Economics and Business Review, Vol. 8 (22), No. 2, 2022: 7-28
DOI: 10.18559/ebr.2022.2.2

How Google Trends can improve market predictions—the case of the Warsaw Stock Exchange[1]

***Paweł Kropiński*[2], *Marcin Anholcer*[3]**

Abstract: The aim of this paper is to investigate interdependencies between the WIG20 index and economic policy uncertainty (EPU) related keywords quantified by a Google Trends search index. Tests for two periods from January 2015 till December 2019 and from June 2016 till May 2021 have been performed. This allowed the period of relative stability from the time of economic shock caused by the COVID-19 pandemics followed by various restrictions imposed by the governments to be distinguished.

A bivariate VAR model to selected search terms and the value of the WIG20 index was applied. After using AIC to establish the optimal number of lags the Granger causality test was performed. The increased empirical relationship has been confirmed between twelve EPU related terms and changes in the WIG20 index in the second period versus six terms for the pre-COVID period. It was also found that in the post-COVID period the intensity of reverse relations increased.

Keywords: Granger causality, Warsaw Stock Exchange, economic policy uncertainty, WIG20, Google Trends, predictions.

JEL codes: C12, C32, G17, G41.

Introduction

Optimizing investment strategies in financial markets is one of the most difficult and most important problems faced by both scientists and practitioners working in the area of finance. It is mainly because of the complexity and entanglement of the factors influencing the behaviour of financial instruments.

[1] Article received 23 November 2021, accepted 25 June 2022. The authors are indebted to three anonymous referees and to the editor for their valuable comments that allowed the improvement in the quality of this contribution.

[2] Department of Operations Research and Mathematical Economics, Poznań University of Economics and Business, al. Niepodległości 10, 61-875 Poznań, Poland, corresponding author: pawel.kropinski@phd.ue.poznan,pl, ORCID: https://orcid.org/0000-0003-2129-299X.

[3] Department of Operations Research and Mathematical Economics, Poznań University of Economics and Business, al. Niepodległości 10, 61-875 Poznań, Poland, m.anholcer@ue.poznan.pl, ORCID: https://orcid.org/0000-0001-7322-7095.

Numerous experts have been developing more and more sophisticated models including various factors to better measure and predict how they affect the prices of assets and other financial indicators.

However, the research potential of classical market prediction methods has been exhausted in the last few years. Methods such as technical analysis, fundamental analysis, or time series analysis are used in the practice of prediction. However, there is not much new significant research that could be performed in the area. New models and methods become more and more complex while they do not provide any significant improvement in prediction quality. These classical price predicting methods assume that historical events are likely to repeat in the future and that future prices depend on their historical values as well as on the values of certain economic indicators such as GDP, inflation, interest rates or performance indices of chosen branches of the economy. Since the number of such factors is finite this is the second reason why the classical approach cannot promise too much further improvement in prediction.

One can propose two possible directions of research that may lead to an improvement in the situation. The first is to apply non-classical methods using the fact that the computational power of computers has increased significantly in the last few decades. Also the capability of collecting, storing and processing large amounts of data continues to increase rapidly. For those reasons the use of various novel approaches including Artificial Intelligence (AI) and Machine Learning (ML) methods has become more and more popular. A few examples are mentioned below. Ticknor (2013) used Bayesian regularized artificial neural networks in a novel method to forecast financial market behaviour. Podsiadlo and Rybinski (2016) decided to use a rule-changing trading system based on rough set theory and have included a time-weighted rule voting method that accounts for information ageing to increase returns of the trading strategy. Using ML (deep learning, to be more specific) to find trading signals Ding, Zhang, Liu and Duan (2014) were able to create a model for which 60% of the fifteen randomly selected stocks from the S&P500 obtained an accuracy above 60% as to the buy/sell decision. Liu (2018), using a modified approach achieved the highest results (65.53% accuracy) on selected stocks from the S&P 500. Liu, Chao, Lin and Lin (2019) considered a deep learning model for predicting stock prices. Colliri and Zhao (2019) presented a network-based methodology using machine learning to optimize returns in the stock market. Anghinoni Zhao, Ji and Pan (2019) studied the application of complex network topology analysis in market time series forecasting and in particular they used community detection and simple network metrics to develop a trend-detection algorithm. Gera and London (2019) used the so-called asset graph, defined on the correlation matrix to reduce the complexity of the portfolio selection problem. Gałązka (2011) studied the network structure of the Polish Stock Market using Minimum Spanning Tree and Weighted Random Graph models. Patil, Wu, Potika and Orang (2020) analysed advantages in the market prediction of the model using a hybrid approach including graph-theoretical concepts and deep learning.

The second approach that could improve prediction is by the inclusion of new, non-standard explanatory factors. One of these could be the digital footprints left by the potential investors on the Internet. Gathering information on entertainment, shopping and many other social activities conducted through digital devices has become increasingly popular in the last two decades mostly due to the rapid development of technology. Recently their use intensified even further (Ali, 2020) due to the lockdowns and social distancing protocols caused by the outbreak of SARS-Cov-2 which has also shown with certainty that social media panic travels faster than the COVID-19 spread (Depoux et al., 2020). Knowledge of human behaviour on social media gathered through Big Data tools helped to make the Internet a vital marketing channel for businesses, organizations and institutions alike (Appel, Grewal, Hadi, & Stephen, 2020) while successful attempts have been made to use social media sentiment in searching for better stock returns (Tan & Tas, 2020). In one of the earliest studies in this area Gruhl, Guha, Kumar, Novak and Tomkins (2005) examined the belief that web discussions and blog postings may be an early indicator of real-world behaviour and proved it to be true for anticipating purchase spikes on Amazon. Goel, Hofman, Lahaie, Pennock and Watts (2010) used search query volume to forecast the opening weekend box-office revenue for feature films, first-month sales of video games and the rank of songs on the Billboard Hot 100 chart. In all cases they proved search counts to be highly predictive.

This study focuses on factors that improve prediction of the behaviour of financial markets. Price changes are caused by the buy/sell decisions of the investors that are influenced by the information that they possess. It follows that with the correct choice of hypotheses about how prices, opinions and information interact it should be possible to model market dynamics (Gusev et al., 2015). This coincides with the assumptions made by Simon (1955) suggesting that information gathering must always precede decision-making.

The Internet is nowadays the first choice source for many kinds of information including information about the financial markets. This observation was used e.g. by Da, Engelberg and Gao (2011), who analysed search frequency in Google (SVI). In a sample of Russell 3000 stocks from 2004 to 2008 they showed that SVI was correlated with existing proxies for investor attention as well as capturing the attention of retail investors. In particular the authors found that an increase in SVI predicted higher stock prices in the next two weeks.

It is worth observing that not only the search results about a specific company may be correlated with the future prices of its shares. Using the fact that a vast amount of data reflecting all aspects of humans' lives became available for analysis thanks to Google Preis, Moat and Stanley (2013) proved that based on search result statistics for the word "debt" one can create an investment portfolio model that is far ahead (326% vs 16%) of the results of the "buy and hold" strategy. A further collaboration allowed an improvement in the results by incorporating views and edits of Wikipedia articles into a predictive model (Moat et al., 2013).

Bloom (2009) proposed a structural framework for measuring uncertainty shocks. The model when used to simulate a large macro uncertainty shock produces a rapid drop and rebound in the employment, output and productivity growth. This research was later adopted and developed by Baker, Bloom and Davis (2016) in developing a new index for measuring economic policy uncertainty (EPU) based on newspaper coverage frequency. In the article three groups of terms are proposed relating to uncertainty, economy and policy-relevant terms. These words are broad enough to cover a large topic range and having no knowledge of the future shocks one can easily assume that some of these words will be used when trying to solve future crises. Baker and others (2016) also provide a translation to a few other languages and introduce similar indices for India, Italy or Germany. Following Baker and others' method EPU has been created for Croatia by Sorić and Loić (2017). Later Belgian EPU was brought to life by Algaba, Borms, Boudt and Van Pelt (2020) followed by Bergman and Worm (2020) for Danish uncertainty index.

Since the outbreak of COVID-19 researchers focused on how various aspects of the economic policy uncertainty have influenced the economy. Karanasos, Yfanti and Hunter (2021) focused on studying U.S. economic fundamentals and looking at the impact on emerging stock market volatility. The paper shows the power of the economic uncertainty channel by describing how much policy uncertainty levels increase the adverse effect on the volatility of the emerging markets. A more theoretical approach has been proposed by Aljanabi (2021) and demonstrates how uncertainty and information overload affect purchasing decisions. The study refers to information processing and motivation theory to investigate shifts in customer decisions under high uncertainty levels. Zebrowska-Suchodolska, Karpio and Kompa (2021) review the economies of Eastern Europe under extreme conditions such as COVID-19 pandemics. Using the GARCH model they find that Eastern European economies responded in an analogous way to high uncertainty which is consistent with the findings of this paper.

It is noteworthy that the EPU index for Poland is proposed by Hołda (2019) and the National Bank of Poland and the term sets proposed in this article are a basis for the research. Although the method has been applied successfully by Hołda for the Polish economy it has several disadvantages and no research exists that would address these drawbacks. To begin with the data gathering proposed requires complex web scraping and data mining process. In addition, the data is presented quarterly which has little hedging value against economic shocks as the event has already happened when the information is obtained. Furthermore to the authors' best knowledge there is no research that could investigate causality between uncertainty measures and changes in the Polish stock market index.

As far as the authors are aware this is the first study that will attempt to overcome the limitations listed above with the proposed Granger-causality method, simplified data extraction using the Google Trend index and weekly timestamps assuming predictive value of the information. The results of this research could be valuable for institutional investors as well as policy-makers and institutions with an exposure to the Polish stock market.

It was assumed after Jun, Yoo and Choi (2018) that the public will maximise their sense of well-being by reducing any uncertainties, financial or psychological, and try to avoid risk relating directly or indirectly to the economic shocks. This will prompt the public to engage in information-seeking behaviour which thanks to the increased use of smartphones and Internet accessibility will involve searching for information on the Internet. The public's ability to behave irrationally which may elicit and reveal higher uncertainty levels is not discounted. In the context of financial markets this irrational behaviour can adopt different forms (Huang, Rojas, & Convery, 2019) and provides diverse precursors to future changes in asset prices and the authors hope to observe it in movements of WIG20 index prices.

The purpose of this study was twofold. Firstly to examine how 34 EUP related terms might reveal predictive relationships to the movement of the WIG20 index. Secondly to investigate if economic shocks increase the attention on certain economic or policy-related words and how this may help predict the Warsaw Stock Exchange (WSE) fluctuations. The first hypothesis proposed is that the broad term set proposed by Baker and others (2016) and later by Hołda (2019) will offer significant help in explaining WIG20 index movements. While the second hypothesis is that in a time of market volatility caused by an economic shock of a global scale the same term set will contain even more information that helps to predict the WSE main index.

The paper is organized as follows. In Section 2 the data analysed in the paper are presented and in particular the method of data collection is explained and the list of words presented. Section 3 provides a brief description of the applied methodology. In Section 4 the results of the statistical research is presented. The paper concludes with some possible directions of further research and open problems.

1. Data

As already mentioned in the previous section the authors focused on the words related to EPU and the value movements of the WIG20 index. WIG20 as well as all Google Trends queries are marked with weekly timestamps and all show the change versus the previous observation which was achieved by the first-order difference. The first reason for this is that the raw numbers are rather difficult to interpret. Another is that it is planned to provide in future an efficient investment strategy using among other things, the findings of this article. Such a strategy would rely on information about whether some explanatory indicator goes up or down rather than on its absolute value.

It was decided not to use any terms corresponding with the COVID-19 epidemic as search terms (the reasons were explained in the Introduction). However, the authors still wanted to see whether including this factor as being a cause of social and economic shock would have some significant influence on the results of this research. For that reason the data were analysed independently for two periods. The first one was 01.01.2015–01.12.2019 (period A). The second one was 01.06.2016–01.06.2021 (period B). In both cases, an approximately 5-year period was analysed and the end of period A intentionally ends before the first human cases of COVID-19 infection were recorded in China.

In order to collect the values of the WIG20 index the free web service Stooq.pl was used on the 25th of August 2021. Closing prices of the WIG20 index at the end of a week were used and since the authors were interested in the relative changes of index prices first-order differences of the time series were calculated.

The data about the frequency of searching chosen terms were collected with Google Trends. These data have become an interest point for scientists since they were made publicly available. It allows access to the volume of queries users enter into Google in a given geographic area. Google Trends has shown some predictive ability and was proven to predict customer behaviour for products such as box-office revenue for feature films, first-month sales of video games or the ranking of songs on the Billboard Hot 100 chart (Goel et al., 2010). Choi and Varian (2012) used Google Trends to predict selected close-term economic indicators such as car and motor parts sales, unemployment benefit claims, journey destination planning or the Consumer Confidence Index for Australia. What is more interesting in the authors' opinion is that Google Trends can significantly improve the profitability of proposed investment strategies compared to the buy-and-hold strategy (Preis et al., 2013; Huang et al., 2019).

Google defines weeks as ending on Sunday relative to the number of searches carried out in that week. The data is anonymized and categorised by topic for each search query. The search results are normalised for a given time period and scaled on a range of 0 to 100 based on a topic's proportion to all searches on all topics (Choi & Varian, 2012). It is worth noting that only a sample of Google searches are used in Google Trends and for that reason query results may slightly vary from one day to another. Hence for each term the results over five separate, non-consecutive independent data queries between the 25th of July 2021 and the 5th of August 2021 were averaged. It was assumed that variability across different dates is irrelevant for the presented results and that the methodology used is consistent with earlier work by Preis and others (2013). For the current research “PL” as a geographic region and “web” as a Google product for which the trend has been performed were chosen. As the time frame of the queries is close to five years results in weekly time-frames were obtained. In order to quantify changes in the search results the relative change in search volume was used by calculating the first-order difference of each averaged time series.

Relationships between the Warsaw WIG20 stock index delta and changes of interest according to Google Trend results were tested for thirty four words closely related to those proposed by Hołda (2019) in line with the Economic Policy Uncertainty Index built by Baker and others (2016). Some words repeat with different suffixes to eliminate potential limitations as Google search is not as accurate when it comes to non-English queries (Bar-Ilan & Gutman, 2005). Also optionally words with and without polish letters such as, ‘ą’ or ‘ż’, ‘ś’ and ‘ć’

Table 1. Term set for the general economic policy uncertainty

Term set	English translation	Implementation in Polish
E	economic economy WIG20 credit	gospodarcza gospodarka WIG20 kredyt
P	sejm senat parliament government bill legislation regulation fiscal budget deficit tax VAT CIT (corporate income tax) PIT (personal income tax) NBP (National Bank of Poland), MPC (Monetary Policy Council), ECB Central Bank FED SNB Bundesbank	sejm senat parlament rząd, rzad ustawa legislacja regulacja fiskalne, fiskus budżet, budzet deficyt podatek, podatki, podatkowy, podatkowe VAT CIT PIT NBP RPP EBC bank centralny FED SNB Bundesbank
U	uncertainty	niepewność, niepewnosc
[PL]	Poland or Polish	Polska

Source: Based on (Hołda, 2019; Baker et al., 2016).

have been tested. The WIG20 index has been used as it is the main stock index in the Polish Stock Exchange that includes the largest twenty companies by market cap and liquidity. Since the EPU-broad index as proposed by Hołda (2019), see Table 1, contains a range of words relating to politics the WIG20 index will be most appropriate index for the purpose of this paper since the State Treasury influence over the largest companies amounts to nearly 75% of the index as at 20th of March 2015.

The use of the word "debt" in the work presented by Preis and others (2013) is a good example of the application of post-factum knowledge to predict events going back in time. As Preis analysed the period between 2004 to 2011 which overlaps with a debt crisis using the word "debt" is hardly appropriate when trying to improve the predictability of events that are unfolding. For that reason search terms such as "COVID-19" or "pandemics" are not included in the analysis. Instead of that, it was decided to focus on a set of words related to economic uncertainty. To be more specific the EPU index for Poland proposed by Hołda (2019) and the National Bank of Poland and the term sets proposed in this article were the starting point for the current research.

Additionally, two words are included in the research: "credit" and "wig20". Both are allocated to the "E" term set as "credit" replaces the word "debt" highlighted in the work of Preis and others (2013) and "wig20" is arbitrarily added assuming that some relationship will exist between the index and the name of that index being input into Google Trends. The terms used in the research are listed in Table 1.

The Polish language was used for the original search of the keyword terms as shown in Table 1. Within the body of the paper English equivalents will be used with a note indicating grammatical forms that do not exist in English. Query words without Polish diacritic marks are marked with (n.d.), the informal form of the word Public Treasury are marked (inf.) and an adjective in neutral form for the word "tax" are marked as (n).

2. Methodology

To establish if there was a causality relation between the selected terms and the value of the WIG20 index it was decided to use the well-known causality of Granger defined in his seminal paper (Granger, 1969). The main idea applied here is that the cause must come before the effect and for that reason when a variable x influences another variable y it should improve y's predictions. Since pairs of variables are analysed and the respective series are stationary (see below) the Wald test has been applied.

Assume that a VAR(p) system (i.e. the system with p lags) is considered. In the unrestricted form, it is:

$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i x_{t-i} + \varepsilon_t$$

$$x_t = \sum_{i=1}^{p} \alpha_i' y_{t-i} + \sum_{i=1}^{p} \beta_i' x_{t-i} + \varepsilon_t'$$

The restricted form of the equation explaining the value of y_t (where one assumes that x does not affect y) is

$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t$$

The null hypothesis states that $\beta_i = 0$ for all $i = 1, \ldots, p$. On the contrary the alternative hypothesis states that there is $i^* \in \{1, \ldots, p\}$ such that $\beta_{i^*} \neq 0$. The test statistic follows a χ^2 distribution. (For more details, see e.g. Lütkepohl, 2005, pp. 102–104).

A necessary assumption that needs to be satisfied when testing Granger's causality is the stationarity of both tested time series. Since it is good practice is to have a cross-check the Augmented Dickey-Fuller (ADF) test and Kwiatkowski-Philips-Schmidt-Shin (KPSS) test were applied for that purpose. In the ADF test the null hypothesis is the presence of a unit root or in other words the fact that the series is non-stationary. The alternative hypothesis is that the series is stationary. The test statistic follows the Dickey-Fuller t distribution (For more details see Dickey & Wayne, 1979; Fuller, 1996, pp. 546–578). In the case of the KPSS test, the procedure is somewhat opposite: the null hypothesis is that the series is trend-stationary while the alternative hypothesis states that there is a unit root. The test statistic of this Lagrange Multiplier (LM) test has been described by Kwiatkowski, Phillips, Schmidt and Shin (1992) where more details can be found.

Another issue is establishing the number of lags p. The following procedure was used. First it was assumed that a delay longer than approximately one quarter (thirteen weeks) does not need to be analysed since this period would be long enough to make a decision after searching the information (as will be seen in the following section this period is much shorter in reality). Then the VAR(p) model was estimated for every analysed pair of variables and the Akaike Information Criterion (AIC) was computed using the formula

$$AIC(p) = \ln\left(\det\left(\tilde{\Sigma}_u(p)\right)\right) + \frac{2pK^2}{T}$$

where $\tilde{\Sigma}_u(p)$ is the Maximum Likelihood estimate of the white noise covariance matrix for the estimated model VAR(p), K is the number of free variables (in this case $K = 2$) and T is the sample size (see Lütkepohl, 2005, pp. 146–147) for more details). Finally the value of p was chosen for which $AIC(p)$ took the minimum value.

All the computations were performed using the R statistical software. The data extraction was performed with the use of the "gtrendsR" package. The KPSS and ADF stationarity tests were performed using the "tseries" package. The VAR(p) models have been estimated with "vars" package but the values of $AIC(p)$ were computed directly from the formula. Finally the Granger causality has been examined with the "lmtest" package.

3. Results

The outcome of the ADF and KPSS tests shown in Table 2 and Table 3 indicates the stationarity of all the time series under investigation. In the first column the analysed series have been listed. Columns 2 and 3 describe the values corresponding with the ADF test (test statistics and p-value). The last two columns contain the values corresponding with the KPSS test (trend and p-value). As one can see in each of the seventy two cases the ADF null hypothesis that the unit root exists in the series can be rejected and one has no reason to reject the KPSS null hypothesis that the time series is stationary. This implies that each of the analysed series is stationary.

Table 2. Stationarity tests for time series for period A

Variable for period A	ADF		KPSS	
	statistics	p-value	KPSS Trend	p-value
Δ WIG20 (price)	–5.732	< 0.01	0.108	> 0.1
Δ Central Bank	–9.001	< 0.01	0.013	> 0.1
Δ budget (n.d.)	–8.583	< 0.01	0.011	> 0.1
Δ budget	–8.314	< 0.01	0.011	> 0.1
Δ Bundesbank	–11.504	< 0.01	0.022	> 0.1
Δ Corporate Income Tax	–8.927	< 0.01	0.013	> 0.1
Δ deficit	–7.382	< 0.01	0.012	> 0.1
Δ ECB	–9.725	< 0.01	0.012	> 0.1
Δ FED	–9.806	< 0.01	0.014	> 0.1
Δ fiscal	–6.930	< 0.01	0.087	> 0.1
Δ treasury (inf.)	–8.599	< 0.01	0.014	> 0.1
Δ economic	–7.566	< 0.01	0.018	> 0.1
Δ economy	–7.940	< 0.01	0.015	> 0.1
Δ credit	–8.711	< 0.01	0.014	> 0.1

Variable for period A	ADF		KPSS	
	statistics	*p*-value	KPSS Trend	*p*-value
Δ legislation	–9.756	< 0.01	0.013	> 0.1
Δ National Bank of Poland	–11.214	< 0.01	0.014	> 0.1
Δ uncertainty (n.d.)	–9.091	< 0.01	0.014	> 0.1
Δ uncertainty	–6.853	< 0.01	0.015	> 0.1
Δ parliament	–8.920	< 0.01	0.014	> 0.1
Δ Personal Income Tax	–7.400	< 0.01	0.022	> 0.1
Δ tax	–6.865	< 0.01	0.024	> 0.1
Δ taxes	–7.936	< 0.01	0.017	> 0.1
Δ taxing (n)	–7.337	< 0.01	0.023	> 0.1
Δ taxing	–8.564	< 0.01	0.017	> 0.1
Δ Poland	–8.962	< 0.01	0.012	> 0.1
Δ regulation	–8.951	< 0.01	0.015	> 0.1
Δ Monetary Policy Council	–9.848	< 0.01	0.017	> 0.1
Δ government (n.d.)	–9.069	< 0.01	0.013	> 0.1
Δ government	–8.807	< 0.01	0.013	> 0.1
Δ sejm	–8.601	< 0.01	0.014	> 0.1
Δ senat	–10.127	< 0.01	0.012	> 0.1
Δ SNB	–8.682	< 0.01	0.015	> 0.1
Δ bill	–9.458	< 0.01	0.014	> 0.1
Δ VAT	–9.341	< 0.01	0.014	> 0.1
Δ WIG20	–9.466	< 0.01	0.015	> 0.1

Source: Own calculations on the basis of Google Trends and Stooq.pl.

Table 3. Stationarity tests for time series for period B

Variable for period A	ADF		KPSS	
	statistics	*p*-value	KPSS Trend	*p*-value
Δ WIG20 (price)	–6.004	< 0.01	0,102	> 0.1
Δ Central Bank	–9.154	< 0.01	0,012	> 0.1
Δ budget (n.d.)	–8.353	< 0.01	0,011	> 0.1
Δ budget	–8.255	< 0.01	0,011	> 0.1
Δ Bundesbank	–9.852	< 0.01	0,013	> 0.1
Δ Corporate Income Tax	–9.461	< 0.01	0,013	> 0.1

Variable for period A	ADF		KPSS	
	statistics	*p*-value	KPSS Trend	*p*-value
Δ deficit	–8.013	< 0.01	0,021	> 0.1
Δ ECB	–9.540	< 0.01	0,025	> 0.1
Δ FED	–9.194	< 0.01	0,016	> 0.1
Δ fiscal	–7.305	< 0.01	0,017	> 0.1
Δ treasury (inf.)	–9.741	< 0.01	0,013	> 0.1
Δ economic	–8.027	< 0.01	0,017	> 0.1
Δ economy	–7.338	< 0.01	0,034	> 0.1
Δ credit	–8.405	< 0.01	0,013	> 0.1
Δ legislation	–10.298	< 0.01	0,013	> 0.1
Δ National Bank of Poland	–10.435	< 0.01	0,013	> 0.1
Δ uncertainty (n.d.)	–8.461	< 0.01	0,013	> 0.1
Δ uncertainty	–6.882	< 0.01	0,015	> 0.1
Δ parliament	–9.058	< 0.01	0,015	> 0.1
Δ Personal Income Tax	–7.590	< 0.01	0,020	> 0.1
Δ tax	–6.214	< 0.01	0,021	> 0.1
Δ taxes	–8.580	< 0.01	0,017	> 0.1
Δ taxing (n)	–6.522	< 0.01	0,023	> 0.1
Δ taxing	–8.510	< 0.01	0,021	> 0.1
Δ Poland	–8.756	< 0.01	0,013	> 0.1
Δ regulation	–8.597	< 0.01	0,019	> 0.1
Δ Monetary Policy Council	–9.342	< 0.01	0,013	> 0.1
Δ government (n.d.)	–7.989	< 0.01	0,018	> 0.1
Δ government	–7.342	< 0.01	0,025	> 0.1
Δ sejm	–8.506	< 0.01	0,015	> 0.1
Δ senat	–8.997	< 0.01	0,012	> 0.1
Δ SNB	–9.756	< 0.01	0,015	> 0.1
Δ bill	–9.325	< 0.01	0,016	> 0.1
Δ VAT	–9.377	< 0.01	0,014	> 0.1
Δ WIG20	–8.299	< 0.01	0,017	> 0.1

Source: Own calculations on the basis of Google Trends and Stooq.pl.

The results presented in Table 2 and Table 3 allowed the application of the Granger causality test and verification as to whether any change in the Google

Trends queries may Granger-cause the changes in WIG20 prices. As the first step the optimal number of lags p was found following the procedure described in the previous section. Then the causality itself was tested. Granger indicates towards causality looking at the ability to better predict the results. In the first test it was assumed that the delta of a Google Trend term Granger-causes changes in the WIG20 index when Google Trends values provide significant additional information about future changes of WIG20 and effectively allows for better forecasting. Also the opposite test was performed to identify if bidirectional causality exists between time series. The results of the computations are presented in Table 4 and Table 5.

Table 4. Causality tests for period A

Factor	optimal p	AIC(p)	p-value: Factor → Δ wig20 (price)	p-value: Δ wig20 (price) → Factor
Δ Central Bank	5	12.936	0.403	0.275
Δ budget (n.d.)	1	12.692	0.234	0.747
Δ budget	1	12.188	0.922	0.323
Δ Bundesbank	7	13.144	0.890	0.733
Δ Corporate Income Tax	2	12.211	0.967	0.075
Δ deficit	2	13.056	0.346	0.278
Δ ECB	6	12.146	0.690	0.302
Δ FED	10	12.360	0.467	0.542
Δ fiscal	12	11.567	0.986	0.910
Δ treasury (inf.)	6	13.051	0.135	0.150
Δ economic	3	12.004	0.023	0.769
Δ economy	2	12.362	0.694	0.317
Δ credit	4	11.106	0.346	0.258
Δ legislation	5	13.027	0.400	0.361
Δ National Bank of Poland	6	11.560	0.306	0.597
Δ uncertainty (n.d.)	5	13.088	0.035	0.515
Δ uncertainty	2	13.027	0.338	0.857
Δ parliament	4	12.038	0.726	0.609
Δ Personal Income Tax	1	12.531	0.058	0.653
Δ tax	2	11.627	0.023	0.335
Δ taxes	4	11.057	0.767	0.580
Δ taxing (n)	2	12.280	0.185	0.932
Δ taxing	3	11.769	0.898	0.721
Δ Poland	4	12.110	0.162	0.706
Δ regulation	4	11.672	0.946	0.016

Factor	optimal *p*	AIC(*p*)	*p*-value: Factor → Δ wig20 (price)	*p*-value: Δ wig20 (price) → Factor
Δ Monetary Policy Council	13	11.729	0.325	0.991
Δ government (n.d.)	6	11.937	0.003	0.728
Δ government	6	11.676	0.014	0.240
Δ sejm	3	12.372	0.212	0.352
Δ senat	3	11.701	0.679	0.381
Δ SNB	13	12.249	0.635	0.339
Δ bill	3	12.135	0.007	0.856
Δ VAT	12	12.184	0.098	0.885
Δ WIG20	6	12.720	0.268	0.748

Source: Own calculations on the basis of Google Trends and Stooq.pl.

Table 5. Causality tests for period B

Factor	optimal *p*	AIC(*p*)	*p*-value: Factor → Δ wig20 (price)	*p*-value: Δ wig20 (price) → Factor
Δ Central Bank	3	13.642	0.870	0.669
Δ budget (n.d.)	2	13.080	0.241	0.800
Δ budget	2	12.470	0.664	0.882
Δ Bundesbank	4	13.430	0.605	0.658
Δ Corporate Income Tax	2	12.540	0.818	0.263
Δ deficit	2	13.097	0.860	0.039
Δ ECB	8	13.483	0.804	0.771
Δ FED	6	13.305	0.022	0.787
Δ fiscal	3	12.611	0.006	0.173
Δ treasury (inf.)	6	13.460	0.142	0.086
Δ economic	2	12.442	0.830	0.074
Δ economy	3	12.697	0.190	0.074
Δ credit	4	11.832	0.049	0.042
Δ legislation	5	13.297	0.766	0.306
Δ National Bank of Poland	6	11.851	0.834	0.025
Δ uncertainty (n.d.)	2	13.690	0.097	0.817
Δ uncertainty	2	13.207	0.657	0.513
Δ parliament	3	12.274	0.818	0.097
Δ Personal Income Tax	11	12.900	0.009	0.146
Δ tax	2	11.928	< 0.001	0.038
Δ taxes	2	12.648	0.004	0.471

Factor	optimal p	AIC(p)	p-value: Factor → Δ wig20 (price)	p-value: Δ wig20 (price) → Factor
Δ taxing (n)	2	12.522	0.001	0.176
Δ taxing	3	12.671	0.217	0.514
Δ Poland	5	12.465	0.046	0.161
Δ regulation	4	11.960	0.321	0.695
Δ Monetary Policy Council	12	12.539	0.026	0.004
Δ government (n.d.)	3	12.415	0.009	< 0.001
Δ government	3	12.697	0.140	0.001
Δ sejm	4	12.761	0.339	0.001
Δ senat	3	12.022	0.942	0.655
Δ SNB	5	13.741	0.123	0.348
Δ bill	5	12.679	0.003	0.885
Δ VAT	3	12.578	0.366	0.378
Δ WIG20	3	11.303	< 0.001	0.075

Source: Own calculations on the basis of Google Trends and Stooq.pl.

According to the results presented in Table 4, in the case of period A the tests show that one-directional causality exists for the words:

- "economic" with a three-week delay,
- "uncertainty" written without Polish diacritic signs with a five-week delay,
- "tax" with a two-week delay,
- "government" with and without Polish diacritic signs both with a six-week delay,
- "bill" with a three-week delay.

These results suggest that the listed words could be used as early predictors for change in the WIG20 level. What appears very interesting is the words are in fact more often connected with politics than with the economy itself. It could suggest for instance that a growing discussion about the decisions of government and changes in the law is one of the more important factors increasing the predictability of changes of stock indices and it is better than discussion about central banks and other institutions involved in monetary or fiscal policy. This is what the authors managed to discover for the period not overlapping with any major global economic crisis.

The situation changes when one considers the period that overlaps with the time of the SARS-Cov-2 pandemic with all its consequences, such as lockdowns, a decrease in demand for services caused by social distancing, significant changes in purchasing behaviour and others. The tests performed for this period were intended to include the recent major economic shock caused by the COVID-19. It was hypothesised that with significant social disruption one would observe some regulatory and economic changes that would urge society and investors to gather more information and potentially reveal the link between WIG20 index changes. Interestingly more than in period 'A' queries appear to have bidirectional properties specifically between government and tax-related searches. According to the results presented in Table 5, this time the list of the words for which the non-causality hypothesis can be rejected is longer:

- "fed" with a six-week delay,
- "fiscal" with a three-week delay,
- "credit" with a three-week delay; notice that in this case also the reverse relationship should be considered as statistically significant,
- "personal income tax" with an eleven-week delay,
- "tax", "taxes", "taxing", all with a two-week delay, in the first case the relationship is mutual,
- "Poland" with a five-week delay,
- "Monetary Policy Council" with twelve-week delay; also the opposite influence is statistically significant,
- "government" with a three-week delay,
- "bill" with a five-week delay,
- "WIG20" with a three-week delay.

These results suggest that in the times when an economic shock occurs (and thus general uncertainty increases), one can observe many more possible predictors. This shows that in uncertain times before making investment decisions that eventually affect the financial markets the market actors gather much more information, this time also including fiscal policy and taxes or the possible actions of the institutions responsible for monetary policy such as the FED or MPC. Also an increasing interest in loans turns out to be a good predictor of the movements of stock indices. An interesting case is the word "bill" which is country-specific as the Polish system is not precedent-based and every major change in regulation must be predated by a bill. In general in period B a much larger spectrum of causal relationships was observed which allows the confirmation of the hypothesis that within the proposed methodology periods of economic shock appear to be more predictable than periods of relative stability.

As already mentioned several significant bi-directional causality relationships have been discovered in period B. In particular this can be observed for "Monetary Policy Council") which is not very surprising since the interest rates were reduced three times within three months between March 2020 and May 2020. As the government stepped in with a series of restrictions every decision had a huge impact on the economy and hence bidirectional causality for the word "government" is rather apparent.

What was not at the core of the authors' interest but is also interesting in their opinion is the change in reverse causality, that is how the changes in the WIG20 influence the popularity of chosen terms. Only the frequency of the word "regulation" can be considered as significantly influenced by the changes in WIG20 prices in the first period. The situation changes dramatically in the period overlapping the pandemic. Here (including the terms discussed above) the list is much longer: "deficit", "credit", "National Bank of Poland", "tax", "Monetary Policy Council", "government" and "sejm". This difference suggests that in uncertain times of economic shock and after observing changes in the stock indices investors are more likely to demand action from the government than in the "normal" times. However as was already mentioned this is not in the core of the considerations so this discussion will not be developed here.

Conclusions

In the proposed research it has been shown that in a period of economic shock represented by the COVID-19 pandemic changes in Google Trends EPU related term sets show significant Granger-causality with changes in the WIG20 index for twelve out of thirty-four words under investigation. This is twice as many as in the five years before the pandemic. Also four out of twelve causality relationships were bidirectional showing in particular strong interdependence between the stock index and government decisions regarding economic policy.

Besides that several causality relationships describing the influence of the stock index level on the search frequencies of selected terms have been proved by the research presented in this paper. Although the latter kind of causality was not the most interesting thing for the authors of this study it also brings some interesting conclusions. In particular one can deduce that in times of increased economic uncertainty changes in the financial markets increase the citizens' expectancies about actions performed by the government including the areas of fiscal and monetary policy.

Following the research of Hołda (2019), the authors are aware that word choice is to some degree arbitrary and only in a broad way can be related to other periods. Still it is noteworthy that fifteen out of thirty-four tested words show some degree of relationship significance and Granger-cause changes in WIG20 index even more if one also includes the reverse relationship. It becomes apparent that in many direct or indirect ways knowledge about social attention and information gathering by citizens could become advantageous in predicting future changes in financial markets. The scope of the research although not overly sophisticated proved to show a lot of promise in a future ability to predict significant shifts in a business cycle of the Polish stock market.

Some possible directions of further research and selected open problems are now discussed.

As already mentioned above the choice of search terms is always arbitrary. Thus one of the possible areas that could be improved is the word selection method. This could include increasing their number but also grouping that was not applied in this study. There are numerous possible sources of new words for instance some more detailed expressions corresponding to monetary policy ("interest rate", "increase", "decrease", "expansionary", "loose", "restrictive" etc.). However, at some point going too deeply into the details can cause losses in the significance of individual terms. For that reason a grouping of words could be necessary. Then a time series to compare with the level of some financial indicator would be derived as a number of queries containing any of the words from a group. Unfortunately some queries can contain several words from a group and thus the new series cannot be simply a sum of the series for individual terms. Some other method must be used in order to derive the search frequencies for the expressions which are unions of the individual terms. Grouping can be also useful owing to language-specific reasons. For instance the words "*podatkowy*" and "*podatkowa*" which are translated as "tax" or "taxing" will differ in Polish depending on the context and are two forms (masculine and feminine) of the same adjective. In this case the grouping is straightforward. Unfortunately there are also cases when the grouping is not that obvious, including the words that can have several meanings. A good example is "*rząd*", which declines in the genitive as "*rządu*", meaning "government" and the same word declined in the genitive as "*rzędu*", meaning among others "order" (such as "order of magnitude"). Of course grouping of both words should be different and would depend on the context.

Another issue concerning terms and words is their real meaning. It is very easy to imagine a situation where someone writes "it was a brilliant decision of the company's management", but the statement is in fact sarcastic. Of course it is not of high importance in the case of search queries (if a user tries to find some information one may assume that the user seeks the straightforward meaning of the statements) but if one would like to analyse also the other side of Internet activity (that is not only gathering but also publishing information that is very likely also related to the market's behaviour), the knowledge about the authors' intentions would be crucial to deduce the real implications of the occurrence of the chosen expressions for the movements of selected financial instruments. It is quite easy in some cases (for instance sarcastic messages on Reddit are marked with the "\s" tag) but in many others it would be necessary to perform sentiment analysis or at least some basic sarcasm detection. This would involve the application of some Machine Learning algorithms.

In this study only a causality relationship between selected terms and the level of the WIG20 stock index has been found. The authors did not focus much on the form of the relationship and did not investigate the influence of any other factor such as seasonal changes or other exogenous variables. A bivariate Vector Autoregression model for each search term and WIG20 level was assumed and the authors were not interested in the actual form of the model since the goal of this study was only looking for the presence of causality relationships. In future research one could consider more complex models such as vector variants or ARIMA, SARIMA or SARIMAX including in particular some other independent variables, neither connected with the search term frequency nor with the stock index level. These could improve the understanding of the factors influencing the price movements in financial markets. Of course instead of the main stock index one could consider also other financial indicators or even prices of individual assets or user-defined portfolios.

The ultimate purpose is the construction of an efficient trading strategy similar to the one proposed by Preis and others (2013). This could be a system continuously checking the frequencies of search queries of chosen words and based on them taking into consideration individual influence of each term on the selected financial value and the respective delays, giving recommendations such as "sell", "hold" or "buy". Of course such recommendation rules could rely also on some other variables such as chosen economic indicators. The more precise such a recommendation system is supposed to be, the more detailed should be the investigation of the form and strength of the relationships between the search frequencies of chosen keywords and values of selected financial indicators.

The results of this paper may not be applicable to other economies and further research is to draw wider picture of the causality between Google Trend terms and stock markets globally. One could expect that similar results would be obtained for countries having similar conditions, political culture and historical background meaning in particular the countries from Eastern Europe but not necessarily to the countries with a long history of democracy and liberal economy (such as in Western Europe) or completely different political and historical background (like the countries of Southern America, Asia, or Africa). These are however only suppositions and in order to confirm them a comparative study would be necessary.

Another direction of further research and probably the simplest one is repeating the investigations described in this study for other time periods. It would be interesting to analyse previous periods overlapping with some economic shocks (for example around the years 2001 or 2007) and checking whether they had a similar influence on the predictive power of search frequencies of selected terms to the influence of the COVID-19 outbreak. On the other hand one can hope that the current situation will stabilise soon due to vaccinations and an improvement in health care. Thus it would be also interesting to check whether some time after the situation improves the relationships described in this paper will go back to the situation of the pre-COVID period.

The last idea is to combine the presented method with some other classical or non-classical methods to use the relationships between Google Trends and the changes in stock markets to predict the latter. An example of such an approach joining various methodologies can be found in Hu, Tang, Zhang and Wang (2018) where the authors combined neural networks with an improved version of the sine cosine algorithm proposed by Mirjalili (2016). In the case of the research presented in this paper it could be a hybrid of the Granger's causality or other econometric methods and some ML algorithms—the former allows the identification of relationships, while the latter are often better in predicting the future, however they both widely rely on the quality of the input data and proper calibration of the parameters. These in turn can be supported with applying appropriate classical tools such as the ones used by Hu and others (2018) or econometric methods like those used in this paper. An obvious example of such an application of the methodology used in this paper to improve the performance of AI methods is the reduction in the size of the input dataset by elimination of all the observation concerning insignificant words (meaning, the words for which the causality is not statistically significant).

References

Algaba, A., Borms, S., Boudt, K., & Van Pelt, J. (2020, July 3). The Economic Policy Uncertainty Index for Flanders, Wallonia and Belgium. *SSRN Electronic Journal*, *6*, 1–16. https://doi.org/10.2139/ssrn.3580000

Ali, B. J. (2020). Impact of COVID-19 on consumer buying behavior toward online shopping in Iraq. *Economic Studies Journal*, *18*(42), 267–280. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3729323

Aljanabi, A. R. A. (2021). The impact of economic policy uncertainty, news framing and information overload on panic buying behavior in the time of COVID-19: A conceptual exploration. *International Journal of Emerging Markets*. https://doi.org/10.1108/IJOEM-10-2020-1181

Anghinoni, L., Zhao, L., Ji, D., & Pan, H. (2019). Time series trend detection and forecasting using complex network topology analysis. *Neural Networks*, *117*, 295–306. http://doi.org/10.1016/j.neunet.2019.05.018

Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. *Journal of the Academy of Marketing Science*, *48*(1), 79–95. https://doi.org/10.1007/s11747-019-00695-1

Baker, S., Bloom, N., & Davis, S. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Bar-Ilan, J., & Gutman, T. (2005). How do search engines respond to some non-English queries? *Journal of Information Science*, *31*(1), 13–28. https://doi.org/10.1177/0165551505049255

Bergman, M. U., & Worm, C. H. (2021). Economic Policy Uncertainty and consumer perceptions: The Danish case. Retrieved from https://www.researchgate.net/publication/343832963_Economic_Policy_Uncertainty_and_Consumer_Perceptions_The_Danish_Case

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, *77*(3), 623–685. https://doi.org/10.3982/ecta6248

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, *88*(1), 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Colliri, T., & Zhao, L. (2019). *A network-based model for optimizing returns in the stock market*. (2019 8th Brazilian Conference on Intelligent Systems (BRACIS), IEEE). http://doi.org/10.1109/BRACIS.2019.00118

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Depoux, A., Martin, S., Karafillakis, E., Preet, R., Wilder-Smith, A., & Larson, H. (2020). The pandemic of social media panic travels faster than the COVID-19 outbreak. *Journal of Travel Medicine*, *27*(3). http://doi.org/10.1093/jtm/taaa031

Dickey, D. A., & Fuller, W. A. (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366a), 427–431, http://doi.org/10.1080/01621459.1979.10482531

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2014). *Using structured events to predict stock price movement: An empirical investigation*. (2014 Conference on Empirical Methods in Natural Language Processing EMNLP), 1415–1425). http://doi.org/10.3115/v1/D14-1148

Fuller, W. A. (1996). *Introduction to statistical time series*. New York: John Wiley & Sons. http://doi.org/10.1002/9780470316917

Gałązka, M. (2011). Characteristics of the Polish Stock Market correlations. *International Review of Financial Analysis*, *20*(1), 1–5. http://doi.org/10.1016/j.irfa.2010.11.002

Gera, I., & London, A. (2019). *Portfolio selection based on a configuration model and hierarchical clustering for asset graphs*. (Proceedings of the MATCOS 2019, 39–42).

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). *Predicting consumer behavior with web search*. (Proceedings of the National Academy of Sciences of the United States of America, *107*(41), 17486–17490). https://doi.org/10.1073/pnas.1005962107

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3). http://doi.org/10.2307/1912791

Gruhl, D., Guha, R., Kumar, R., Novak, J., & Tomkins, A. (2005). *The predictive power of online chatter*. (Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 78–87). https://doi.org/10.1145/1081870.1081883

Gusev, M., Kroujiline, D., Govorkov, B., Sharov, S. V., Ushanov, D., & Zhilyaev, M. (2015). Predictable markets? A news-driven model of the stock market. *Algorithmic Finance*, *4*(1–2), 5–51. https://doi.org/10.3233/AF-150042

Hołda, M. (2019). *Newspaper-based economic uncertainty indices for Poland*. (NBP Working Papers No. 310, 1–49). Retrieved from https://ideas.repec.org/p/nbp/nbpmis/310.html

Hu, H., Tang, L., Zhang, S., & Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with Google Trends. *Neurocomputing*, *285*, 188–195. https://doi.org/10.1016/j.neucom.2018.01.038

Huang, M. Y., Rojas, R. R., & Convery, P. D. (2019). Forecasting stock market movements using Google Trend searches. *Empirical Economics*, *59*, 2821–2839. https://doi.org/10.1007/s00181-019-01725-1

Jun, S. P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, *130*, 69–87. https://doi.org/10.1016/j.techfore.2017.11.009

Karanasos, M., Yfanti, S., & Hunter, J. (2021). Emerging stock market volatility and economic fundamentals: The importance of US uncertainty spillovers, financial and health crises. *Annals of Operations Research*, *313*, 1077–1116. https://doi.org/10.1007/s10479-021-04042-y

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, *54*(1–3), 159–178. http://doi.org/10.1016/0304-4076(92)90104-Y

Liu, H. (2018). *Leveraging financial news for stock trend prediction with attention-based recurrent neural network*. Retrieved from https://arxiv.org/abs/1811.06173

Liu, J., Chao, F., Lin, Y., & Lin, C. (2019). *Stock prices prediction using deep learning models*. Retrieved from https://arxiv.org/abs/1909.12227

Lütkepohl, H. (2015). *New introduction to multiple time series analysis*. Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-540-27752-1

Mirjalili, S. (2016). SCA: A sine cosine algorithm for solving optimization problems. *Knowledge-Based Systems*, *96*, 120–133. https://doi.org/10.1016/j.knosys.2015.12.022

Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Eugene Stanley, H., & Preis, T. (2013). Quantifying Wikipedia usage patterns before stock market moves. *Scientific Reports*, *3*, 1801. https://doi.org/10.1038/srep01801

Patil, P., Wu, C. S. M., Potika, K., & Orang, M. (2020). *Stock market prediction using ensemble of graph theory, machine learning and deep learning models*. (Proceedings of the 3rd International Conference on Software Engineering and Information Management. ACM, 85–92). http://doi.org/10.1145/3378936.3378972

Podsiadlo, M., & Rybinski, H. (2016). Financial time series forecasting using rough sets with time-weighted rule voting. *Expert Systems with Applications*, *66*(30) 219–233. http://doi.org/10.1016/j.eswa.2016.08.066

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, *3*, 1684. https://doi.org/10.1038/srep01684

Simon, B. H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118. http://doi.org/10.2307/1884852

Sorić, P., & Lolić, I. (2017). Economic uncertainty and its impact on the Croatian economy. *Public Sector Economics*, *41*(4), 443–477. https://doi.org/10.3326/pse.41.4.3

Tan, S. D., & Tas, O. (2021). Social media sentiment in international stock returns and trading activity. *Journal of Behavioral Finance*, 22(2), 221–234. https://doi.org/10.1080/15427560.2020.1772261

Ticknor, J. L. (2013). A Bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*, *40*(14), 5501–5506. http://doi.org/10.1016/j.eswa.2013.04.013

Zebrowska-Suchodolska, D., Karpio, A., & Kompa, K. (2021). COVID-19 pandemic: Stock markets situation in European Ex-communist countries. *European Research Studies Journal*, *24*(3), 1106–1128. https://doi.org/10.35808/ersj/2408