrainian war, underlines this even more. On the other hand, such conflicts are a worldwide discussed topic and, as discussed, have impact far across the border. Also, with a higher amount of seedterms compared to we used, as well as tweets, we might get a more unbiased public sentiment.
The within-dependency of our variables are also a point of improvement. First we did not anticipate for these correlations. During our research we came to the conclusion that the within-dependency of the variables could affect our future correlations and therefore our aim to answering our research questions. Besides that we removed the highly autocorrelated variables, one could argue that coefficients around 0.4 give a significant enough within-dependency to take another direction in the research.
Although we found significant correlations coefficients between the public and market sentiment, it is important to note that confounders may influenced the observed relationships. The (unmeasured) presence of confounders may give another perspective on the these relationships [Roh18]. The relationship may be influenced, or even caused, by an unknown third variable.

The obtained correlation coefficient between the market and public sentiment were mostly around 0.3. As discussed in section 3.3 such coefficients are context interpretable. As a result, there may be drawn different conclusions about our coefficients. The dependency of market sentiment and economic sentiment.

6 Conclusion & Further Research

In this research we investigated the relationship between the market and public sentiment. Our goal was to find the correlations between the market sentiment, the economic public sentiment and the Russo-Ukrainian public sentiment. We found significant correlations around $|\rho| = 0.3$ between the market sentiment and the economic public sentiment. Such coefficients are, in the context of our research, weak. Nevertheless they are indicating a relationship between people's expressions on Twitter and market trends. Therefore this sentiment is an indicator of financial activities and behaviour, although it may not be of any significant worth, i.e. useful in trading strategies. Also, we found one correlation of 0.25 between the market sentiment and the negative Russo-Ukrainian war sentiment. Due to the fact that there is only one relationship, we conclude there is a relation between the market and the Russo-Ukrainian war, but to a lesser extent than the market has with the economic public sentiment.

When we zoom into the different periods we see fewer significant coefficients but higher values for the market-public economic sentiment. Also, we see a correlation between all the Russo-Ukrainian war sentiments with the market volume in the pre-war period, but no correlation in the post-invasion period.

We investigated the relationship between expressions on Twitter about the economy with expressions about the Russo-Ukrainian war. We again found correlations around $|\rho| = 0.3$ between the negative sentiment about the Russo-Ukrainian war with all the economic sentiments. These correlations do indicate a relationship between the two public sentiments and show that people are more sensitive to negative sentiment. In the pre-war period all the public sentiments correlate with each other with higher coefficients, while none of them correlate in the post-invasion periods. This indicates a period-dependent relationship i.e. in the shaky period before the invasion there is a relationship between the economic sentiment and the Russo-Ukrainian war sentiment.

In future research a more specific correlation could be investigated. The high dependency of Russia's gas could be a point of investigation, through the correlation between gas prices and the Russo-Ukrainian war. Also, the dependency of wheat from Ukraine or Russia and the consequences thereof could be investigated. Furthermore, in further research more economic indicators should be taken into account. The market sentiment could be measured on a variety of indicators and therefore there is a lot to gain to get there. The influence of confounding variables in further research is also a interesting subject to analyze. In particular this would provide a more clear perspective of the relationship between the public and market sentiment. Finally, the economic effects on people's daily life of the Russo-Ukrainian war is a subject worth of research.

References

[ADA22] Dow Theory Explained: What It Is and How It Works, 6 2022.

[AF04] Werner Antweiler and Murray Z Frank. Is all that talk just noise? the information content of internet stock message boards. *The Journal of finance*, 59(3):1259–1294, 2004.

[Ame15] Jacob Amedie. The impact of social media on society. 2015.

[Arm14] Richard A Armstrong. When to use the b onferroni correction. *Ophthalmic and Physiological Optics*, 34(5):502–508, 2014.

[BGM90] Francesco Brioschi, Luca Ghezzi, and Rocco Mosconi. The link between stock prices and fundamental values: some evidence from the italian stock exchange. *Giornale degli economisti e annali di economia*, pages 363–392, 1990.

[BMZ11] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8, 2011.

[Cha22] Dave Chaffey. Global social media statistics research summary , 8 June 2022.

[CL16] Suk Bong Choi and Myung Suh Lim. Effects of social and technology overload on psychological well-being in young south korean adults: The mediatory role of social network service addiction. *Computers in Human Behavior*, 61:245–254, 2016.

[CPS88] David M Cutler, James M Poterba, and Lawrence H Summers. What moves stock prices?, 1988.

[Dic] Public sentiment definition and meaning | collins english dictionary.

[Fam65] Eugene F Fama. The behavior of stock-market prices. *The journal of Business*, 38(1):34–105, 1965.

[Gar13] Diego Garcia. Sentiment during recessions. *The journal of finance*, 68(3):1267–1300, 2013.

[Gro22] World Bank Group. Russian Invasion to Shrink Ukraine Economy by 45 Percent this Year, 4 2022.

[LMBS22] Gaye-Del Lo, Isaac Marcelin, Théophile Bassène, and Babacar Sène. The russo-ukrainian war and financial markets: the role of dependence on russian commodities. *Finance Research Letters*, 50:103194, 2022.

[May08] Antony Mayfield. What is social media. 2008.

[Mej09] Yelena Mejova. Sentiment analysis: An overview. *University of Iowa, Computer Science Department*, 2009.

[MG12] Anshul Mittal and Arpit Goel. Stock prediction using twitter sentiment analysis. *Standford University, CS229 (2011http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf)*, 15:2352, 2012.

[MRQ+15] Mahmoud Maqableh, Lama Rajab, Walaa Quteshat, Ra'ed Moh'd Taisir Masa'deh, Tahani Khatib, and Huda Karajeh. The impact of social media networks websites usage on students' academic performance. *Communications and Network*, 7(04):159–171, 2015.

[RLPM15] Andrijana Rebekić, Zdenko Lončarić, Sonja Petrović, and Sonja Marić. Pearson's or spearman's correlation coefficient-which one to use? *Poljoprivreda*, 21(2):47–54, 2015.

[Roh18] Julia M Rohrer. Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in methods and practices in psychological science*, 1(1):27–42, 2018.

[SBS18] Patrick Schober, Christa Boer, and Lothar A Schwarte. Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia*, 126(5):1763–1768, 2018.

[SDS66] Philip J Stone, Dexter C Dunphy, and Marshall S Smith. The general inquirer: A computer approach to content analysis. 1966.

[Ser22] Congressional Research Service. The Economic Impact of Russia Sanctions, 12 2022.

[ST06] Gerald Schneider and Vera E Troeger. War and the world economy: Stock market reactions to international conflicts. *Journal of conflict resolution*, 50(5):623–645, 2006.

[Tre21] Kat Tretina. What Are The Nasdaq Composite And Nasdaq 100 Indexes?, 7 2021.

[VC12] G Vinodhini and RM Chandrasekaran. Sentiment analysis and opinion mining: a survey. *International Journal*, 2(6):282–292, 2012.

[Wil14] Will. Autocorrelation – Learn by Marketing, 1 2014.