Journal of Social Media Research
ISSN: 3062-0945
2025, 2(2), 122-138
https://jsomer.org

REVIEW ARTICLE OPEN ACCESS

Multilingual X/Twitter sentiment analysis of geopolitical risk using granger causality focusing on the Ukraine war and financial markets

John Corcoran Burns[1*], Tom Kelsey[1], and Carl Donovan[2]

1 School of Computer Science, University of St Andrews, St Andrews, Scotland, United Kingdom
2 School of Mathematics and Statistics, University of St Andrews, St Andrews, Scotland, UK
* **Correspondence**: Jack Cole Building, North Haugh, St. Andrews, KY16 9SX, United Kingdom, Tel: +1-718-702-7724, e-mail: jb370@st-andrews.ac.uk

Article Info

DOI: 10.29329/jsomer.23

Article History:
Received: 10/02/2025
Revised: 28/03/2025
Accepted: 29/04/2025

Keywords
X / Twitter
Ukraine War
Sentiment Analysis
Financial Market Analysis
Multilingual Analysis

Highlights:

- Possible to build a geopolitical risk index from X / Twitter historical data.
- Can use this index to identify large events such as the Ukraine war
- This index also provides predictive information on changes in financial markets.
- These changes occur on the daily level and the hourly level.

Abstract

This paper investigates the changes in financial assets and markets from December 1st, 2021, to April 30th, 2022, during the start of the Ukraine War. These dates roughly correspond to the prelude to the War in December 2021 to a few weeks after Russian troops withdrew from the Kyiv area on April 7th, 2022. We used the Goldstein 1992 Results Table to create Positive and Negative Geopolitical Risk bigrams (Goldstein, 1992). With these bigrams, we collected over 3.6 million tweets during our research period in seven different languages (English, Spanish, French, Portuguese, Arabic, Japanese, and Korean) to capture worldwide reaction to the Ukraine War. Using various sentiment analysis methods, we constructed a time series of changes in the daily Geopolitical Risk sentiment. We explored its relationship to 39 financial assets and markets at various time lags. We found through Granger causality that the geopolitical risk time series contained predictive information on several assets and market changes.

Citation: Burns, C. B., Kelsey, T., & Donovan, C. (2025). Multilingual X/Twitter sentiment analysis of geopolitical risk using granger causality focusing on the Ukraine war and financial markets. *Journal of Social Media Research, 2*(2), 122-138. https://doi.org/10.29329/jsomer.23

1. Introduction

On February 24th, 2022, Russia launched an invasion of Ukraine, formally starting the Ukraine War. This invasion was telegraphed months ahead of time, and contrary to the Russian expectation of a short conflict, the Ukraine War has continued up to the time of writing our study, more than three years later. This war represents one of the most important increases in geopolitical risk in our world today. Caldara and Iacoviello define "geopolitical risk as the risk associated with wars, terrorist acts, and tensions between states that affect the normal and peaceful course of international relations. Geopolitical risk captures both the risk that these events materialize, and the new risks associated with escalating existing events" (Caldara & Iacoviello, 2022, p. 2). This definition effectively describes the Ukraine War, and as Caldara and Iacoviello show, geopolitical risks impact various financial markets and assets (Caldara & Iacoviello, 2022). However, getting up-to-date information on geopolitical events on a large scale can potentially be time-consuming (Caldara & Iacoviello, 2022). Thus, we wanted to test whether there was a low-cost, quicker way to evaluate geopolitical risk. Thus, we turned to social media, specifically X / Twitter.

Founded in 2006, the microblogging platform X / Twitter has become one of the most popular social networking platforms globally, boasting 611 million active monthly users and ranking seventh in worldwide daily engagement (Zote, 2025). For research purposes, Pak and Paroubek put best why X / Twitter is an effective resource:

> Microblogging platforms are used by different people to express their opinion about different topics, thus it is a valuable source of people's opinions. Twitter contains an enormous number of text posts and it grows every day…Twitter's audience varies…Therefore, it is possible to collect text posts of users from different social and interests groups. Twitter's audience is represented by users from many countries….it is possible to collect data in different languages (Pak & Paroubek, 2010, p. 1).

Their last point is especially important for our study, as we aim to track worldwide sentiment; therefore, we need text written in multiple languages. Vicinitas states that English language tweets comprise only 30% of tweets posted. This means that a significant portion of all tweets will be excluded if we only collect English-language tweets. However, by including Japanese, Spanish, French, Portuguese, Arabic, and Korean, we can obtain approximately 85% to 90% of all tweets posted to X / Twitter. Therefore, using these languages will give us a larger corpus of tweets and better understand the overall sentiment surrounding the Ukraine War and associated geopolitical risks. To collect and analyze tweets related to the Ukraine War, we employed a combination of the X / Twitter API, sentiment analysis techniques, Granger causality, and finally the "Goldstein Index," which we define later in this paper. The rest of our paper is outlined as follows: Section 2 describes the key concepts employed for our analyses, Section 3 provides a literature review of previous work on geopolitical risk, media, and social media, and how they can affect financial markets. Section 4 details our methodology, while Section 5 displays our results. Section 6 discusses our findings, and Section 7 concludes.

2. Key Concepts

Three key concepts—the "Goldstein Index," sentiment Analysis, and Granger Causality—are the backbone of our research.

2.1. Goldstein Index

The "Goldstein Index" is a concept that comes from the 1992 paper A Conflict–Cooperation Scale for WEIS Events Data by Goldstein, who made use of the World Events Interaction Survey ("WEIS") data set. The WEIS data was developed by McClelland, which is "a record of the flow of action and response between countries (as well as non-governmental actors, e.g., NATO) reflected in public events reported daily in the New York Times from January 1966 through December 1978" (McClelland, 2006). The individual WEIS events can be grouped into "61 event types" (Goldstein, 1992, p. 2). Goldstein constructed a panel of eight International Relations faculty at USC to analyze and score the WEIS events (Goldstein, 1992, p. 6). This panel was individually given 61 cards with each WEIS event type and asked to "sort the cards into cooperative (friendly) actions and conflictual (hostile) ones" (Goldstein, 1992, p. 7) and rank them on a scale with -10 as the most conflictual and +10 as the most cooperative. The resulting rankings from each panel member were then averaged, creating what we refer to as the "Goldstein Index," a table of all 61 WEIS event types ordered from most conflictual to most cooperative. This table is the basis for our data gathering procedure, further described in our Methodology section. One potential bias of the "Goldstein Index" to note is mentioned by Goldstein as at the time of his writing: the table "seems to reflect the continuing emphasis placed on military affairs by international relations scholars" (Goldstein, 1992). However, this bias does not concern us greatly as our study revolves around the Ukraine War, a military affair.

2.2. Sentiment Analysis

Sentiment Analysis, as defined by Sim et al., is a field of document classification that classifies subjective impressions, sensibilities, attitudes of textual documents, individual opinions, on a topic, unlike text mining, which extracts information from text (Sim, 2021). Sentiment analysis programs, thus, try to define a given text as positive or negative, or potentially some other emotion such as anxiety (Gilbert & Karahalios, 2010). There are two main ways to accomplish sentiment analysis: Rules-based methods and Machine Learning based methods (Pota, 2021). Rules-based methods are typically lexicon dictionaries that assign specific values to certain words. The lexicon is then compared to the given text, and any matching words between the lexicon and the text are counted, and a sentiment score is given. As Cambria states, this is a popular sentiment analysis method "because of its accessibility and economy" (Cambria, 2013). As for the second method, machine learning based models, while more computationally heavy, have "the best results … obtained by deep learning approaches, using neural networks with various architectures, based on convolutional layers, …, recurrent layers, or the most recent transformers, constituting the layers of prominent systems employing BERT (Bidirectional Encoder Representation from Transformers)" (Pota, 2021, p. 1). For our study, we employed a combination of different methods including Rules-based methods, recurrent layers becoming recurrent neural networks ("RNNs") and transformer BERT models (Hutto & Gilbert, 2014; Devlin, et al, 2017; Géron, 2019; Inoue, et al, 2021). This combination proved necessary as different languages responded better to different sentiment analysis models.

2.3. Granger Causality

First detailed by Granger in his 1969 paper, Granger causality aims to find "the direction of causality between two related variables and also whether or not feedback is occurring" (Granger, 1969, p. 1). Since then, Granger causality tests have been used in various studies, including the Thurman and Fisher study, which aim to predict whether eggs Granger cause chickens or chickens Granger cause eggs (Thurman & Fisher, 1988). However, it should be noted that causality in this case does not mean the typical definition of causality, i.e., a change in one variable causes the change in another, but rather as Gilbert and Karahalios put it: "Although the technique has the word "causal" in it, we are not testing true causation. We can only say whether one time series has information about another." (Gilbert & Karahalios, 2010, p. 4). Moreover, as Granger himself states, his definition of causality mentions that "1. The cause occurs before the effect, and 2. The cause contains information about the effect that is unique and is not found in any other variable. A consequence of these statements is that the causal variable can help forecast the effect variable after other data has first been used." (Granger, 2003, p. 6). Thus, the null hypothesis for the Granger Causality Test is that the two time series are unrelated or provide any predictive information about each other. While the alternative hypothesis, which is accepted at a p-value less than 0.05, is that one tested time series does provide predictive information about the other time series. For this case study, we followed the lead of Bollen et al., who used Granger causality to test "whether one time series has predictive information about the other or not" (Bollen et al., 2011, p. 4). We chose to use Granger causality over traditional correlation to examine the relationship between the change in the sum of sentiment trend and the financial asset, as traditional correlation tests for a linear relationship between the variables, in other words, it checks to see if the variables change together at a constant rate. Granger causality works better for this paper as it tests if one series contains predictive information about the other, i.e., if one trend moves, does the other also move in the future. Since social media news reacts faster than the financial markets change their prices, the two time series will have a lag between them and not vary at the same time, thus Granger causality is a better statistical test for this paper.

3. Literature Review

Multitudes of studies use sentiment analysis, especially with X / Twitter. For example, Pak and Paroubek showed in their study how to effectively use X /Twitter to construct a corpus of tweets and use sentiment analysis on those tweets to derive insights (Pak & Paroubek, 2010). Additionally, Rajput et al used X / Twitter to analyze sentiment analysis around the Coronavirus pandemic (Rajput et al., 2020). Baker et al. (2021) used X / Twitter to "construct a database of more than 14 million tweets that contain a keyword related to 'uncertainty'…from June 1st, 2011, and March 1st, 2021". They transformed the count of these tweets into a time series and used that time series to measure economic uncertainty in the US during their research period. The Baker study was important for us as their use of keywords also provided a basis for us to use keywords to gather data for our analysis with the "Goldstein Index".

Many papers have also explored the relationship between news media and the effect on various financial markets through sentiment analysis. Using sentiment analysis to key in on anxiety-related terms in a large online blog, LiveJournal, Gilbert and Karahalios found through using Granger causality analysis that "increases in expressions of anxiety…predict downward pressure on the S&P 500 index" (Gilbert & Karahalios, 2010, p. 1). Uhl also showed that using a corpus of Reuters news articles, the sentiment analysis of those articles over time could "predict changes in stock returns better than macroeconomic factors" (Uhl, 2014, p. 1). Tetlock, et al., 2008, takes a more expansive approach to the returns of specific firms in the S&P 500 index by using sentiment analysis on articles from the Wall Street Journal and the Dow Jones News Service from 1980 to 2004 (Tetlock, et al., 2008, p. 2) to show that the number of negative words used in the articles about the firms can forecast lower earnings for the firms.

Three papers influenced our research: Bollen et al. (2011), Amen (2020), and Caldara and Iacoviello (2022). While Bollen et al. provided a framework about how to work with Granger causality and X/Twitter, Caldara and Iacoviello, and Amen provided the theoretical basis for working with trends in geopolitical risk data and the assets we should investigate that might be affected by geopolitical events, such as the Ukraine War. Bollen et al. researched whether the change in moods and the change in the Dow Jones index were linked. To do so, they complied a X / Twitter corpus of tweets containing "author's mood states" (Bollen et al., 2011, p. 2) and analyzed them through sentiment analysis programs to identify six moods: "Calm, Alert, Sure, Vital, Kind, and Happy" (Bollen et al., 2011). Creating a time series from the tweets' sentiment, Bollen et al. then used Granger causality to find if the change in mood sentiment that predates a change in the Dow Jones index. They found that out of the six moods, only Calm passed the Granger causality test and had information that predicted the change in the Dow Jones from 2 – 6 lags (Bollen et al., 2011).

On the other hand, Caldara and Iacoviello, and Amen focused specifically on geopolitical risk. Caldara and Iacoviello built the Geopolitical Risk (GPR) index, which used the count of news articles that mentioned their keyword indicators for geopolitical risks across 11 different English language newspapers starting from 1985 (Caldara & Iacoviello, 2022, p. 7). This GPR index captured the changes in geopolitical risks, and Caldara and Iacoviello were able to show how the increases in the GPR index predicted lower stock returns (Caldara & Iacoviello, 2022). Lastly, Amen built the Thorfinn Sensitivity Index (TSI), which uses "over 30,000 daily feeds" (Amen, 2020) to construct a daily index of the weight average of 12 geopolitical risk groups which experts have scored based on the news feeds that have come in for that day (Amen, 2020, p. 2). Amen then compares the changes in the TSI to changes in various "safe havens" and "risky assets" (Amen, 2020, p. 6) to develop trading strategies. Caldara and Iacoviello, and Amen had a wide range of assets and markets that provided a starting point for our analyses. Appendix A contains Table A.1, which shows the different financial assets we considered and their sources, while Table A.2 regroups them into the asset class categories we used.

Finally, we explored more recent research to compare our methods to. Niu et al. (2023) followed a similar data-gathering procedure as Caldara and Iacoviello, collecting news stories from ten English-language papers looking for key geopolitical words to build a time series showing the change in geopolitical risk (Niu et al., 2023, p. 4). Building on Caldara and Iacoviello, they used various machine learning methods to predict changes to the S&P 500 based on the geopolitical risk time series data. They found that Support Vector Regressions provided the highest predictive ability of their methods tested (Niu et al., 2023). Yilmazkuday (2024), constructed a study showing how geopolitical risk affected the stock prices worldwide to different degrees. For example, a one-unit increase in geopolitical risk caused a 0.8 decrease in stock prices in Latvia (Yilmazkuday, 2024). They also studied the Ukraine War and found that most affected countries' stocks were near the source of the geopolitical risk, i.e., Ukraine (Yilmazkuday, 2024). They also used the geopolitical risk keywords from Caldara and Iacoviello to build the geopolitical risk timeline and compare the stock values of the markets in different countries worldwide (Yilmazkuday, 2024). These two studies helped solidify our data collection methodology, as we also used a variation of the Caldara and Iacoviello method.

We aim to extend the literature by combining the "Goldstein Index" with X / Twitter to see if we can capture large geopolitical events, such as the Ukraine War and see how the sentiment around a geopolitical event can affect different financial assets on an equivalate or smaller time scale than both Caldara and Iacoviello, whose index captures both daily and monthly data, and Amen, whose index is only for daily. Additionally, we aim to capture the global impact of a geopolitical event by using multiple languages. While many studies only look at English tweets [Pak and Paroubek, 2010; Baker et al., 2021], or perhaps one additional language like Italian for Pota et al. (2011), or Dutch for Kleinnijenhuis. (2013), we aim to capture a more expansive, worldwide sentiment using the seven languages we study.

4. Method

Our methodology for this paper consists of three parts. The first was data gathering, the second was sentiment analysis, and the last was financial market analysis with Granger causality.

4.1. Data Gathering

As described earlier, we used the "Goldstein Index" as the basis for our data gathering. Many studies that worked with X / X/Twitter (Abouzahra & Tan, 2021; Baker et al., 2021; Beykikhoshk et al., 2015) have made use of keywords to collect tweets through the X / X/Twitter API, and so we decided to follow these methodologies. However, the "Goldstein Index" does not fit neatly into the X / Twitter API framework, as shown below in Table 1.

Table 1. A recreation of the portion from the Goldstein Paper showing the table Goldstein created. As can be seen, many of the phrases Goldstein uses would not work with the Twitter / X API as they are too long or awkward.

New Weights for WEIS Events		
Event Type	**Weight**	**SD**
223-Military attack; clash; assault	-10.0	0.0
211-Seize position or possessions	-9.2	0.7
222-Nonmilitary destruction / injury	-8.7	0.5
221-Noninjury destructive action	-8.3	0.6
182-Armed force mobilization, exercise, display; military buildup	-7.6	1.2
195-Break diplomatic relations	-7.0	1.3
173-Threat with force specified	-7.0	1.1
174-Ultimatum; threat with negative sanction and time limit	-6.9	1.4

To address this issue, we split the phrases in the index into single terms (such as "attack," "clash," "assault") and bigrams (two-term phrases like "military attack," "military clash," "military assault"). These smaller phrases are more manageable for the X/Twitter API, which allowed us to collect more data. Through experimentation, we found that while the single terms gathered more data, these tweets addressed various topics rather than the geopolitical tweets we searched for. The bigrams, on the other hand, provided a better corpus of tweets for geopolitics, even though there were fewer of them. Thus, we chose to use bigrams for our study, as their increased precision over the data collected was more valuable for our research.

As described earlier, we investigated not only English tweets but also incorporated additional languages to understand worldwide sentiment regarding the Ukraine War better. To do this, we had our "Goldstein Index" bigrams for French, Portuguese, Arabic, Japanese, and Korean translated by Gengo, a professional translation company. For Spanish, we used an independent translator. We deemed it important to use a human translator over machine translation because, as described by Pearse: "while [machine] fluency improves, mistranslation still occurs, so it is still vital to have a human translator check and edit the machine translation" (Pearse, 2020). Appendix B shows a world map highlighting the coverage we gained using multiple languages.

Next, we gathered the tweets from December 1st, 2021, to April 30th, 2022, the timeframe around the start of the Ukraine War. To do so, we implemented "Twarc" (2022), which collects and stores tweets from the X/Twitter API[1] from specific periods that use keywords such as our "Goldstein Index" bigrams. For our research, we used the top ten negatively and positively weighted bigrams that returned a non-zero number of tweets, where the majority of tweets focused on geopolitics. For example, while "call truce" ranked below "policy support" (2.9 and 4.5, respectively), many of the tweets we obtained for "policy support" focused more on internal politics than geopolitics than the "call truce" tweets, thus "call truce" was used. We also removed tweet duplicates removed by the "text" variable and the "created_at" variable generated from the X/Twitter API from our tweet data, as we viewed anything retweeted within the exact second after the original posting as most likely a bot. However, this removal did not cause significant data loss. This method collected over 3.6 million tweets for our research period. After collecting the tweets, we moved on to the sentiment analysis of the tweets' text.

[1] https://developer.twitter.com/en/docs/twitter-api/tweets/filtered-stream/introduction

4.2 Sentiment Analysis

Three different methods were applied for the sentiment analysis process. For English, we implemented the VADER Lexicon developed by Hutto and Gilbert. VADER is a rules-based sentiment analysis lexicon that is highly accurate, especially on short English texts (96%), such as tweets (Hutto & Gilbert, 2014, p. 9). For Arabic, we turned to CamelBERT, developed by Inoue et al. Based on the BERT model, a multi-layer Transformer-based model used for various natural language processing tasks, CamelBERT achieves the same functionality as BERT, including accurate sentiment analysis results for Arabic. However, for Spanish, French, Portuguese, Japanese, and Korean, we created our own RNN models to obtain the sentiment for tweets (Géron, 2019). While there were Transformer models trained for these languages (such as BETO by Cañete et al for Spanish, BERTimbau by Souza et al for Portuguese, CamemBERT by Martin et al for French, KR-BERT by Lee et al for Korean, and Bert-Base-Japanese by Tohoku-nlp), we found through testing that our X/Twitter data received either poor sentiment accuracy or had an extended processing time when evaluated by these BERT models. While these models are generally accurate overall, there were a few reasons why they did not work for our data. First, there was a mismatch between our X / Twitter data and the pre-training data that the various BERT models use. X/Twitter data is short text only, while the pre-trained data did not use this short text data exclusively, which could cause lower sentiment accuracy. Second, our computational limitations prevented the BERT models from executing in a reasonable time, as they are generally large models, and we had a vast amount of data. Thus, we turned to creating RNNs. RNNs are effective in this task because they can remember short sequences, such as tweets, and return a sentiment analysis score. We found sentiment analysis short text datasets for each language (TASS 2020 for Spanish; Gamebusterz for French; Augustop for Portuguese; Darkmap for Japanese; Park for Korean) and thus could train effective RNNs off these datasets. These RNNs overcame the pretraining data problem of the BERT models and are significantly smaller, which improved runtime while maintaining a high sentiment accuracy.

4.3. Financial Markets and Granger Causality

With the sentiment obtained for each tweet in the different languages, we compiled the tweets into one data frame and grouped the tweets by day, attaining the sum of the sentiment for each day, thus creating a time series for the change in the sentiment by day. We compared this time series to the various financial assets mentioned in Appendix A. We obtained the financial asset data for December 1st, 2021, to April 30th, 2022, through the Python package yfinance (Aroussi, 2023) and various other websites that contained the relevant financial asset data (CNBC, 2023; GoldHub, 2023; Investing.com, 2023a; Investing.com, 2023b; LiveCharts.co.uk, 2023; Monitor, 2023). However, these financial assets all trade at different times, and certain regions observe different holidays that close the markets. Thus, we needed to alter our sum of sentiment time series so that each date would line up correctly. Additionally, the financial data time series is needed to check for stationarity. Stationarity for a time series is defined as "a flat-looking series, without trend, constant variance over time, a constant autocorrelation structure over time, and no periodic fluctuations (seasonality)" (National Institute of Standards, 2023). Stationary checks are important as Granger causality only works under the assumption of stationary time series (Granger, 1969). Thus, differencing (where the current observation is subtracted from the previous observation) makes a time series stationary. If differencing is not done, the Granger causality results would be useless; they are applied to the financial and sentiment time series. Afterwards, Granger causality was tested to see if the sentiment analysis of the tweets obtained with the "Goldstein Index" bigrams can provide information about forecasting the change in the various financial assets. It should be noted that we did not create a Vector Autoregression (VAR) model as Caldara and Iacoviello did, as they focused more on trying to forecast the change of the model variables. In contrast, we were more interested in investigating the relationship between the sentiment on social media and various financial assets. So, Granger causality applied better in our case.

For the Granger causality model specifications, we were testing to see if the financial asset is exogenous to the sum of sentiment, i.e. if the sum of "Goldstein Index" sentiment trend Granger causes the change in the price of the financial asset, which makes the financial asset price the endogenous variable and the change in the sum of sentiment the exogenous variable. Once we made our time series stationary through differencing, we could apply Granger causality to see if the sentiment analysis of the tweets obtained with the "Goldstein Index" bigrams can provide information about forecasting the change in the various financial assets.

For the hourly time frame, the whole methodology is repeated. The only change is that for the time series, we sum the sentiment at the hourly level instead of the daily level and obtain the financial data at the hourly level as well.

5. Results

Table 2 below shows each bigram's tweet count with a negative weight ("Goldstein Negative") for each language.

Table 2. The number of tweets that contained each Bigram from December 1st, 2021, to April 30th, 2022, for the "Goldstein Negative" Category.

Bigrams	English	Spanish	French	Portuguese	Arabic	Japanese	Korean	Total
Military Invasion	931,256	180,118	23,357	20,398	3,234	4,514	662	1,163,539
Military Attack	722,575	116,651	19,825	14,585	7,029	40,735	545	921,945
Military Clash	44,495	18,172	742	13,332	1,567	5,390	2,475	86,173
Military Assault	180,108	8,650	2,986	426	396	395	0	192,961
Seize Position	3,774	5	691	14,934	2,614	4	68	22,090
Seize Possession	1,460	27	4	782	478	0	1	2,752
Non-Military Destruction	7	5,813	38	1,641	66	0	0	7,565
Non-Military Injury	1	23	0	96	29	0	0	149
Force Mobilization	3,795	8,953	3,421	1,274	2,210	63	714	20,430
Force Exercise	50,138	50,117	1,650	1,263	3,959	4,339	366	111,832
Total	1,937,609	388,529	52,714	68,731	21,582	55,440	4,831	2,529,436

For our research period, we collected 2,529,436 tweets that used one of these bigrams across all 7 languages, of which 591,547 were non-English. We included "Military Invasion" as a bigram for this specific study that was not included in the original Goldstein Index. While the original Goldstein Index would work in all other contexts for different events, since the Ukraine War was an invasion, leaving out the bigram synonym of "Military Invasion" would have missed over a million tweets from this study, which would have been severely detrimental. Table 3 shows the tweets count for each bigram with a positive weight ("Goldstein Positive") for each language.

Table 3. The number of tweets that contained each Bigram from December 1st, 2021, to April 30th, 2022, for the "Goldstein Positive" Category.

Bigrams	English	Spanish	French	Portuguese	Arabic	Japanese	Korean	Total
Military Assistance	443,872	37,954	33,368	3,369	3,327	4,914	8,248	535,052
Economic Aid	111,168	137,250	9,997	6,552	390	55,645	2,388	323,390
Substantive Agreement	1,219	1,553	17,996	65	309	69	234	21,445
Suspend Sanctions	34,326	7,222	1,518	689	1	0	158	43,914
Diplomatic Recognition	16,062	4,500	493	199	183	11	202	21,650
Grant Privilege	9,608	1,572	204	519	1,055	3,421	4,420	20,799
Call Truce	55,011	1,477	233	1,347	716	7	1,899	60,717
Material Assistance	9,127	2,357	323	377	6,944	0	198	19,326
Endorse Position	7,278	6,907	1,457	81	503	3,243	15,881	35,350
Verbal Support	10,330	2,446	10,337	304	57	3	553	24,030
Total	698,001	203,238	75,926	13,529	13,485	67,313	34,181	1,105,673

For the "Goldstein Positive" category, we collected 1,105,673 tweets, of which 407,672 were non-English. This brings the total number of tweets during our study period to 3,635,109, with nearly 70% of tweets coming in the "Goldstein Negative" Category and 27% being non-English.

Figure 1 is the daily count of tweets captured by the "Goldstein Index" bigrams from December 1st, 2021, to April 30th, 2022. This represents a total of 151 days. Each day is the total count of tweets captured by each topic with the Goldstein Negative and Goldstein Positive bigrams. The count of Goldstein's bigram tweets starts rising around two weeks before the start of the Invasion. He remains, on average, three times higher than before the War started, indicating increased discussion of geopolitical events on X / Twitter.

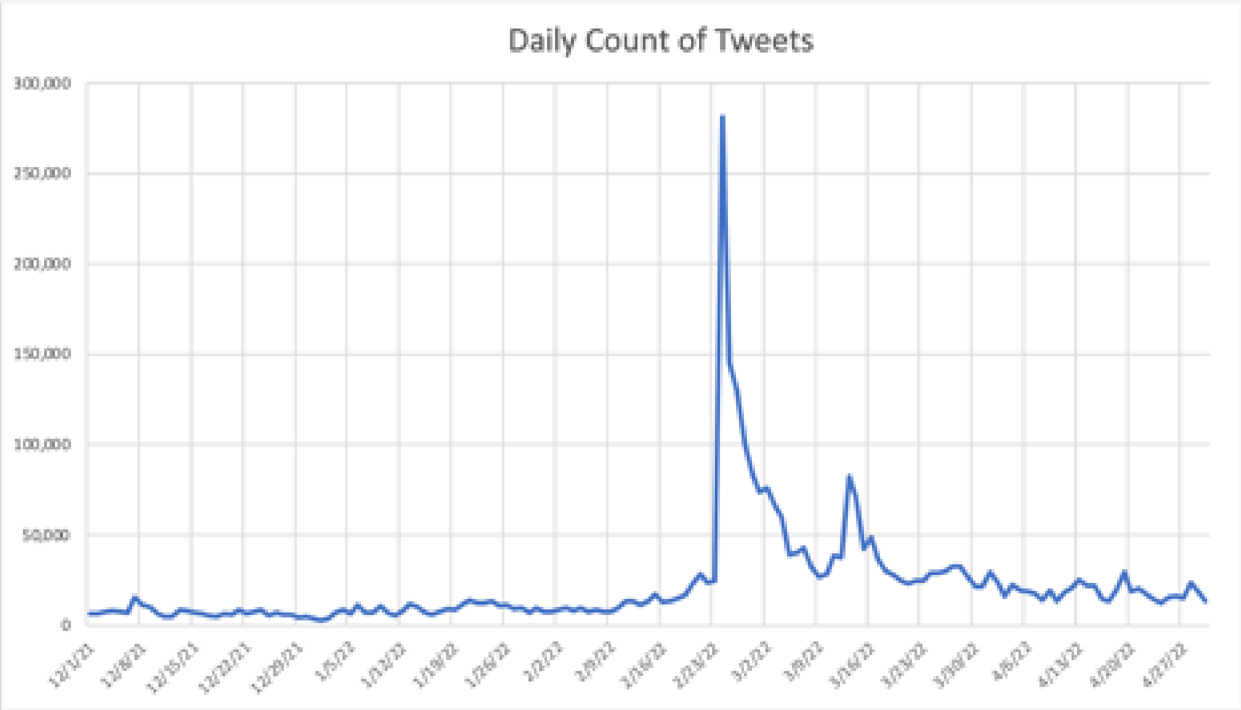

Figure 1. Daily Count of Tweets in both "Goldstein Positive" and "Goldstein Negative" Bigram Categories

Figure 2 displays the change in the Daily Sum of Sentiment, the start of the Ukraine War is on February 24th, 2022, which is evident by the large decrease in sentiment. The average sum of sentiment after the start of the War, like the count, was around three times more negative than before the start of the War. This indicates that, while more people were talking about geopolitics, that they were talking about it in a more negative way than usual, which is unsurprising given the scale and the devastation the Ukraine War caused during this time.

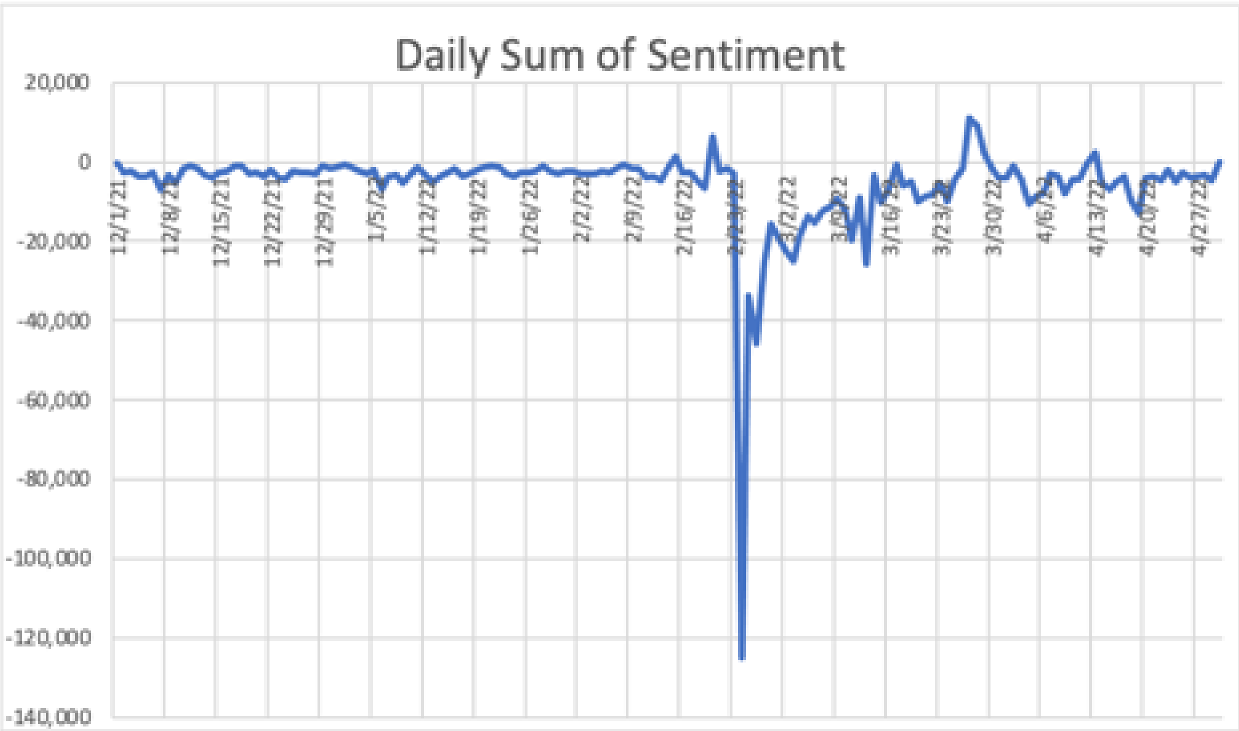

Figure 2. Daily Sum of Sentiment for all Tweets in both "Goldstein Positive" and "Goldstein Negative" Bigram Categories

Following a similar procedure to C. Pop et al., we chose to examine the Granger causality relationship between the daily sum of sentiment and each financial asset at different lagged values. In C. Pop et al., they used "the Granger causality was considered for one lag, five lags (a typical trading week), and 20 lags (the average number of trading days within a month)" (C. Pop et al., 2016, p. 132). Tables 4 – 6 show the results from the Granger Causality tests, where Table 4 shows the results of one lag (which represents one day), Table 5 shows the results of five lags, and Table 6 shows the results of 10 lags. Note that these tables only display the financial assets for which the sum of the sentiment time series provides predictive information of the change in the financial asset at the lagged value. Any asset that doesn't appear in the table, but does in Appendix A, either did not have any Granger causality with the sum of the sentiment, or it did, but the "feedback" (Granger, 1969, p. 5), denoted by "reverse okay" in our tables also passed. To check this "feedback", we see if the financial asset could provide information about the change in the sum of sentiment and pass the Granger causality test. Any test with a p-value less than 0.05 means that one can reject the Null Hypothesis of the Granger causality test and say that the daily sum of the sentiment time series does have forecasting, predictive information for the financial asset at the indicated lag. However, if the feedback test (which is when the variables of the original Granger causality test are reversed and tested) also passes, this would make the original Granger causality test meaningless as this would mean that the asset price trend and the change in the sum of sentiment trend would Granger cause each other, thus neither variable would contain predictive information about the other. Important to note here on the structure of the following tables, for organization, we included all the passing results for each lag, all in one table. Each entry in the tables below is the individual result for the Granger causality test between the change in the sum of sentiment and the financial asset.

Table 4. Granger Causality Results for Lag 1 Against the Different Financial Assets Separated by Asset Class.

Granger Causality	Sent_Sum	P-Value	Reverse Okay	Reverse P-Value
Gold Price	Yes	0.004988	Yes	0.438339
Gold Futures	Yes	0.040277	Yes	0.53574
Wheat Futures	Yes	0.002161	Yes	0.599003
German 10y Bond	Yes	0.043168	Yes	0.618356
FTSE 100	Yes	0.003308	Yes	0.880743
10 Year US Treasury	Yes	0.035314	Yes	0.82779
10 Year US Futures	Yes	0.022427	Yes	0.909091
EUR	Yes	0.035038	Yes	0.08545
GBP	Yes	0.006035	Yes	0.080256
AUD	Yes	0.044733	Yes	0.154349
MXN	Yes	0.021253	Yes	0.149475

Table 5. Granger Causality Results for Lag 5 Against the Different Financial Assets Separated by Asset Class.

Granger Causality	Sent_Sum	P-Value	Reverse Okay	Reverse P-Value
Oil Price	Yes	0.042714	Yes	0.904926
Oil Futures	Yes	0.03906709	Yes	0.928638
Wheat Futures	Yes	6.46596E-08	Yes	0.944131
German 10y Bond	Yes	0.005677	Yes	0.845183
FTSE 100	Yes	0.001391	Yes	0.860171
10 Year US Treasury	Yes	0.003009	Yes	0.714947
IG-ETF	Yes	0.044506	Yes	0.518171
10 Year US Futures	Yes	0.001307	Yes	0.708251
Bitcoin-Futures	Yes	0.013295	Yes	0.384176
2Y-Treasury Yield	Yes	0.030406	Yes	0.593837
GBP	Yes	0.005202	Yes	0.702139
MXN	Yes	0.04897	Yes	0.793235
RUB	Yes	0.000424	Yes	0.903236
Bitcoin	Yes	0.016982	Yes	0.333437

Many studies have looked at the Monthly (Caldara & Iacoviello, 2022; Niu et al., 2023; Yilmazkuday, 2024) or Daily levels (Pop, C. et al., 2016; Amen, 2020). However, we wanted to see if we could find forecasting information on an even smaller time scale. Thus, we redivided our tweet data into individual hours and reran the Granger Causality tests on a smaller subset of financial assets exhibited in Table 7 below.

Table 6. Granger Causality Results for Lag 10 Against the Different Financial Assets Separated by Asset Class.

Granger Causality	Sent_Sum	P-Value	Reverse Okay	Reverse P-Value
Gold Price	Yes	0.003813	Yes	0.489212
Oil Price	Yes	3.09826E-07	Yes	0.936389
Gold Futures	Yes	0.000121234	Yes	0.969781
Oil Futures	Yes	8.57092E-08	Yes	0.927111
Wheat Futures	Yes	1.3972E-14	Yes	0.608716
Nikkei 225	Yes	0.032658	Yes	0.680651
German 10y Bond	Yes	0.000445	Yes	0.826079
FTSE 100	Yes	0.001059	Yes	0.806077
10 Year US Treasury	Yes	0.000575	Yes	0.23006
Defense-ETF	Yes	0.005718	Yes	0.803637
Metals-ETF	Yes	0.039279	Yes	0.977376
10 Year US Futures	Yes	0.000117	Yes	0.212961
Bitcoin-Futures	Yes	0.010613	Yes	0.727189
EUR	Yes	0.000284	Yes	0.984743
GBP	Yes	0.000997	Yes	0.943754
MXN	Yes	0.000004	Yes	0.857881
RUB	Yes	0.000002	Yes	0.98621
Bitcoin	Yes	0.004504	Yes	0.214014

Table 7. Granger Causality Results for the Hourly Sum of Sentiment Time Series Against the Different Financial Assets Separated by Asset Class.

Granger Causality	Sent_Sum	Number of Lags (up to 24)	P-Value	Reverse Okay	Reverse P-Value
EUR	Yes	10	0.0196	Yes	0.0578
JPY	Yes	12	0.0262	Yes	0.7542
RUB	Yes	1	0.0019	Yes	0.3641
GBP	Yes	8	0.0362	Yes	0.1647
MXN	Yes	12	0.006	Yes	0.2507
EURGBP	Yes	5	0.0431	Yes	0.1953
AUD	Yes	10	0.018	Yes	0.35
ZAR	Yes	12	0.022	Yes	0.3978
BNB	Yes	1	0.0111	Yes	0.082
Metals-ETF	Yes	12	0.0443	Yes	0.5334
CSI-300	Yes	8	0.0386	Yes	0.5287
Sensex	Yes	5	0.0019	Yes	0.6829
FTSE 100	Yes	4	0.0013	Yes	0.2928
Gold Futures	Yes	6	0.0005	Yes	0.2823
Oil Futures	Yes	21	0.0179	Yes	0.2311

It should be noted that while the sum of sentiment was shown not to provide any predictive information for HY_ETF, the IG_ETF, and the Nikkei 225 within the 24 lags (representing at least one full day of data), it did show outside this limit, at 48 lags, 60 lags, and 30 lags respectively.

6. Discussion

With our initial tweet gathering, we found it unsurprising that more than twice the number of tweets were in the “Goldstein Negative” category, given the nature of the Ukraine war. However, this shows that not only does the “Goldstein Index” find significant geopolitical events, but when our sentiment analyses are run on the captured tweets, they return an accurate sentiment result as shown by the significant decrease in sentiment at the onset of the Ukraine War, followed by a sustained increase in negative sentiment relative to the before the War.

While we investigated 39 different financial assets time series, we found that only 11 assets were Granger causal with the sum of the sentiment from the “Goldstein Index” tweets at Lag 1, with the most immediate lag representing one day of trading. However, as we increased the number of lags, we found that more assets were Granger causal, i.e., the change in sum of sentiment provided predictive information for the change in the asset value (14 for Lag 5, a week of trading, and 19 for Lag 10, roughly two weeks of trading). One explanation for this is that news can take time to disperse and affect the market, especially with " sticky " assets, meaning their prices do not move quickly (Hayes, 2021). As Kleinnijenhuis et al. describe, “news impact may not be limited to short-term effects, however. Long-term graphs showed that hope versus fear sentiments in financial news preceded actual economic developments.” (Kleinnijenhuis et al, 2013). This means it may take time for certain financial asset prices to change in response to big geopolitical events such as the Ukraine War. Thus, as we increase the number of lags, which represent the number of days after the change in the sum of “Goldstein Index” sentiment, it might have provided time for the changes in the finance assets’ price to be realized and thus increasing the number of financial assets that the “Goldstein Index” sentiment change is predictive of. For example, the change in sentiment was Granger causal to Steel Futures at 20 lags, nearly a month of trading after the change in sentiment. That said, there were some assets even when the maximum number of lags were used, the “Goldstein Index” sentiment never showed any predictive information, such as USD vs CNY Foreign Exchange Rate, which means that the “Goldstein Index” sentiment time series would not have any use in predicting the change in value of the asset.

As for the successful analyses, our findings match Caldara and Iacoviello, who “document that stock returns experience a short-lived but significant drop in response to higher geopolitical risk. The stock market response varies substantially across industries, with the defense sector experiencing positive excess returns, and with sectors exposed to the broader economy, for instance, steelworks and mining, experiencing negative returns” (Caldara & Iacoviello, 2022, p. 3). This was shown by both the Defense ETF and the “Metals and Mining” ETF time series, which relate to the “Goldstein Index” sentiment, which was our proxy for geopolitical risk. Also, we discovered that both Oil Price and 2 Year US Treasury Bond Yield time series had a relationship with the “Goldstein Index” which Caldara and Iacoviello stated that their Geopolitical Risk Index had as well (Caldara & Iacoviello, 2022, p. 19). In addition, there was a mix of both the “risky” and the “haven” assets describe by Amen appeared (Amen, 2020, p. 6). However, the FTSE 100 was only the “risky asset” (Amen, 2020, p. 6), to appear in all three different lag tests. While the US Treasury 10 Year Yield, was the only “haven asset” (Amen, 2020, p. 6) to appear in all three lag tests. A few of our assets that we wanted to investigate appeared as well. Out of our assets, we found that both GBP/USD and USD/MXN appeared in all three lag tests.

One surprise in our Granger causality analyses was that only Bitcoin and BNB emerged among the cryptocurrencies. As Baur et al. find, “Bitcoin is mainly used as a speculative investment” (Baur et al, 2018, p. 2). Thus, we assumed that Bitcoin and the other cryptocurrencies would experience a price change due to their status as a “risky asset” (Amen, 2020, p.6) and the massive change in sentiment generated by the start of the Ukraine War. While this was not the case, these results are consistent with the findings of Rognone et al., which “suggest investor enthusiasm for Bitcoin irrespective of the sentiment of the news” (Rognone et al., 2020, p. 1). This also aligns with the results from Abraham et al. (2018, who found that tweet volume was a better indicator of Bitcoin and Ethereum price changes than tweet sentiment (Abraham et al., 2018, p. 2). However, their methodology specifically collected tweets with keywords for Bitcoin and Ethereum and only in English (Abraham et al., 2018, pp. 8 - 9). This difference in methodologies might explain the slight variation we observe with the sum of sentiment for the “Goldstein Index” tweets time series, having predictive information for the Bitcoin time series.

After completing the Daily Level analyses, we wanted to see if we could capture predictive information about financial assets at a smaller time interval when the sum of sentiment from “Goldstein Index” tweets was broken down to the hourly level. Amen’s Thorfinn Sensitivity Index analyses were conducted only on the daily level, as were those of Bollen et al., and Caldara and Iacoviello’s GPR Index (Caldara & Iacoviello, 2022, p. 1). Unlike the Daily Granger causality analyses, we found that many Forex assets, which trade 24 hours a day, responded to changes in sentiment within less than half a day. This result aligns with Rognone et al., who found that “Forex comoves and reacts homogeneously to news” (Rognone et al., 2020, p. 1). It reinforces Nofsinger's findings, which state that “financial markets adjust to changes in mood faster than real markets” (Nofsinger, 2005, p. 3). This discovery is important as it suggests that changes in sentiment can provide predictive information about shifts in the Forex time series over shorter time intervals than other geopolitical risk indices, potentially informing different trading options in Forex markets.

The USD vs RUB exchange rate was crucial, as the Ruble is the Russian currency. We found that at the Daily Level, a change in Goldstein Index sentiment had predictive power regarding changes in the USD vs RUB after five lags (approximately a week of trading). However, at the Hourly Level, changes in the Goldstein Index sentiment contained predictive information for USD vs RUB changes within one hour. The nature of the conflict and the varying time scales could explain this difference between the daily and hourly lags. News about the Ukraine War updated frequently, especially at its onset, resulting in rapid changes in Goldstein Index information. As previously mentioned, Forex markets tend to respond to news, so the low lag value at the Hourly Level was unsurprising. However, at the Daily Level, the predictive information from the Goldstein Index could possibly be explained by the amalgamation of data at that level. At the Daily Level, smaller hourly changes would be averaged out. While the Goldstein Index may have predictive information at a smaller time interval, at the Daily Level, aggregating all more minor changes might diminish the predictive information. Nevertheless, over time at the Daily Level, the overall trends between USD vs RUB and the Goldstein Index become clearer, explaining how the Goldstein Index demonstrates predictive power at five and ten lags, but not at the first lag (i.e., one day).

Two minor issues should be mentioned. The first is that we encountered the same issue as Bollen, et al., who detailed: “we have no knowledge of the 'ground truth' for public mood states, nor in fact for the particular subsample of the population represented by the community of Twitter.com users. This problem can only be addressed by increased research into direct assessments of public mood states vs. those derived from online communities such as Twitter” (Bollen et al., 2011). This is also related to a lack of a baseline econometric model for this data without any social media variables. Without the baseline model, the difference created by adding the change in the sum of sentiment time series as a variable to predict the change of the financial asset could not be found. However, this issue was outside of the purview of our study, as we were only investigating the relationship and not creating a prediction of how the financial asset would change. Additionally, by including a significant lead time (nearly three months) before the start of the Ukraine War, mitigates the effect described by Bollen, et al., as we were able to develop a baseline “ground truth” for X/Twitter sentiment regarding geopolitical risk. The second issue is with X/Twitter itself. While X/Twitter’s demographics have slightly balanced out over time, X/Twitter users are more often younger and male. Thus, while capturing more sentiment worldwide, we could capture an uneven demographic, potentially skewing our results.

Lastly, we see several avenues for extending our research further. Adding Russian and Ukrainian, while not as popular on X / Twitter, to the languages we captured and analyzed could change our results. These additions may capture changes in financial markets and assets more specific to the Eastern European and Central Asian markets that were greatly affected by the Ukraine War, like Yilmazkuday’s study. Another analysis examining the tweets captured by the “Goldstein Index” from May 2022 to the Present could prove interesting. Investigating how sentiment has changed since the initial outbreak of the war and seeing if the geopolitical risk sentiment still provides predictive information on the assets and markets in this study. There is the potential for more Arabic tweets in this period as Iran gets more involved in the Ukraine War. Also, focus on other big geopolitical risk events in the X / X/Twitter Age, such as Covid or the first Ukrainian invasion, to see if the “Goldstein Index” bigrams tweets produce similar results. Lastly, a study into sunflower seed futures could yield interesting results, as Ukraine was the largest producer of sunflower seeds before the War. Thus, sentiment around the Ukraine War might have predictive information about prices (Association, 2023).

7. Conclusion & Practical Implications

Using X/Twitter and sentiment analysis, we identified the start of the Ukraine War using the generic geopolitical bigrams from the “Goldstein Index.” We also show that the increased negative sentiment lasted for months, relating to the heightened geopolitical risk caused by the invasion. This rise in negative sentiment was also reflected in various financial assets and markets through Granger causality. Some immediate effects, like with many Foreign Exchange Rates, showed differences after a change in X/Twitter sentiment in only a few hours, while other markets, such as the Nikkei 225, took almost two weeks of trading before the change in sentiment provided predictive information relevant to changes in the financial market.

Statement of Researchers

Researchers' contribution rate statement:

John Corcoran Burns: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Validation; Visualization; Writing - original draft. **Tom Kelsey:** Funding acquisition; Project administration; Supervision; Writing - review – editing. **Carl Donovan:** Funding acquisition; Project administration; Supervision; Writing - review - editing

Conflict statement:

The authors declared no conflicts of interest.

Data Availability Statement:

Data can be made available upon request. Please get in touch with the corresponding author.

Acknowledgements:

We would like to thank Irlanda Márquez Cruz for her Spanish translations of the 'Goldstein Index' bigrams for this study. Also, the translators at Gengo.com for their work for the French, Portuguese, Arabic, Japanese, and Korean.

Funding:

This work was supported by the School of Computer Science of the University of St Andrews.

Use of AI Statement:

The authors submit that no generative AI tools or models were not used in any part of this study.

Ethical Considerations:

Approved by the School of Computer Science Ethics Committee for the University Teaching and Research Ethics Committee (UTREC) at the University of St Andrews.

The Authors' Biographies

John Corcoran Burns, PhD, graduated from the School of Computer Science at the University of St Andrews. Specializes in natural language processing, multilingual analysis, and geopolitical risk.

Tom Kelsey is a Professor in the School of Computer Science at the University of St Andrews. Specializes in data-driven research, using advanced computational techniques to identify, retrieve, classify, and analyze data to generate hypotheses and develop models. His primary interest is the modelling of human fertility both for the healthy population and for survivors of childhood cancer.

Carl Donovan, Honorary Research Statistician and Lecture in the School of Mathematics and Statistics at the University of St Andrews and the Director and Consultant at DMP Statistical Solutions. Specializes in statistical modelling with an interest in Earth and Planetary Sciences.

8. References

2018 Research on 100 Million Tweets: What It Means for Your Social Media Strategy for Twitter. (2018) Vicinitas https://www.vicinitas.io/blog/twitter-social-media-strategy-2018-research-100-million-tweets#language.

Twarc. (2023) Twarc https://twarc-project.readthedocs.io/en/latest/.

World Map: Simple. (2022) Map Chart https://www.mapchart.net/world.html.

Abouzahra, M., & Tan, J. (2021) Twitter Vs. Zika—the Role of Social Media in Epidemic Outbreaks Surveillance. *Health Policy and Technology*, vol. 10, no. 1, pp. 174-81, doi:https://doi.org/10.1016/j.hlpt.2020.10.014.

Abraham, J., Higdon, D., Nelson, J., & Ibarra, J. (2018) Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis. *SMU Data Science Review*, vol. 1, no. 1, https://scholar.smu.edu/datasciencereview/vol1/iss3/1.

Altig, D., Baker, S. R., Barrero, J. M., Bloom, N., Bunn, P., Chen. S., Davis. S. J., Leather, J., Meyer. B.H., Mihaylov, E., Mizen, P., Parker, N. B., Renault .T., Smietanka, P., & Thwaites, G. (2020) Economic Uncertainty before and During the Covid-19 Pandemic. Working Paper Series, *National Bureau of Economic Research*, doi:10.3386/w27418.

Amen, S. (2020) Political Market Making: Trading Financial Markets Using Thorfinn Political Indices. Data, Foreign Exchange, General, vol. 2023, *Cuemacro*, https://www.cuemacro.com/2020/06/26/political-market-making/.

Aroussi, R. (2023) Reliably Download Historical Market Data from with Python. *Ran Aroussi* https://aroussi.com/post/python-yahoo-finance.

Association, National Sunflower (2023) *World Supply & Disappearance*. National Sunflower Association https://www.sunflowernsa.com/stats/world-supply/.

Augustop. (2019) *Portuguese Tweets for Sentiment Analysis*. Kaggle https://www.kaggle.com/datasets/augustop/portuguese-tweets-for-sentiment-analysis?select=TweetsWithTheme.csv.

Baker, S. R., Bloom, N., Davisc, S. J., & Renaultd, T. (2021) Twitter-Derived Measures of Economic Uncertainty. Twitter-based Uncertainty Indices, Economic Policy Uncertainty, May 13th, 2021, pp. 1-14. general editor, *Economic Policy Uncertainty*, https://www.policyuncertainty.com/media/Twitter_Uncertainty_5_13_2021.pdf.

Baur, D. G., Hong, K., & Lee, A. D. (2018). Bitcoin: Medium of Exchange or Speculative Assets? Journal of International *Financial Markets, Institutions and Money*, vol. 54, pp. 177-89, doi:https://doi.org/10.1016/j.intfin.2017.12.004.

Beykikhoshk, A., Arandjelovic, O., Phung, D., & Venkatesh, S. (2015). Using Twitter to Learn About the Autism Community. *Social Network Analysis and Mining*, vol. 5, no. 1, doi:10.1007/s13278-015-0261-5.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, vol. 2, no. 1, pp. 1-8, doi:10.1016/j.jocs.2010.12.007.

Brady, W. J., Will,s J. A., Jost, J. T., & Van Bavel, J. J. (2017). Emotion Shapes the Diffusion of Moralized Content in Social Networks. *Proceedings of the National Academy of Sciences*, vol. 114, no. 28, pp. 7313-18, doi:10.1073/pnas.1618923114.

Burns, J. C. (2024). *Automatic-GR* GitHub. https://github.com/jb370/Automatic-GR

Caldara, D., & Iacoviello, M., (2022). Measuring Geopolitical Risk. *American Economic Review*, 112(4) ,1194-225, doi:10.1257/aer.20191823.

Cambria, E. (2013). An Introduction to Concept-Level Sentiment Analysis. Advances in Soft Computing and Its Applications. *MICAI 2013*, Heidelberg, Berlin, doi:https://doi.org/10.1007/978-3-642-45111-9_41.

Cañete, J., Chaperon, G., Fuentes, R., Ho, J. H., Kang, H., & Pérez, J. (2020) BETO: Spanish BERT. *ICLR 2020*, https://github.com/dccuchile/beto?tab=readme-ov-file

CNBC. (2023). *U.S. 2 Year Treasury*. CNBC https://www.cnbc.com/quotes/US2Y.

Darkmap. (2016) *japanese_sentiment/data*, Github.com, https://github.com/Darkmap/japanese_sentiment/tree/master/data

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Google AI Language*. doi:10.48550/arxiv.1810.04805.

Engelberg, J., & Parsons, C. A. (2009). The Causal Impact of Media in Financial Markets. Workshop in Behavorial Finance, *Yale University*, pp. 1-44. http://www.econ.yale.edu/~shiller/behfin/2009_11/engelberg-parsons.pdf.

Gamebusterz. (2017) *xac*, French-Sentiment-Analysis-Dataset, Github.com, https://github.com/gamebusterz/French-Sentiment-Analysis-Dataset/bl ob/master/xac

Gamebusterz. (2017) *xaj,* French-Sentiment-Analysis-Dataset, Github.com, https://github.com/gamebusterz/French-Sentiment-Analysis- Dataset/blob/master/xaj

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly

Gilbert, E., & Karahalios, K. (2010). Widespread Worry and the Stock Market. *Proceedings of the International AAAI Conference on Web and Social Media,* vol. 4, no. 1, 2010, pp. 58-65, doi:10.1609/icwsm.v4i1.14023.

GoldHub. (2023). *Gold Spot Prices*. GoldHub https://www.gold.org/goldhub/data/gold-prices..

Goldstein, J. S. (1992). A Conflict-Cooperation Scale for Weis Events Data. *The Journal of Conflict Resolution*, vol. 36, no. 2, pp. 369-85, https://www.jstor.org/stable/174480.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, vol. 37, no. 3, p. 424, doi:10.2307/1912791.

Granger, C. W. J. (2003). Time Series Analysis, Cointegration, and Applications. *Nobel Prize*. https://www.nobelprize.org/uploads/2018/06/granger-lecture.pdf.

Hayes, A. (2023). *What Is Price Stickiness? Definition, Triggers, and Example*. ECONOMICS. Investopedia https://www.investopedia.com/terms/p/price_stickiness.asp#:~:text=%22Sticky%22%20is%20a%20general%20economics,that%20is%20resistant%20to%20change.

Hutto, C. J. & Gilbert, E, (2014). Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *International AAAI Conference on Weblogs and Social Media (ICWSM),* http://eegilbert.org/papers/icwsm14.vader.hutto.pdf.

Inoue, G., Alhafni, B., Baimukan, N., Bouamor, H., & Habash, N., (2021). The Interplay of Variant, Size, and Task Type in Arabic Pre-Trained Language Models. Proceedings of the Sixth Arabic Natural Language Processing Workshop, *Association for Computational Linguistics*, https://huggingface.co/CAMeL-Lab/bert-base-arabic-camelbert-mix-sentiment.

Investing.com. (2023a) *Germany 10-Year Bond Yield*. Investing.com https://www.investing.com/rates-bonds/germany -10-year-bond-yield-historical-data.

Investing.com. (2023b). *United States 2-Year Bond Yield*. Investing.com https://www.investing.com/rates-bonds/u.s.-2 -year-bond-yield-historical-data.

Kleinnijenhuis, J., Schultz, F., Oegema, D., & van Atteveldt, W. (2013) Financial News and Market Panics in the Age of High-Frequency Sentiment Trading Algorithms. *Journalism*, vol. 14, no. 2, pp. 271-91, doi:10.1177/1464884912468375.

Lee, S., Jang, H., Baik, Y., Park, S., & Shin, H. (2020). KR-BERT: a small-scale Korean-specific language model, *arXiv*, https://doi.org/10.48550/arXiv.2008.03979

Lhessani, S. (2023). *Python: How to Get Live Market Data (Less Than 0.1-Second Lag)* Medium https://towardsdatascience.com/python-how-to-get-live-market-data-less-than-0-1-second-lag-c85ee280ed93.

LiveCharts.co.uk. (2023) *Live Charts - Crude Oil Chart*. LiveCharts https://www.livecharts.co.uk/MarketCharts/crude.php.

Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., de la Clergerie, É., Seddah, D., & Sagot, B. (2020). CamemBERT: a tasty french language model, *Proceedings of the 58th Annual Meeting of the Association of Computer Linguistics*, https://aclanthology.org/2020.acl-main.645/

McClelland, C. (2006). World Event/Interaction Survey (WEIS) Project, 1966-1978. *Inter-university Consortium for Political and Social Research* [distributor], doi: 10.3886/ICPSR05211.v3

Monitor, Markets. (2023) *Metals & Mining Overview.* ETF.com https://www.etf.com/topics/metals-mining.

Niu, Z., Wang, C., & Zhang, H. (2023). Forecasting stock market volatility with various geopolitical risks categories: New evidence from machine learning models. *International Review of Financial Analysis*, vol. 89, October 2023, https://doi.org/10.1016/j.irfa.2023.102738

Nofsinger, J. R, (2005). Social Mood and Financial Economics. *Journal of Behavioral Finance*, 6(3), 144-60, https://doi.org/10.1207/s15427579jpfm0603_4

Pak, A., & Paroubek, P. (2010) Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Seventh International *Conference on Language Resources and Evaluation ({LREC}'10),* European Language Resources Association (ELRA), http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf.

Park, L. (2015) *Nsmc*. GitHub.com https://github.com/e9t/nsmc.

Pearse, B. (2021) *Human and Machine Translation: Both Alive and Kicking — and Here to Stay*. Translation & Localization Blog. Smart Cat https://www.smartcat.com/blog/human-and-machine-translation-both-alive-and-kicking-and-here-to-stay/

Pop, C., Bozdog, D., Calugaru, A., & Georgescu, M. A. (2016) *Chapter 7 - an Assessment of the Real Development Prospects of the Eu 28 Frontier Equity Markets.* Handbook of Frontier Markets, Academic Press, 2016, pp. 117-46, https://doi.org/10.1016/B978-0-12-803776-8.00007-0

Pota, M., Ventura, M., Fujita, H., & Esposito, M. (2021) Multilingual Evaluation of Pre-Processing for Bert-Based Sentiment Analysis of Tweets. *Expert Systems with Applications*, vol. 181, p. 115-119, https://doi.org/10.1016/j.eswa.2021.115119.

Prabhakaran, S. (2023) *Granger Causality Test in Python*. Machine Learning Plus https://www.machinelearningplus.com/time-series/granger-causality-test-in-python/.

Rajput, N. K., Grover, B. A., & Rathi, V. K. (2020) Word Frequency and Sentiment Analysis of Twitter Messages During Coronavirus Pandemic. *Computing Research Repository (CoRR)*, vol. abs/2004.03925, https://doi.org/10.48550/arxiv.2004.03925.

Rognone, L., Hyde, S., & Zhang, S. S. (2020) News Sentiment in the Cryptocurrency Market: An Empirical Comparison with Forex. *International Review of Financial Analysis*, vol. 69, https://doi.org/10.1016/j.irfa.2020.101462.

Smith, A. (2023) *23 Essential Twitter Statistics to Guide Your Strategy in 2023*. Sprout Blog. sproutsocial https://sproutsocial.com/insights/twitter-statistics/.

Souza, F., Nogueira, R., & Lotufo, R. (2020) BERTimbau: pretrained BERT models for Brazilian Portuguese, Intelligent Systems: *9th Brazilian Conference, BRACIS 2020*, https://dl.acm.org/doi/10.1007/978-3-030-61377-8_28

Standards, National Institute of. (2023) *Engineering Statistics Handbook: Stationarity. Process or Product Monitoring and Control*. National Institute of Standards https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc442.htm#:~:text=Stationarity%20can%20be%20defined%20in,no%20periodic%20fluctuations%20(seasonality).

Strycharz, J., Strauss, N., & Trilling, D. (2017) The Role of Media Coverage in Explaining Stock Market Fluctuations: Insights for Strategic Financial Communication. *International Journal of Strategic Communication*, vol. 12, pp. 1-19, https://doi.org/10.1080/1553118X.2017.1378220.

TASS 2020. (2020) Workshop on Semantic Analysis at Sepln 2020. *TASS 2020,* http://tass.sepln.org/2020/.

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, vol. 62, no. 3, pp. 1139-1168, https://doi.org/10.1111/j.1540-6261.2007.01232.x

Tetlock. P. C., Saar-Tsechansky, M., & MacSkassy, S. (2008) More Than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance,* vol. 63, no. 3, pp. 1437-67, doi:10.1111/j.1540-6261.2008.01362.x.

Thurman, W., & Fisher, M. E. (1988). Chickens, Eggs, and Causality, or Which Came First? *American Journal of Agricultural Economics*, vol. 70, no. 2, pp. 237-38, https://webdoc.agsci.colostate.edu/koontz/arec-econ535/papers/thurman%20fisher%20(ajae%201988).pdf.

Tohoku-nlp (2020). *Bert-base-japanese,* Hugging Face, https://huggingface.co/tohoku-nlp/bert-base-japanese

Uhl, M. W. (2014). Reuters Sentiment and Stock Returns. *Journal of Behavioral Finance*, vol. 15, no. 4, pp. 287-98, https://doi.org/10.1080/15427560.2014.967852.

Wold, C. (2023). *Top 3 Defense Etfs (Ppa, Xar).* Investopedia https://www.investopedia.com/news/top-3-defense-etfs- ppa-xar/.

Yilmazkuday, H. (2024). Geopolitical Risk and Stock Prices. Department of Economics, *Florida International University,* Working Paper 2407, https://economics.fiu.edu/research/working-papers/2024/2407.pdf

YuJeong, S., Yun, D. Y., Hwang, C., & Moon, S. J. (2021). Korean Sentiment Analysis Using Natural Network: Based on Ikea Review Data. *International Journal of Internet, Broadcasting and Communication*, vol. 13, no. 2, pp. 173-78 https://doi.org/http://dx.doi.org/10.7236/IJIBC.2021.13.2.173.

Zach. (2023). *How to Perform a Granger-Causality Test in Python.* statology https://www.statology.org/granger-causality-test-in-python/.

Zote, J. (2025). *45+ Twitter (X) stats to know in marketing in 2025*. Sprout Blog. Sproutsocial https://sproutsocial.com/insights/twitter-statistics/

Appendix A

Table A1 below lists the financial assets and markets we analyzed for our study. The Caldara, Iacoviello, and Amen financial assets or markets come directly from their papers. Our Own assets or markets come from a few different sources. We were interested in expanding on the assets listed in the other papers (such as Gold Futures and Oil Futures), and we also wanted to look at smaller international markets or emerging markets (such as the Sensex or the USD-MXN FX Rate). In addition, we wanted to see if different cryptocurrencies outside of Bitcoin reacted differently to geopolitical events. Lastly, since our study involved the Ukraine War, we wanted to see how the Natural Gas markets and Wheat Market responded to the crisis, as both Russia and Ukraine are two of the world's largest producers of Wheat, and Russia is the primary source of Natural Gas for Europe. Table A2 is a reordering of the assets based on asset class.

Table A1. We analyzed the financial assets and markets and their sources.

Source	Financial Asset or Market
Caldara and Iacoviello (5)	Defense ETF, Metals and Mining ETF, Crude Oil Price, 2 Year US Treasury Yield, Steel Futures
Amen (19)	S&P 500 Index (US Stock Exchange), Morgan Stanley Capital International Index ("MSCI"), CSI 300 Index (Chinese Stock Exchange), FTSE 100 Index (UK Stock Exchange), Nikkei 225 (Japanese Stock Exchange), Bitcoin, USD vs. EUR, JPY, AUD, CNY, RUB, and ZAR FX Rates, VIX Index (Volatility Index), MSCI Futures, Bitcoin Futures, US High Yield (HY) ETF, US Investment Grade (IG) ETF, Gold Price, 10 Year US Treasury Yield
Our Own (15)	Gold Futures, Crude Oil Futures, 10 Year US Treasury Yield Futures, S&P BSE Sensex (Indian Stock Exchange), 10 Year German Bond Yields, USD vs. GBP, MXN FX Rates, EUR-GBP FX Rate, Ethereum (ETH), ChainLink (LINK), Ripple (XPR), Binance Coin (BNB), Algorand (ALGO), Wheat Futures, Natural Gas Futures

Table A2. The Financial Assets and Markets we analyzed grouped by Asset Class.

Asset Class	Asset or Market
Commodity (7)	Gold Price, Crude Oil Price, Gold Futures, Crude Oil Futures, Steel Futures, Wheat Futures, Natural Gas Futures
International Markets and Assets (5)	CSI 300, Nikkei 225, BSE Sensex, FTSE 100, 10 Year German Bond Yield
U.S. Based Markets and Assets (12)	S&P 500, MSCI, VIX, 2 Year US Treasury Yield, 10 Year US Treasury Yield, Defense ETF, Metals and Mining ETF, US HY ETF, US IG ETF, 10 Year US Treasury Yield Futures, Bitcoin Futures, MSCI Futures
Foreign Exchange Markets (9)	USD vs. EUR, JPY, GBP, AUD, MXN, ZAR, RUB, CNY, and EUR-GBP
Crypto Currencies (6)	Bitcoin, ETH, Link, XPR, BNB, ALGO

Appendix B

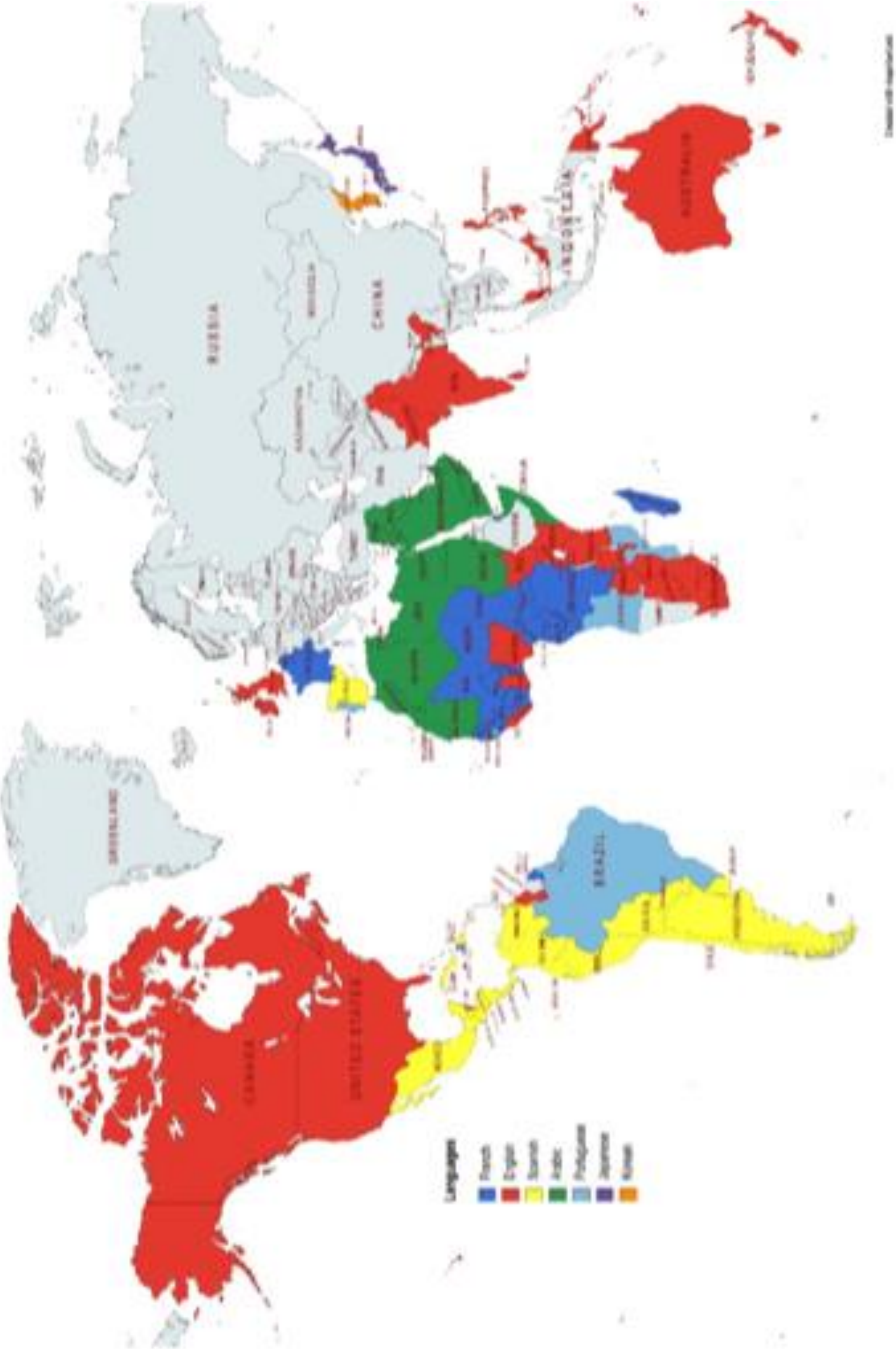

Figure B1. Map of countries that where one of the seven languages is one of the national languages used by that country. From World Map: Simple