The Knightian Uncertainty Effect of Financial Dissemination on Social Media: A Twitter Informed Knightian Uncertainty Approach to Modeling Financial Markets

Buz Galbraith
Siddhant Kumar

February 1st 2022

1: Introduction

Over the past few years Twitter has become a key forum for public discourse, among both institutions and individuals. Further, the discourse on Twitter, especially regarding finance, which has historically been viewed as an inaccessible field, has reached new audiences. There are some that claim, however, that influential users on Twitter can have a disproportionate effect on public opinion and even potentially manipulate the market for their own gain. Many point to Elon Musk's May 1st 2020 tweet[1], "Tesla Stock price is too high imo." as a prime example. This leads one to question whether these tweets, despite their notoriety, actually influence expectations of prices for market participants in a qualitative way.

Among the broad categories of models developed to explain financial market fluctuations, the two most widely known are the Rational Expectations and Behavioral models. Recently, however, there has been a resurgence in interest around Knightian Uncertainty models. In this paper we seek to compare the performance of Bloomberg-informed Rational Expectations Hypothesis, Behavioral and KU Models at predicting movement in the stock market. Additionally, under the assumption that due to its unpredictable nature Twitter can be understood as a Knightian uncertainty factor, we hypothesize that augmenting the standard Knightian model to consider Twitter as a Knighian factor may lead to an increase in predictive accuracy.

2: Data Collection

2.1: Bloomberg Data Collection

The first task of this paper was to collect data which would support qualitative comparisons of the stock market under Rational Expectations, Behavioral and Knightian Uncertainty based interpretations. This was done using Frydman and Mangee's approach to analyzing Bloomberg Market Wraps. Every day over the period of analysis, which comprises the months of August, September, October and November 2021, we read that day's Bloomberg Market Wrap and scored market factors relevant to the three aforementioned models; factors were primarily categorized as Fundamental, Psychological, Technical and Knightian Uncertainty factors. They were encoded on a scoring sheet using three discrete values (1, 2, and 3) depending on the relevant factor's effect on that day's financial market movement, where a score of 2 represents a non-factor event. Similarly, a score of 1 represents an event that negatively impacted stock market prices and a score of 3 represents the opposite.

[1] https://twitter.com/elonmusk/status/1256239815256797184?lang=en, Jan 14th, 2022

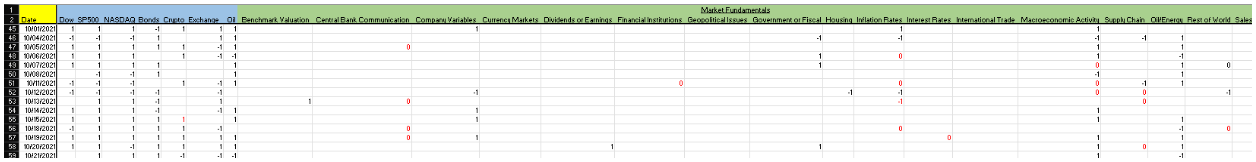


Figure 1: A cross-section of the Bloomberg Market Wrap coding sheet depicting Fundamental factors.

During the month of October, for example, when Earnings Season drove financial markets consistently upwards, the Fundamental factor "Dividends or Earnings" was often scored a 3 to reflect that positive earnings news that directly lifted financial markets, as depicted in figure 1. Similarly, figure 2 details Psychological factors, and additionally included a column which tracked trading days over which Knightian Uncertainty factors occured. Additionally, another sheet was completed containing a more thorough scoring of Knightian Uncertainty factors over the period of analysis, which informs our later models. It should be noted that this second sheet includes several one-off events that did not influence overall market movement, such as changes in company variables.

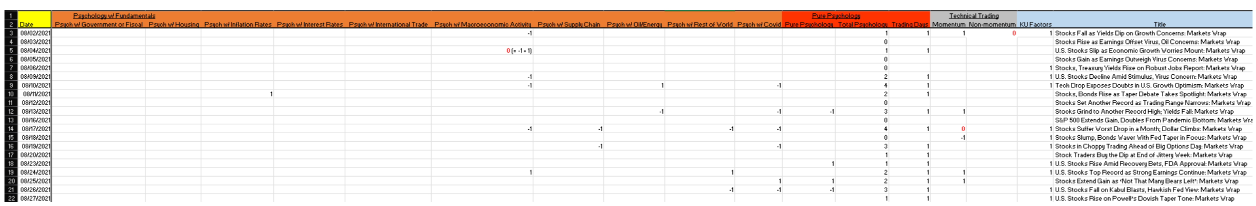


Figure 2: A cross-section of the Bloomberg Market Wrap coding sheet depicting Psychological, Technical, and Knightian Uncertainty factors. Additionally, it lists Market Wrap titles.

2.2: Twitter Data Collection

The second major data collection task of this paper was collecting market sentiment information from Twitter. This was an especially interesting task for us, as unlike the Bloomberg data collection, there was not a standard approach allowing for much more methodological freedom. Our macro level approach to gathering this data was to collect tweets from a certain subset of users over a period of analysis, scan tweets for only those we feel contain terms relevant to market expectations, conduct sentiment analysis on said tweets, and finally aggregate those tweets by time.

2.2.1: Gaining Access to Twitter Data

In order to properly analyze tweets our first task was to actually get access to Twitter Data. We did this through the Twitter Developer API (Application Programing Interface)[2]. To get access to

[2] https://developer.twitter.com/en, Jan 14th, 2022

this API, we had to submit an application for a Student Twitter Developer account which allowed for the scraping of 2,000,000 tweets per month. We used this API to gather JSON files containing tweet data with a granularity of up to 200 uncleaned tweets per month per user. We then conducted further analysis of these tweets in Python as will be discussed later on in this paper.

2.2.2: Choosing Users

A key determining question of how effective our analysis of Twitter as a Knightian Uncertainty factor could be was: what kind of information were our tweets catching? In order to explain our approach to this it is first important to return to our hypothesis, which is that users on Twitter distill information as a Knightian Uncertainty factor which influences their followers' expectations of prices in financial markets. Thus our goal became to use twitter data in as predictive a way as possible, given our technical and time limitations. To do this we aimed to analyze users who have influential effects on as varied an audience as possible. For instance, someone who follows a German Bloomberg correspondent may represent a different subset of market participants than someone who follows a Tiwanese Politician; they may both be interested in markets and their expectations could be altered by the information they are exposed to from users they follow on Twitter.

Our method to capture this variation was firstly when possible to analyze individuals, as opposed to companies, as they are likely to be more opinionated. For instance, the official Tesla Twitter account[3] is run by a social media management team and primarily tweets positive news regarding the company, while Elon Musk, the CEO of Tesla, tweets[4] about a much wider variety of content in a more opinionated manner. Thus, despite the fact that these two accounts are both representative of Tesla, we argue that Elon Musk's Twitter would give us more "information gain". Further we tried when possible to vary users based on geographic region and occupation in order to capture varying perspectives and cast a wider net of sentiment.

User	**Region**	**Occupation**
Tucker Carlson	North America	Reporter for Fox
Ben Shapiro	North America	Conservative Political Commentator
Ezra Klien	North America	Founder of VOX
Qiao Collective	Asia	Chinese Focused Media Collective

[3] https://mobile.twitter.com/tesla?lang=en, Jan 14th, 2022
[4] https://twitter.com/elonmusknewsorg?lang=en, Jan 14th, 2022

Hua Chunying	Asia	Chinese Assistant Minister of Foreign Affairs,
Jack Dorsey	North America	Twitter CEO
Nathan Law	Asia	Exiled Hong Kong Politician
Tsai Ing-wen	Asia	Prime Minister Of Taiwan
Phoebe Bridges	North America	American Musician
Paul Kerry	Asia	Copy Editor of the Korean Herald
David M. Friedman	Middle East	Former US Ambassador to Israel

Figure 3: Depicts a subsection of users who had their tweets scrapped. It is color coded based on region, and lists the occupation of the user as well. The overall breakdown of users were: 15 from North America, 10 from Asia, 5 from the Middle East and 5 from Europe. A full list of users can be found in the appendix.

2.2.3: Choosing Tweet Fields

The Twitter API has a number of potentially useful "Tweet Fields"[5] that could be collected from each API call. For the purposes of this paper, we simply collected User Name, User ID, Tweet ID, and Date Created, which were important for the purposes of sorting and reporting our data. Additionally, we kept track of Public Metrics, a field that contains a number of subfields about engagement including Number of Retweets, Like Counts, etc., as a measure of engagement. It is worth noting that a tweet with 100 likes may have affected less people than one with 1,000,000 likes. We did not weigh our tweets during analysis in any way based on this information; we simply felt it may be useful in a qualitative sense to keep.

2.2.4: Choosing Terms Relevant to Market Expectations

We assert that analyzing the sentiment of all tweets from each user during a period of time would be less effective than analyzing some subset of those tweets which we feel are relevant. We ultimately chose this subset to be only those tweets containing certain terms. For our purposes we chose to read over Bloomberg Market Wraps for the semester and picked around 250 terms we felt were relevant to our data. This was an area that, though important, is entirely subjective. As such, the code has been implemented in a way that one can easily add or remove any term they wish from this analysis with a single function call.

[5] https://developer.twitter.com/en/docs/twitter-api/fields, Jan 14th, 2022

Terms
interest rates
interest rate
central bank
central bank
the fed
fed chairman
inflation
price
prices
deflation

Figure 4: A small subset of the terms we used when scraping tweets. The full list can be found in the appendix under section 8.

2.2.6: Sentiment Analysis for Twitter Data

One of the key areas of interest for us was conducting sentiment analysis on the actual Twitter data after it was collected. It is important to note that, though one of us is a programmer with some experience in general machine learning, neither of us prior to writing this paper had a proper background in Natural Language Processing (NLP). As such, our approach was heavily influenced by *Sentiment analysis algorithms and applications: A survey* (Medhat, Hassan, Korashy, 2013)[6]. This paper further contained figure 5, which suggested two distinct approaches to sentiment analysis, a machine learning based approach and a lexicon based approach. We chose to implement sentiment analysis based on both types as a point of comparison.

[6] https://www.sciencedirect.com/science/article/pii/S2090447914000550, Jan 14th, 2022

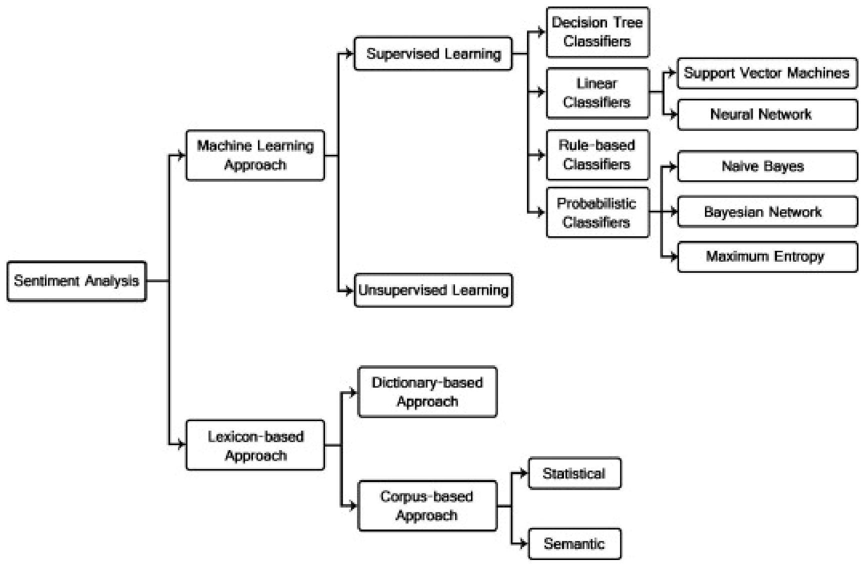


Figure 5: A broad overview of sentiment analysis approaches.

2.2.6 A: Lexicon-based Approach

For a lexicon-based approach we looked into a number of lexicons before settling on one. Ultimately, we chose the VADER (Valence Aware Dictionary and Sentiment Reasoner)[7] Lexicon. We chose this approach because it is specifically designed to analyze social media data, and thus seemed well suited for our project. From an implementation standpoint, this lexicon is dictionary based and thus was highly dependent on data cleaning, meaning that it did not attempt to understand sentences in a sequential manner, but instead as a sum of each word's individual sentiment score.

2.2.6 B: Machine Learning Approach

There were many more options for the implementation of a machine learning approach to sentiment analysis. Researching all of the options was exceedingly interesting for us, and showed us how deep this field is. Ultimately, we settled on using the Google Cloud Sentiment Analysis API[8]. This is an incredibly robust, general purpose sentiment analysis algorithm that has historically performed well, and analyzes sentences in a sequential manner.

2.2.7: Twitter Data Examples

[7] https://github.com/cjhutto/vaderSentiment, Jan 14th, 2022
[8] https://cloud.Google.com/natural-language/docs/analyzing-sentiment, Jan 14th, 2022

Figure 6: Depicts an example of a tweet that was captured by our approach. Though it is not directly related to financial markets, it indicates the users' belief of increasing instability, which may lead market participants following this user to lower their expectations of future prices.

	User Name	orignal texts	created a	retweet cour	reply cour	like coun	quote cour	id
0	Elon Musk	@pcgamer amd has been great to work with!	2021-09-	882	720	21064	130	1.443E+18
1	Elon Musk	@gfilche @freshjiva no guarantees, but i think it will. this is a big part of what i meant by teslaâ€™s long-term competitive advantage being manufacturing technology.	2021-09-	456	289	5737	86	1.443E+18
2	Elon Musk	@sciguyspace @verge ðŸ¤£ðŸ¤£	2021-09-	421	803	15977	36	1.443E+18
3	Elon Musk	@thesheetztweetz spacex has sued to be "allowed" to compete, bo is suing to stop competition	2021-09-	998	527	10147	85	1.443E+18
4	Elon Musk	@spacedotcom maxwell was incredible	2021-09-	907	688	7987	437	1.443E+18
5	Elon Musk	@mortenlund89 really?	2021-09-	142	451	3628	17	1.443E+18

Figure 7: Depicts an example of a few columns of the raw data set. It has: an "index" column; a "User Name" column; an "original text" column with the text of the tweet, (which, pre-data cleaning, can have issues with special characters); "created at" column, which contains the date and time the tweet was written; "retweet count", "like count", and "quote count" columns, which are public metrics, as well as an "id" column, which lists each tweet's unique ID from Twitter.

id	User Name	orignal texts	created at	retweet cc	reply coun	like count	quote cou	cleaned texts	Google Se	Google Ma	VADER Sentimen
1.44E+18	Elon Musk	@thesheetzt	2021-09-1	786	658	13259	52	thesheetztweetz nasakennedy spacex	0	0.9	-0.8
1.44E+18	Elon Musk	@matt_lowne	2021-09-1	168	198	4461	4	matt lowne forward flaps will change a l	0.3	0.3	0
1.44E+18	Elon Musk	@wholemarst	2021-09-1	1171	682	8513	160	wholemarsblog this is written by ford ua	-0.8	0.8	0
1.44E+18	Elon Musk	@icannot_enc	2021-09-1	365	460	6514	67	icannot enough garyblack with 10 1 it w	0.4	0.4	-0.1
1.44E+18	Elon Musk	@wholemarst	2021-09-1	398	701	6437	188	wholemarsblog to be fair investors are g	0.3	0.3	5.1
1.44E+18	Elon Musk	@timsweeney	2021-09-1	350	271	6373	56	timsweeneyepic please challenge tim of	-0.1	0.1	0.2
1.44E+18	Elon Musk	@teslarati @r	2021-09-1	483	566	8733	66	teslarati residentsponge highway stack i	-0.1	0.1	0
1.44E+18	Elon Musk	@spacex thes	2021-09-1	1134	1012	19530	78	spacex these are v1 5 starlinks with laser	0.2	0.2	0

Figure 8: Depicts a few examples of our analyzed data. It has all the original columns discussed, except the "index" column which has been dropped; the "ID" column is used as the index instead to help with data processing. Another new column is the "cleaned texts" column, which lists the tweets after they have been cleaned, making them easier for the sentiment analysis algorithms to analyze. Additionally, a "Google Sentiment" column which represents how positive or negative a tweet is, i.e., the overall emotional leaning of a text on a scale of (-1,1) using the Google API has been added. The "Google Magnitude" column, which is an unbounded score ranging from negative infinity to positive infinity, representing the total amount of emotional content present in the tweet, formed using the Google API has also been added. The final new column is the "VADER Sentiment" column, which represents a similar measure to Google Magnitude using the VADER algorithm.

date	Google Sentiment	Google Magnitude	VADER Sentiment
8/1/2021	0.300000012	5.300000027	33.2
8/2/2021	1.200000039	5.60000003	-3.9
8/3/2021	-1.50000006	12.30000006	20
8/4/2021	-2.100000017	8.000000007	-5.9
8/5/2021	2.599999994	3.99999997	30.8
8/6/2021	2.600000046	12.19999988	51.5
8/7/2021	2.90E-08	6.300000132	24
8/8/2021	1.300000041	7.200000047	30
8/9/2021	-0.700000042	12.49999991	13.1

Figure 9: Depicts the Twitter data aggregate on a daily level.

In most cases, we practically wanted to analyze Twitter sentiment over a day. For this purpose, we simply added all of the scores that took place on that day to get a data frame as depicted in Figure 9. This is helpful as the granularity of our Twitter data now matches that of the Bloomberg Data, allowing for comparison.

2.3: Stock Market Indicator Data Collection

To establish a point of comparison, we needed a means by which to access movement in the stock market over our period of analysis. To accomplish this we used "ytfinance", a Python package that aggregates historical data from Yahoo Finance[9]. Using this package, we collected daily close price data over the period of analysis for the S&P 500. We chose to only analyze the S&P 500 as opposed to multiple indices for two main reasons. Firstly, different indices may be moved in different directions by different effects, thus making analysis more difficult, and secondly that in order to compare indices one would have to somehow standardize the scales which would require distributional assumptions that we hope to avoid. We assume this data to be representative of the true price movement in the stock market.

[9] https://finance.yahoo.com/quote/%5EGSPC/, Jan 14th, 2022

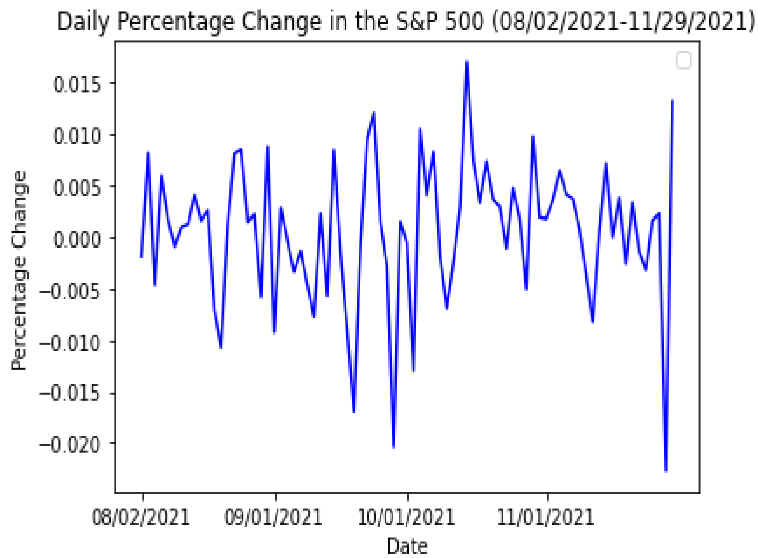


Figure 10: Depicts the percentage change in the S&P 500 over the whole period of analysis.

3: Methodology

3.1: Combining Models

The goal of this paper is to conduct a qualitative analysis on the performance of a number of stock market models over our period of analysis. The word qualitative is key here as we are not ultimately trying to understand a quantitative measure like stock prices, but instead qualitative co-movements in our data in an attempt to compare the accuracy of our chosen models. This is relevant as avoiding quantitative measures allows for the comparison of models without worrying about standardizing scales, avoiding unwanted distributional assumptions. To this end, when representing models we simply took a product of the daily net of the previous model and the daily net of the aggregate of the novel data. The product was a suitable approach for qualitative analysis as regardless of the scale between data, the sign could shift by taking products to reflect changes in accuracy. This process is acceptable because we are comparing models in a primarily qualitative sense, which ignores excessive scales and instead focuses on relative comparisons.

As a concrete example, to represent the Behavioral model, we considered the net of the daily aggregate REH factors times the net of the daily behavioral factors. So, for instance, on our first day of analysis, 08/02/2021, we had a net fundamental factors of 32 and a net Behavioral factor of 37 resulting in a Behavioral model representation of 32*27=1184. We used the same approach for the other models.

Special attention should be paid to our method of representing the augmented KU models, however, as we chose to weight the Twitter Sentiment Analysis data equally as compared to the standard KU factors. This should be considered further when interpreting results.

3.2 Choosing a Metric of Comparison

In order to formalize how well the models fit our S&P 500 stock market data, we need a metric of comparison. As this is a qualitative comparison we chose to work with the Spearman's Rank Correlation Coefficient[10](SCC). As this metric is a non-parametric measure of the relationship between rank variables, it was ideal for this form of analysis. We implemented the metric using the following formula $r_s = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2}\sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2}}$ where R and S are two rank variables. This formula is in effect computing the covariance of the two rank variables divided by the product of their standard deviations. An important assumption of this model is that no two data points are equally ranked. We adjusted our data to deal with this by more heavily ranking instances from later in the period of analysis. So for instance were we to have the ranks [1,1,2,3] we adjust this as [1,2,3,4].

It is also important to note that the Spearman Correlation Coefficient is an ordinal ranking metric. Meaning, it measures correlation in when the data archives highs or lows in effect, and thus it is possible for models to graphically look similar and still have a low Spearman Correlation Coefficient as they achieve extremes at different times. This fact is relevant to keep in mind when interpreting results.

3.3: A Note On Scales

As mentioned above, this analysis is qualitative and the Spearman metric measures correlation in rank variables. As such, for the remaining graphs in this paper the scale of the Y axis can be functionally disregarded, as it has little to do with interpretation. It is instead important to look at the overall shape of the graphs in terms of their movements of the common X axis, representing Date.

[10] https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient

3.4: Models

It is important to formalize how the term "model" will be used throughout this paper. In the spirit of econometrics, we use the term "model" to refer to our attempts to represent established parametric functions used by economists that attempt to model stock prices such as the REH, Behavioral and KU models. In this paper, we take the principles of these models, i.e., the factors upon which they rely to model stock market prices and use the corresponding data gathered from Bloomberg to represent these models. Additionally, we augment our representation of the established Knightian Uncertainty model with Twitter sentiment analysis data in an attempt to gauge the applicability of considering Twitter when modeling stock market prices.

In particular, we constructed a number of model types for this paper. First, the traditional models based off of Rational Expectations, Behavioral and Knightian Uncertainty approaches. Second, Knightian Uncertainty models augmented with Twitter sentiment analysis. These include KU models augmented with Google Sentiment, KU models augmented Google Magnitude and KU models augmented VADER Sentiment. Additionally, KU models augmented with Google Magnitude and Google Sentiment, KU models augmented with Google and VADER Sentiment, KU models augmented with Google Magnitude and VADER Sentiment and finally KU models augmented with all Twitter based models were included.

3.5: General Coding Methodology

The coding section of this paper was done by Galbraith. All code and the data are available at (https://github.com/buzgalbraith/The-Knightian-Uncertainty-Effect-of-Financial-Dissemination-on-Social-Media-A-Twitter-Informed-Kni).

4: REH, Behavioral and Knightian Uncertainty Models

All Bloomberg scoring factors are listed in the appendix under section 7.

4.1: REH Models and the Data

The Rational Expectations Hypothesis, as an interpretation of Muth's pivotal Rational Expectations Hypothesis, rests on the key assumption that the economist's model formalizes how every market participant forecasts outcomes *explicitly* in terms of fundamental market factors, some of which include company earnings, dividends, and Central Bank communications, etc.

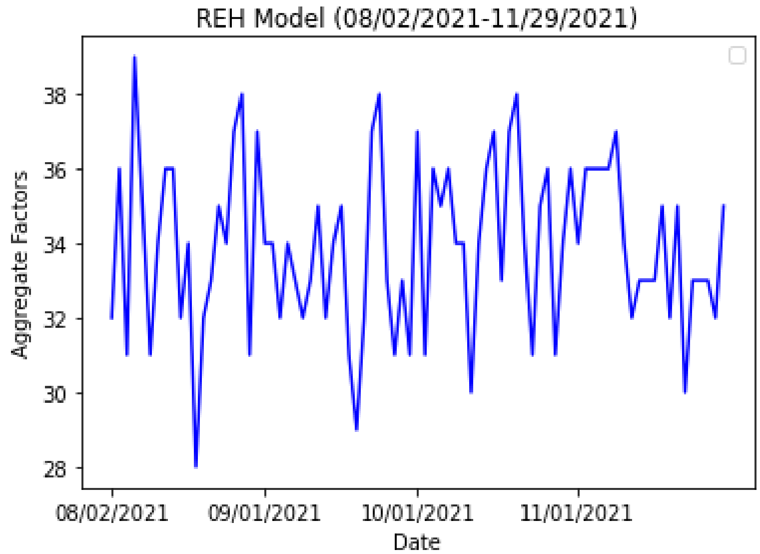


Figure 11: Depicts the REH model over the whole period of analysis.

Our REH approach seeks to model change in the stock market solely through fundamental factors present in Bloomberg Market Wraps from Muth's vision. Our sample period consists of 83 trading days, over which Fundamental factors appeared **100%** of the time.

4.2: Behavioral Models and the Data

Similarly, the Behavioral model rests on the key assumption that the economist's model formalizes how every market participant forecast outcomes based not only on fundamentals, but also psychological factors. In addition to measuring overall market sentiment, which is represented by pure psychology, these additional psychological factors capture change in sentiment around fundamental factor behavior. Some examples of these factors include psychology around earnings or dividends, and psychology around market activity.

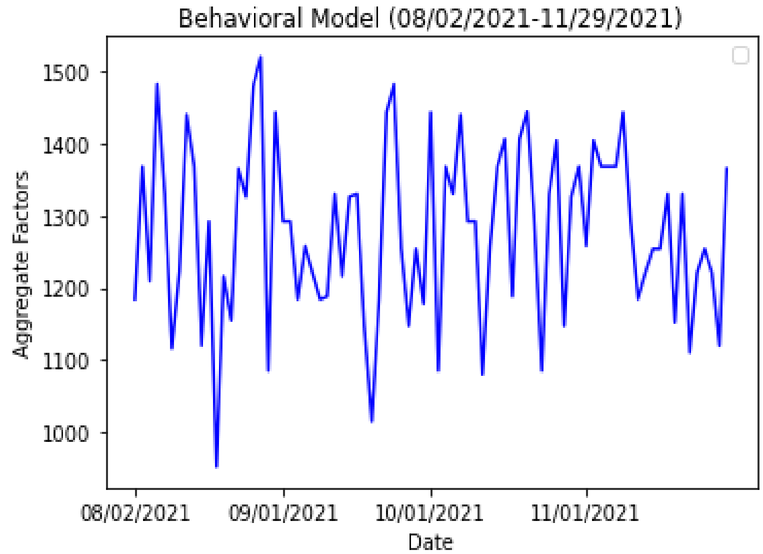


Figure 12: Depicts the Behavioral model over the period of analysis.

Across the 83 trading days, Behavioral factors appeared on 57 days, which equates to **68.7%** of the time.

4.3: Knightian Uncertainty Models and the Data

The Knightian Uncertainty model posits that movements in financial markets cannot be accurately predicted by standard probabilistic models due to unforeseeable change. This is formalized using Knightian Uncertainty factors, i.e., non-repetitive events which may cause non mechanical changes in financial markets, examples of which include the emergence of a new strain of COVID-19 or the appointment of a new Fed Chair. As such, the Knightian Uncertainty model understands movements in financial markets as the combination of fundamental factors, market psychology and these Knightian Uncertainty events.

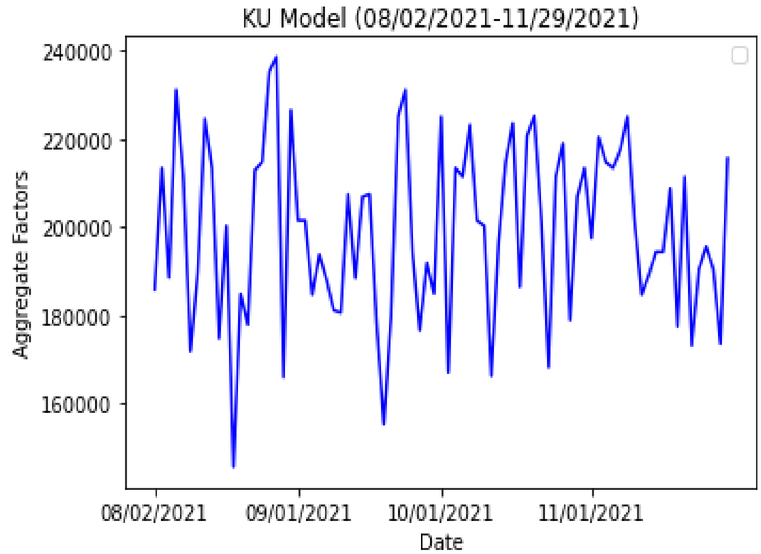


Figure 13: Depicts the KU model plotted against macroeconomic indicator variables over the period of analysis.

Over 83 trading days, Knightian Uncertainty factors appeared on 47 trading days, which is equivalent to **56.6%** of the time.

4.4: Comparing Models

We used the Spearman metric to compare how well each model fits the S&P 500 data. Based on figure 14 below, the REH and Behavioral models perform similarly over the entire period of analysis, resulting in SCCs near **0.62**. In contrast, the KU model performs slightly better, reporting an SCC of **0.64**. These results are consistent with our hypothesis - that Knightian Uncertainty models better capture movement in financial markets as a result of allowing for non-mechanical changes. This is particularly relevant because the period of analysis is affected by several major KU events, one of which is the emergence of the Omicron strain of COVID-19. This leads one to question whether analyzing additional Knightian factors from sources outside of Bloomberg Market Wraps would improve the model.

Model	Spearman Correlation Coefficient
REH Fundamentals	0.6164143135429193
Behavioral Psychology	0.6246818116790729
Knightian Uncertainty Model	0.6447609986424802

Figure 14: Depicts the Spearman Correlation Coefficient of the primary models.

5: Augmented KU Models

5.1: KU Models with Twitter Data Overall

Following evidence that a Knightian Uncertainty based approach more accurately fits our data, we would now like to see if augmenting this approach with Twitter data increases the SCC, thus increasing predictive accuracy.

As described in figure 15 A, we ran a number of models and obtained the following results:

Model	Spearman Correlation Coefficient	Change in Spearman from KU Model (%)
REH Fundamentals	0.6164143135429193	-4.396463985762742
Behavioral Psychology	0.6246818116790729	-3.1142061951146607
Knightian Uncertainty Model	0.6447609986424802	-
KU Model + Google Sentiment	0.018855957521587547	-97.07551207947006
KU Model + Google Magnitude	0.13939231586461392	-78.3807773488007
KU Model + VADER Sentiment	0.01444528573985637	-97.75959064362293
KU Model + Google Sentiment + Google Magnitude	-0.045760076313858616	-107.09721531082133
KU Model + Google Sentiment + VADER Sentiment	0.08866310199633914	-86.24868715958068
KU Model + VADER Sentiment + Google Magnitude	-0.03241022884607473	-105.0267043004017
KU Model + VADER Sentiment + Google Magnitude + Google Sentiment	0.04720908780462384	-92.67804847005003

Figure 15 A: Depicts each model's Spearman Correlation Coefficient and their corresponding percentage changes in error as compared to the unaugmented Knightian Uncertainty model over the whole period of analysis.

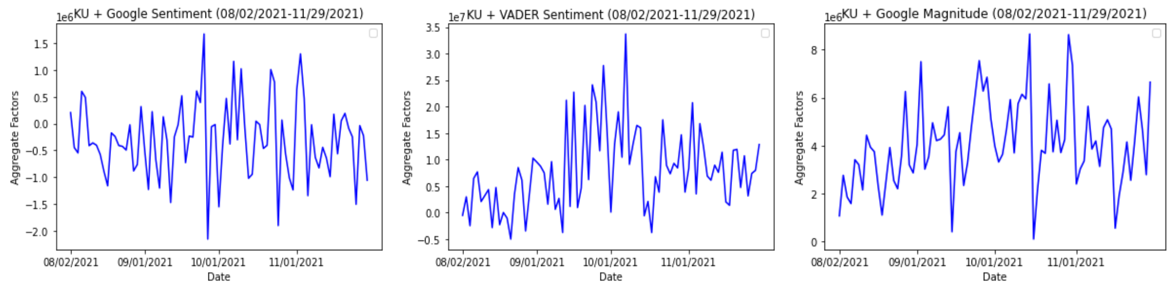


Figure 15 B: Depicts the augmented models against the S&P 500 over the whole period of analysis.

As can be seen from the Spearman metrics in figure 15 A, adding Twitter data as a Knightian Uncertainty factor over the whole period of analysis leads to a decrease in predictive accuracy of the Augmented Knightian Uncertainty models, resulting in a performance substantially worse than not only the standard KU model, but also the REH and Behavioral models. It is, however, worth noting that there were a number of major KU events during the period of analysis. For example, during the months of October and November, when particularly concerning CPI data was released, market participants were concerned over whether the reported high inflation figures were transitory. Similarly, market participants were divided over the possible repercussions of the Fed announcing a bond-tapering program in August.

The theory underpinning Knightian Uncertainty models would predict that such KU events would result in structural changes, necessitating changes in our representation of these models. Thus, constructing models based on a restricted period of analysis may result in higher predictive accuracy.

5.2: KU Models with Twitter Data on a Monthly Basis

We first constructed models on a monthly basis in order to test the hypothesis of whether restricting the period of analysis would increase predictive accuracy.

5.2.1: August

We define August as Month 1. Notice first from figure 16 A that overall the month of August is volatile, with a generally building trend in the stock market during the first portion of the month, followed by a sharp crash in the middle and succeeded by another increasing movement at the end of the month. Despite frequent swings, the overall magnitude of the data remains relatively close to the whole period mean.

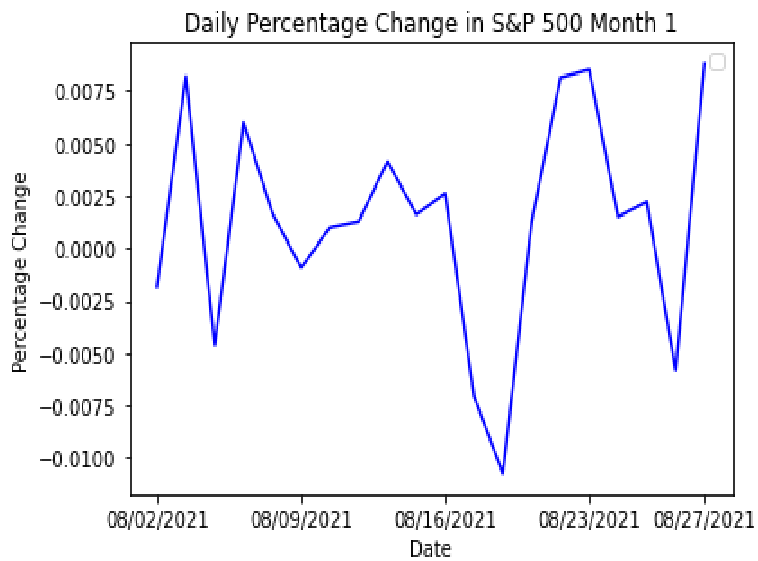


Figure 16 A: Depicts the percentage change in our S&P 500 data over the month of August.

Model	Spearman Correlation Coefficient	Change in Speakman from KU Model (%)
REH Fundamentals	0.7073693258453706	5.414278389278816
Behavioral Psychology	0.6451880027210173	-3.852167110899893
Knightian Uncertainty Model	0.6710374881409934	-
KU Model + Google Sentiment	0.10556554264161433	-84.26830922158052
KU Model + Google Magnitude	0.159668951647083085	-76.20565848125393
KU Model + VADER Sentiment	0.4594714679152688	-31.52819685407381
KU Model + Google Sentiment + Google Magnitude	0.08482669826521039	-87.35887342148204
KU Model + Google Sentiment + VADER	0.05586043592349575	-91.67551188857593
KU Model + VADER Sentiment + Google Magnitude	0.28325186368323263	-57.788965789685086
KU Model + VADER Sentiment + Google Magnitude + Google Sentiment	0.0889812994919421	-86.73974240419098

Figure 16 B: Depicts each model's SCC and their corresponding percentage changes in error as compared to the unaugmented Knightian Uncertainty model over the month of August.

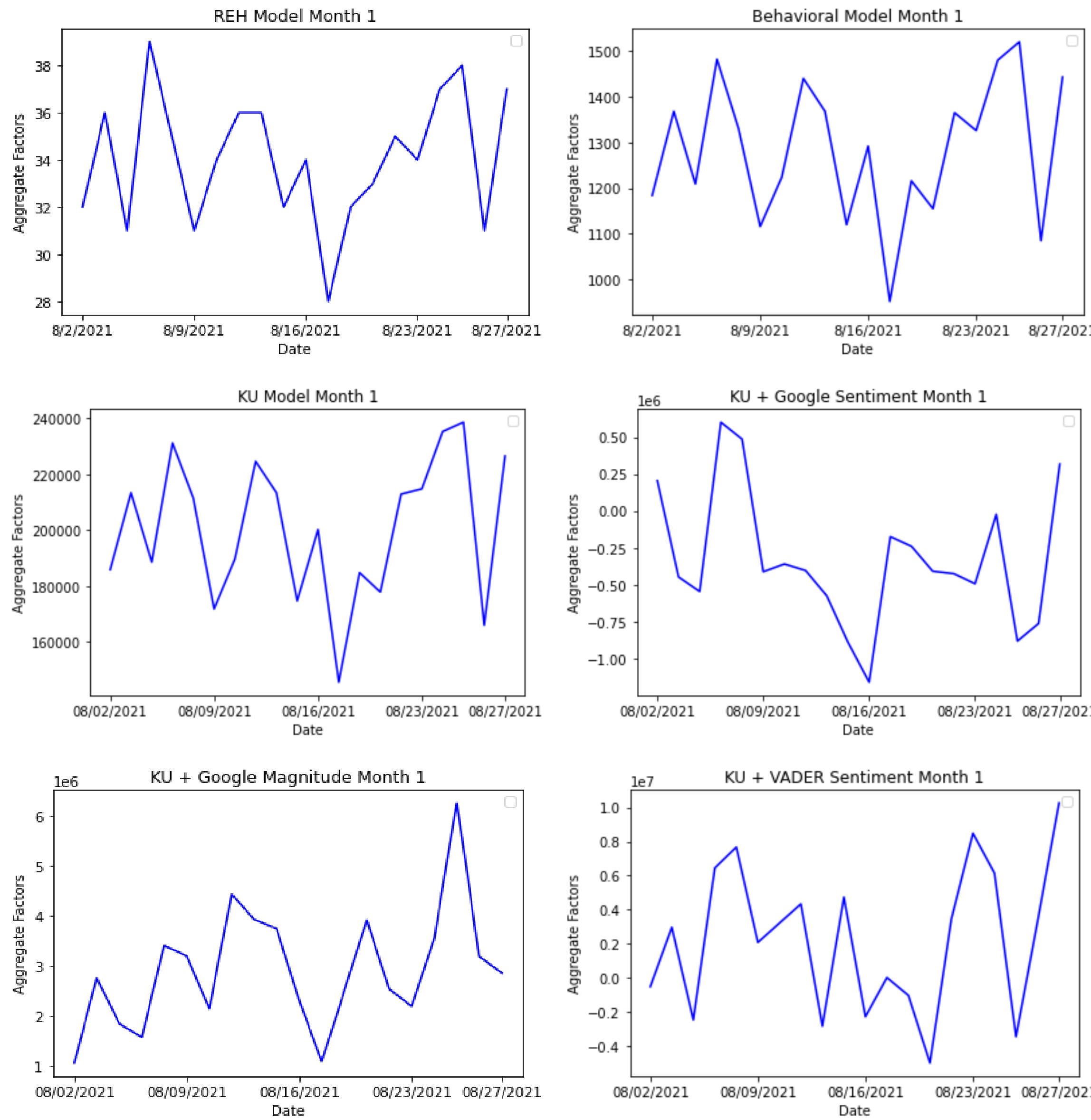


Figure 16 C: Depicts a subsection of the unaugmented models against the S&P 500 over the month of August.

As can be seen by looking at figure 16 B, the unaugmented Knightian model has a SCC of **0.67**. Contrasting this result, the KU models augmented with Twitter data perform significantly worse. Specifically, the KU + Google Sentiment Model results in a SCC of **.10**, which equates to an approximately **84%** reduction of Spearman Correlation. This seems to hold with observation of the graphs. Consider for instance the KU + Google Magnitude graph in figure 16 C. It seems to predict a building, albeit volatile, trend in the S&P 500 over the period of month, which is simply not present in the data. In contrast, the REH model reported a higher SCC then the augmented models with a value of **0.70** which is an approximately **5%** improvement over the KU Model.

5.2.2: September

We define September as Month 2. Looking at figure 17 A, one can see that the first half of the month experienced a generally stable upward trend, followed by a series of downturns and recoveries in the later half of the month.

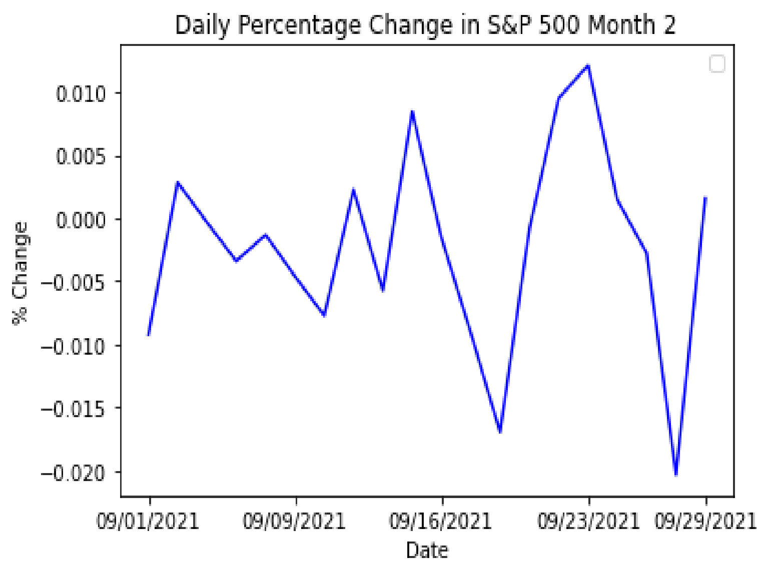


Figure 17 A: Depicts the daily percentage change in the S&P 500 over the month of September.

Model	Spearman Correlation Coefficient	Change in Speakman from KU Model (%)
REH Fundamentals	0.6375312221398862	-8.746179150853456
Behavioral Psychology	0.6784501420145977	-2.8891988902565937
Knightian Uncertainty Model	0.6986350995579697	-
KU Model + Google Sentiment	0.171568195290601	-75.44237393753146
KU Model + Google Magnitude	-0.006831563367655972	-100.97784428122469
KU Model + VADER Sentiment	-0.049922572071428827	-107.14572916577126
KU Model + Google Sentiment + Google Magnitude	0.12857800865013871	-81.59582753121182
KU Model + Google Sentiment + VADER	0.20844307457626493	-70.16424243383305
KU Model + VADER Sentiment + Google Magnitude	-0.11259601897307643	-116.11657058803897

KU Model + VADER Sentiment + Google Magnitude + Google Sentiment	0.166992999023114	-76.09725032012112

Figure 17 B: Depicts each model's SCC and their corresponding percentage changes in error as compared to the unaugmented Knightian Uncertainty model over the month of September.

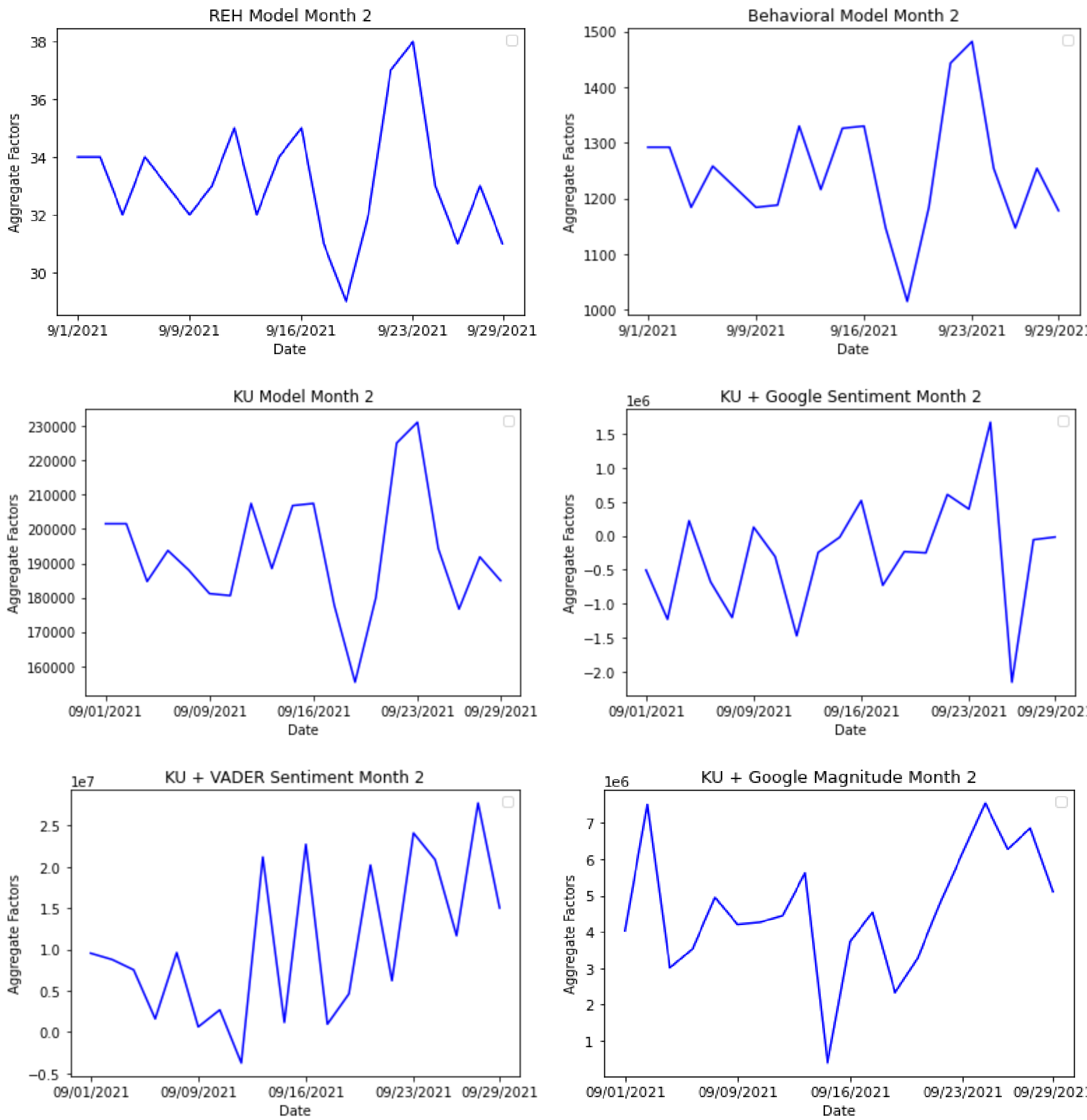


Figure 17 C: Depicts a subsection of the unaugmented models against the S&P 500 over the month of September.

Our unaugmented Knightian Uncertainty model had an SCC of around **0.67**, whereas our augmented Knightian Uncertainty models did poorly overall and largely decreased predictive accuracy. Specifically, the SCC of the KU Model + VADER Sentiment + Google Magnitude + Google Sentiment is **0.16**, resulting in an approximate **76%** decrease in SCC. This interpretation is also supported by the graphs, as the KU Models seem to capture the overall shape of the S&P 500 significantly better over the course of the month. The REH and Behavioral models also reduced this month, albeit to a much lower extent, with the Behavioral model yielding results **0.67** which is an approximately **3%** reducinton in SCC.

5.2.3: October

We define October as Month 3. We notice from figure 18 A that there is a general building trend for much of the first half of the month followed by a mild declining trend in the later half.

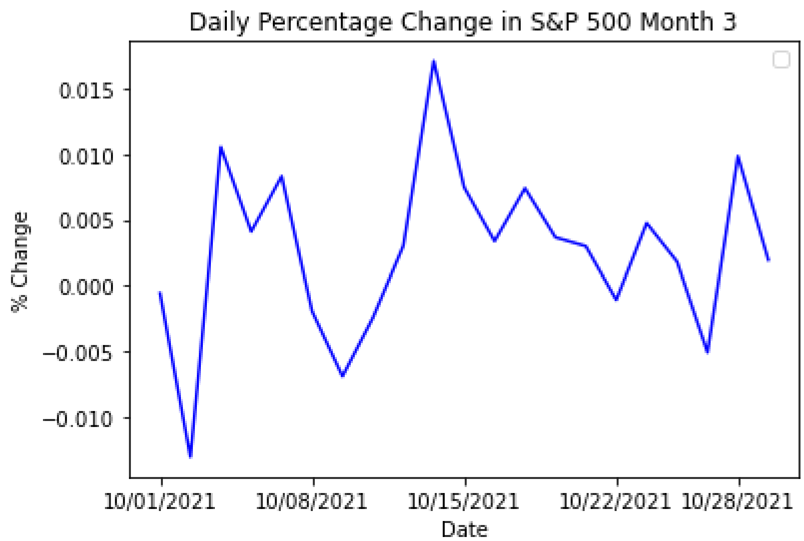


Figure 18 A: Depicts the daily percentage change in the S&P 500 over the month of October.

Model	Spearman Correlation Coefficient	Change in Speakman from KU Model (%)
REH Fundamentals	0.6005517100249415	-4.060389792351277
Behavioral Psychology	0.6064116933149389	-3.124243073790864
Knightian Uncertainty Model	0.6259684698792563	-
KU Model + Google Sentiment	-0.1053265525884387	-116.8261753836827
KU Model + Google Magnitude	0.2715899567346497	-56.61283757837883
KU Model + VADER Sentiment	-0.031412242465239715	-105.01818286012055
KU Model + Google Sentiment + Google Magnitude	-0.22955954595752404	-136.67270110294916
KU Model + Google Sentiment + VADER	0.08593802308802746	-86.27118980855312
KU Model + VADER Sentiment + Google	-0.016138966042740108	-102.57823945123836

Magnitude		
KU Model + VADER Sentiment + Google Magnitude + Google Sentiment	0.015291398657158738	-97.5571615196356

Figure 18 B: Depicts each model's SCC and their corresponding percentage changes in error as compared to the unaugmented Knightian Uncertainty model over the month of October.

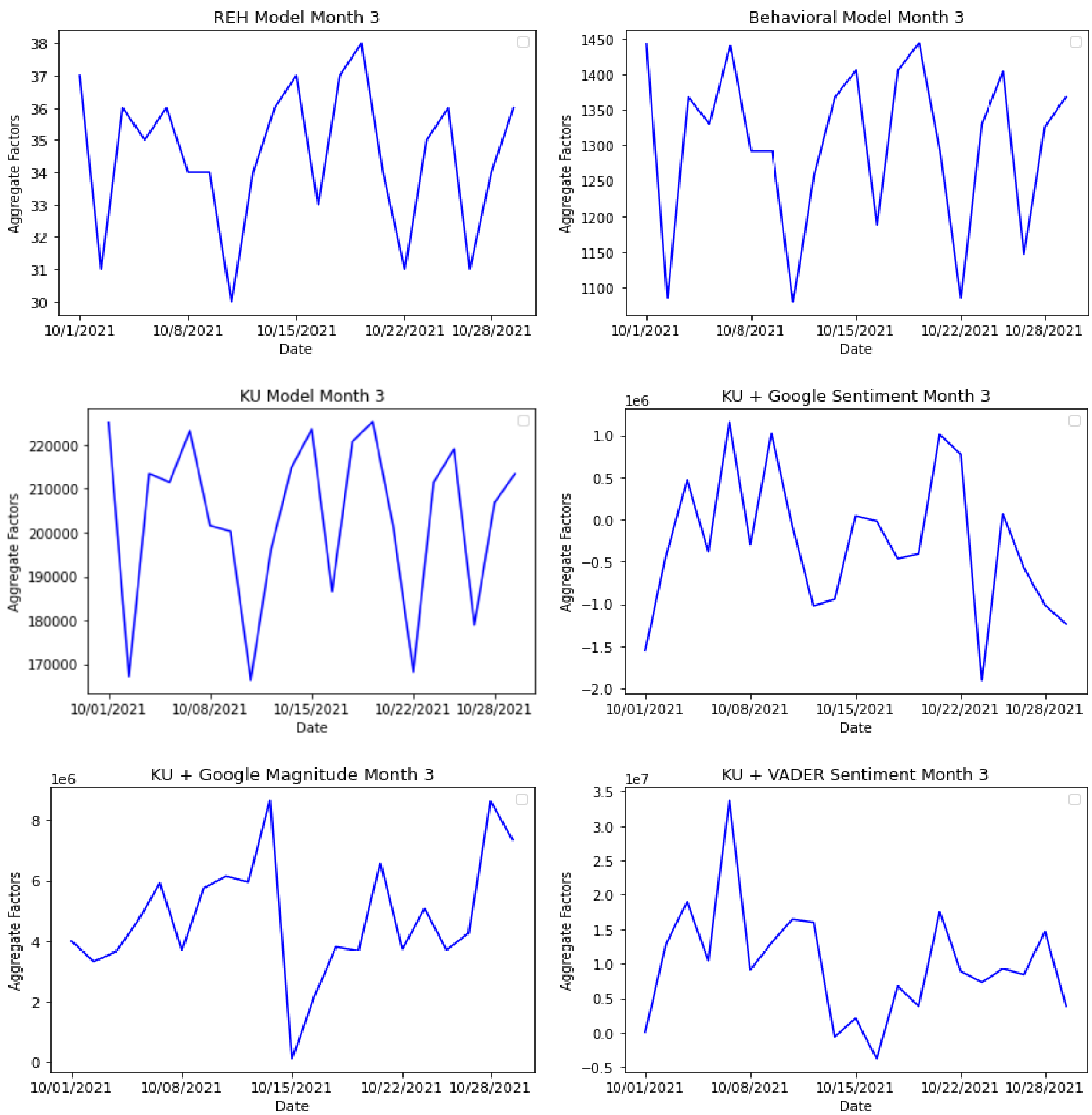


Figure 18 C: Depicts a subsection of the unaugmented models against the S&P 500 over the month of September.

The augmented Knightian Uncertainty model again performed substantially worse this month, with the KU + VADER model having an SCC of approximately **-0.0314** In contrast, the unaugmented Knightian Uncertainty model had an SCC of **0.625**, resulting in a **-105%** change in correlation. Again, the graphs support this conclusion, as in general the Twitter augmented KU models produce graphs notably dissimilar to that of the S&P 500. This can be again compared to the REH and Behavioral models which have only slight reduction in error from the standard KU model. With the Behavioral model having an SCC of **.606**, which is **3%** lower than the KU model.

5.2.4: November

We define November as Month 4. Looking at figure 19 A, we found that November's stock market movement was relatively close to zero with a large downturn and recovery near the end of the month.

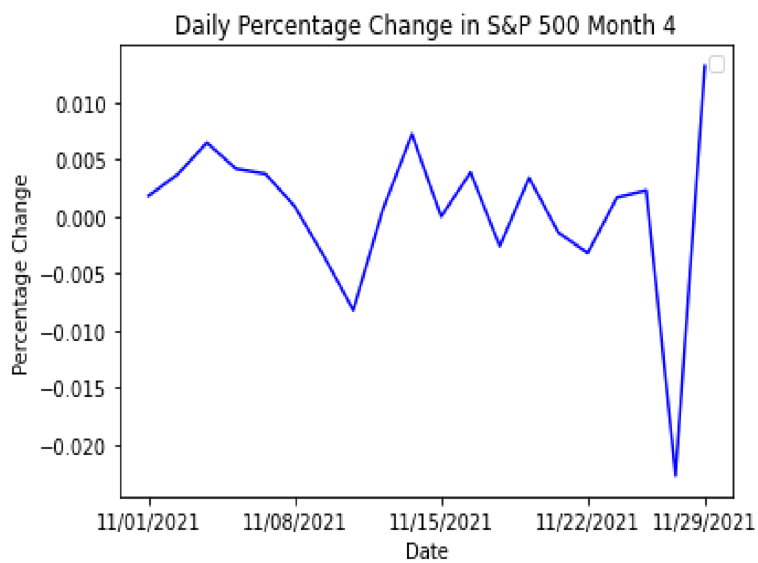
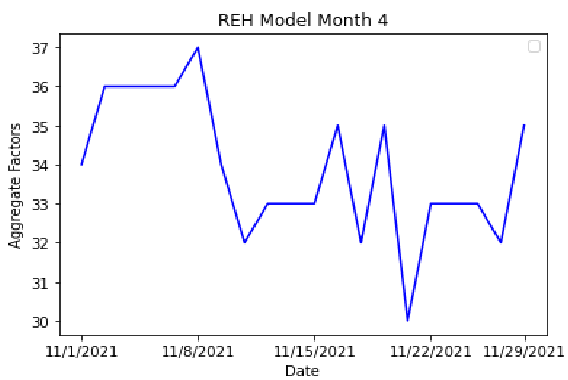


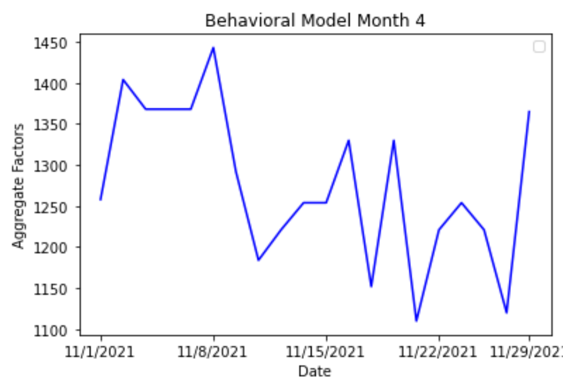
Figure 19 A: Depicts daily percentage change in the S&P 500 over the month of November.

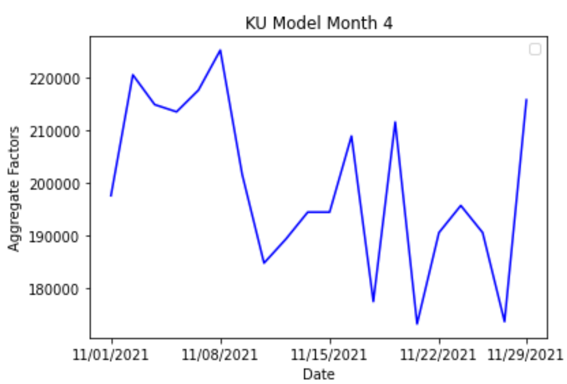
Model	Spearman Correlation Coefficient	Change in Speakman from KU Model (%)
REH Fundamentals	0.5110721964707252	-20.111846416753302
Behavioral Psychology	0.6335046525888414	-0.9738402299924925
Knightian Uncertainty Model	0.6397346459361678	-
KU Model + Google Sentiment	-0.07700048176102395	-112.03631572092578
KU Model + Google Magnitude	0.3224952606383592	-49.58921442085887
KU Model + VADER Sentiment	0.17110885405621581	-73.25315188990267
KU Model + Google Sentiment + Google	0.18456405288982275	-72.25315188990267

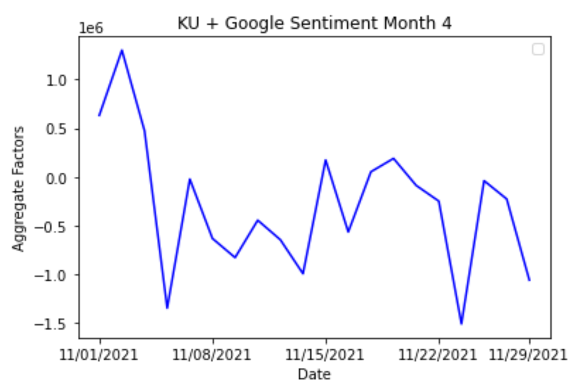
Magnitude		
KU Model + Google Sentiment + VADER	-0.09417485260600636	-114.72092424636371
KU Model + VADER Sentiment + Google Magnitude	0.29323787454718586	-54.16257718572195
KU Model + VADER Sentiment + Google Magnitude + Google Sentiment	-0.21830063078154027	-134.12362174977812

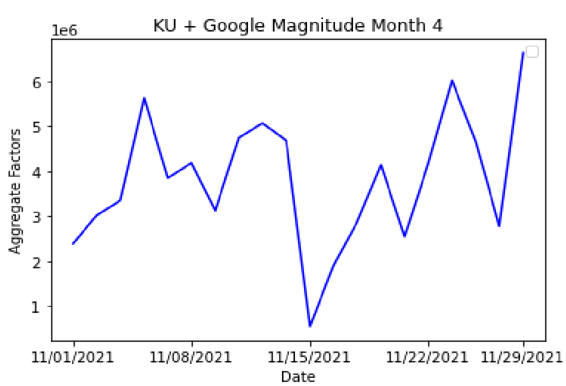
Figure 19 B: Depicts each model's SCC and their corresponding percentage changes in error as compared to the unaugmented Knightian Uncertainty model over the month of November.











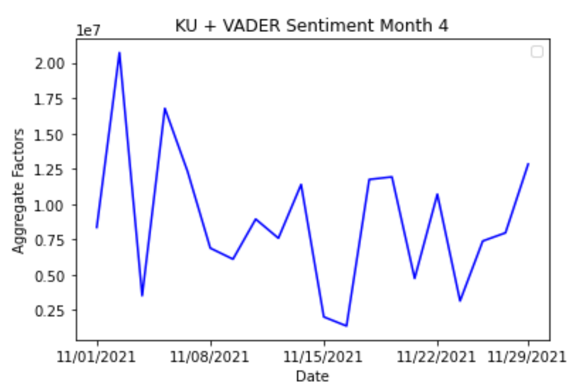


Figure 19 C: Depicts a subsection of the unaugmented models against the S&P 500 over the month of November.

It is interesting to note that results were generally mixed this month. The REH model performed substantially worse then the KU model resulting in a **20%** reduction in SCC, while the behavioral model only slightly reduced SCC with a reduction of around **.1%**. The unaugmented KU model has an SCC of **.639**, and the augmented models decreased SCC. Specifically, with the KU + Google Magnitude model resulting in a SCC of **.32**, this equates to an approximately **49%** decrease in correlation. It is interesting to consider the graphs for this month, as none of them seem to capture movement in the S&P 500 very well. It makes sense, however, that the SCC for KU would be higher, as the relative extrema are achieved around similar times, though the overall shapes look incongruous.

5.2.5: Conclusions

The exercise of reducing granularity on a monthly basis yielded unfavorable results for our hypothesis, with augmented models performing worse than the original Knightian Uncertainty model in every case. One could, however, argue that while Twitter seems to be ineffective at increasing accuracy over long periods of analysis, these results could change if the time period was restricted further. Specifically, during Knightain events, when uncertainty is high and traditional information can not offer reliable answers, influencers on Twitter may have a more direct influence on the stock market expectations.

5.3: KU Models with Twitter Data by Event

To pursue this assertion that twitter augmented models could increase accuracy on a reduced timeline we constructed localized models around Knightian Uncertainty events. For the purposes of this paper we chose four events, one from each month, that we felt gave an interesting variation in KU factors.

5.3.1: Event 1

The first event we chose to analyze was the period of 08/23/2021 - 08/31/2021. We chose to analyze this week because it has low variance and is on average close to the whole sample mean, thus serving as an indicator of “normal macroeconomic activity” across the entire period of analysis.

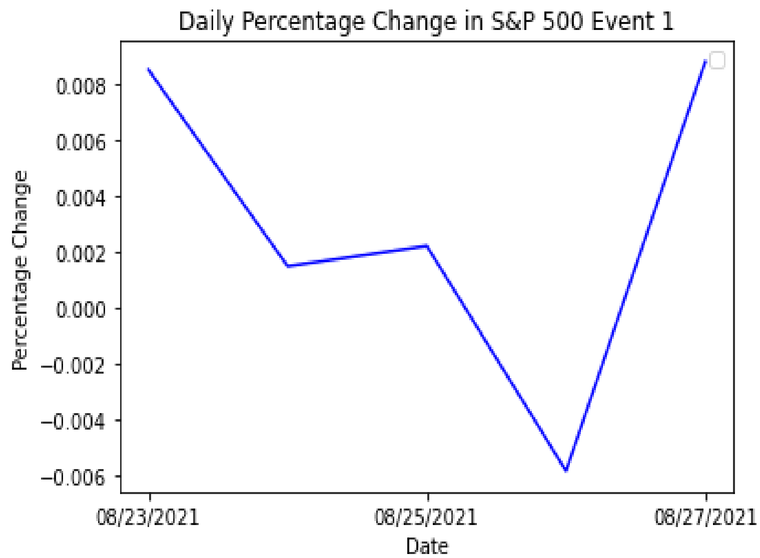


Figure 20 A: Depicts the percentage change in S&P 500 over Event 1.

Model	**Spearman Correlation Coefficient**	**Change in Speakman from KU Model (%)**
REH Fundamentals	0.5182302508803712	-17.834214364982703
Behavioral Psychology	0.5713006363814955	-9.41986589488594
Knightian Uncertainty Model	0.6307129505003024	-
KU Model + Google Sentiment	0.5459740208326083	-13.435419329899032
KU Model + Google Magnitude	-0.2785016029689617	-144.1566330210986
KU Model + VADER Sentiment	0.5392682556514548	-14.498623308164307
KU Model + Google Sentiment + Google Magnitude	0.5474534902677678	-13.200848367944632
KU Model + Google Sentiment + VADER	0.2680452864970505	-57.501223609816776
KU Model + VADER Sentiment + Google Magnitude	0.2870715742422653	-54.484591760078715
KU Model + VADER Sentiment + Google Magnitude + Google Sentiment	0.3625846784875588	-42.5119338044439

Figure 20 B: Depicts each model's SCC and their corresponding percentage changes in correlation as compared to the unaugmented Knightian Uncertainty model over Event 1.

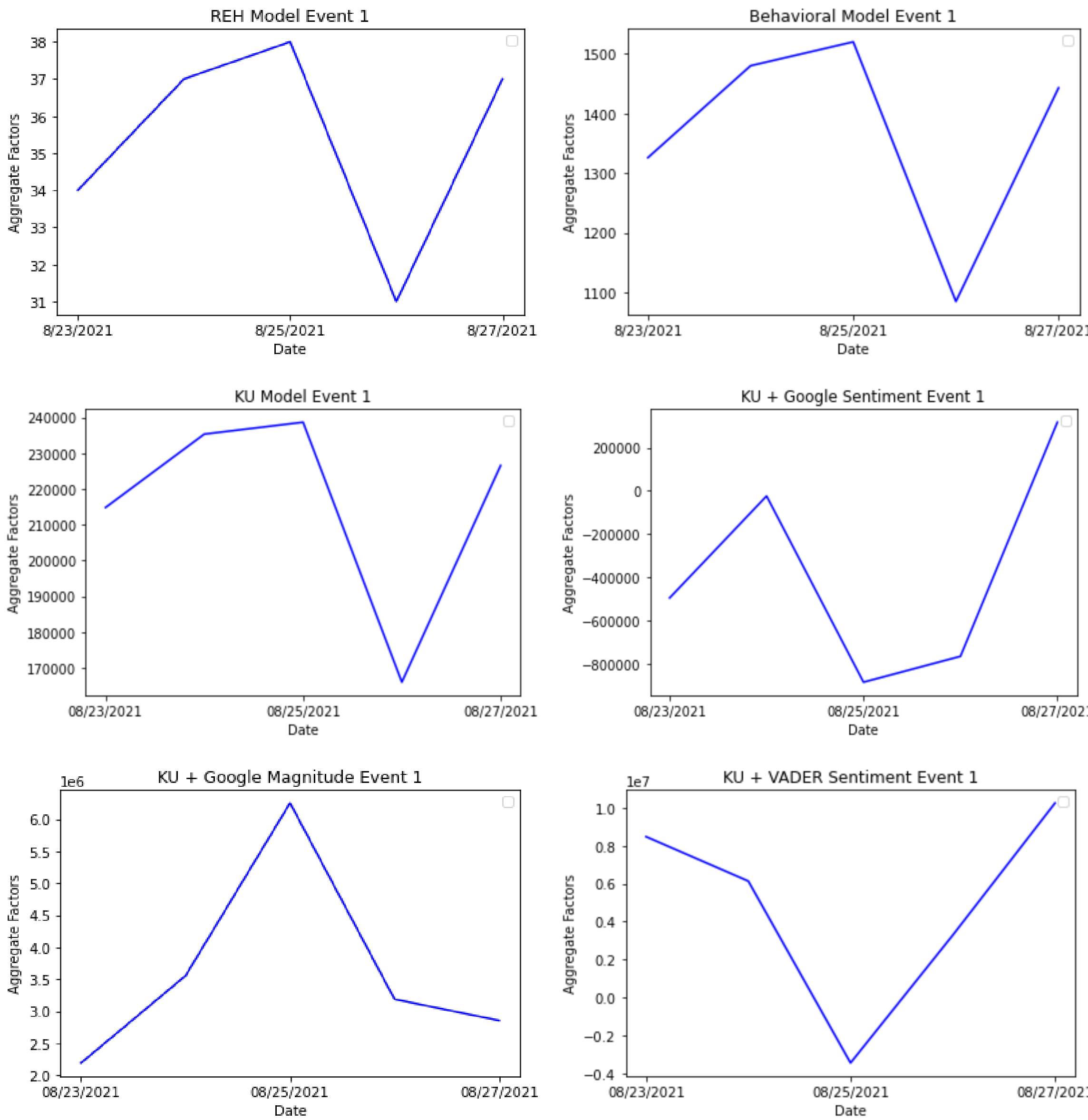


Figure 20 C: Depicts a subsection of the unaugmented models against the S&P 500 over Event 1.

As can be seen from figure 20 B, the unaugmented Knightian Uncertainty model resulted in a SCC of **.63**, which was higher than the errors produced by the augmented KU models. Specifically, the KU + Google Sentient Model had a SCC of **0.54**, resulting in a **13.43%** reduction in correlation. This is interesting as these results are rather comparatively low. Further, looking at the graphs would seem to suggest a different relationship with the graph of the Bloomberg + VADER, looking to capture the “General shape” of the S&P 500 data much more closely. Interestingly these results are not dissimilar to those obtained by the Behavioral model which obtained a **9.4%** reduction in SCC.

5.3.2: Event 2

For the month of September, we chose the near bankruptcy of the Chinese real estate company Evergrande, analyzing the period of 09/15/21 - 09/24/21, as this was a major Knightian Uncertainty Company Variable, which caused substantial repercussions for financial markets.

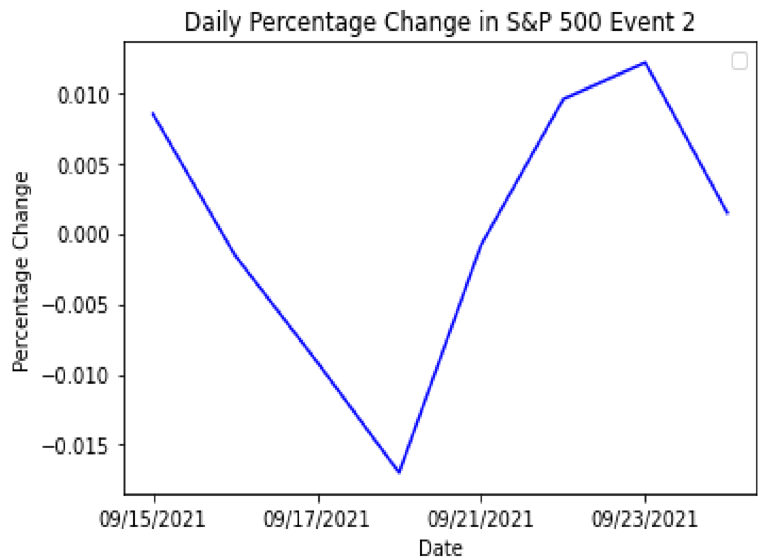


Figure 21 A: Depicts the percentage change in S&P 500 over Event 1.

Model	**Spearman Correlation Coefficient**	**Change in Speakman from KU Model (%)**
REH Fundamentals	0.904127314568554	-2.5757845753299486
Behavioral Psychology	0.9367716838820599	0.9418085967324815
Knightian Uncertainty Model	0.92803140433564969	-
KU Model + Google Sentiment	0.43463849007272026	-53.1655407303953
KU Model + Google Magnitude	0.21297732102023567	-77.05063427538856
KU Model + VADER Sentiment	0.3017637336789506	-67.48345667300188
KU Model + Google Sentiment + Google Magnitude	0.31104059537045203	-66.48382868108683
KU Model + Google Sentiment + VADER	0.19924569576452822	-78.530285200083047
KU Model + VADER Sentiment + Google Magnitude	0.33318487057135915	-64.09767287887456
KU Model + VADER Sentiment + Google Magnitude + Google Sentiment	0.14749622687098787	-84.10654788384278

Figure 21 B: Depicts each model's SCC and their corresponding percentage changes in correlation as compared to the unaugmented Knightian Uncertainty model over Event 2.

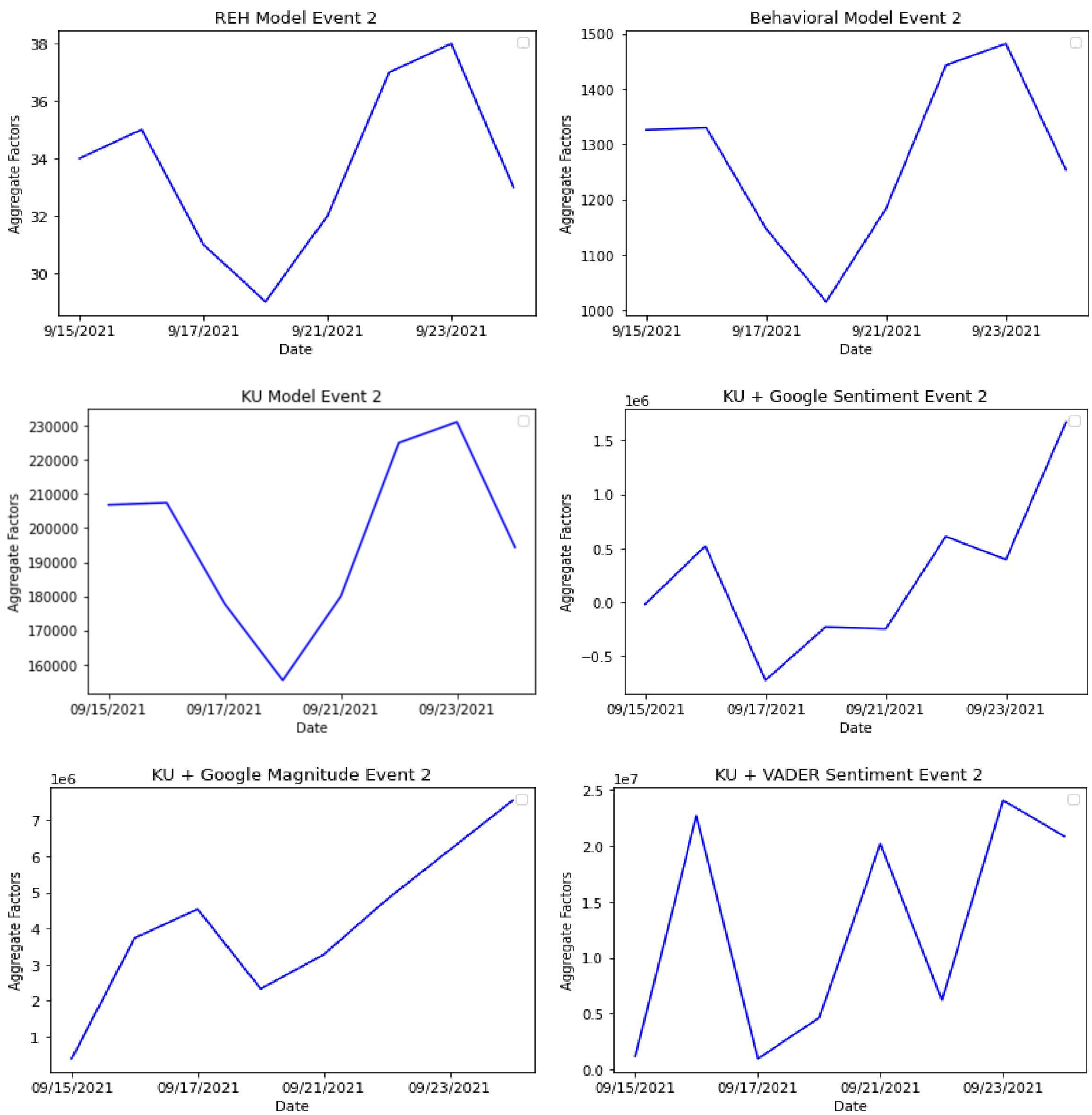


Figure 21 C: Depicts a subsection of the unaugmented models against the S&P 500 over Event 2.

As can be seen from Figure 21B, the unaugmented Knightian Uncertainty model resulted in a SCC of **0.93**, which was higher then the SCC obtained from the augmented KU models. Specifically, the KU+Google Magnitude had a SCC of **0.21**, resulting in a **77.05%** reduction in SCC. In this case, comparing the graphs would lead to a similar conclusion as the SCC analysis, as the KU graphs seem to more effectively capture the movement of S&P 500 in a qualitative sense. It is interesting to note that in this case the Behavioral model alone resulted in a slightly higher SCC then the unaugmented KU model.

It is notable that while this was a KU factor occurring primarily in China with dramatic repercussions for global financial markets, the Chinese government has officially blocked Twitter. Thus, though we intentionally added Chinese users to our dataset, Twitter may be an ill-suited platform to capture KU factors originating within the country.

5.3.3: Event 3

For the month of October, we choose the advent of the Earning Season, analyzing the period of 10/11/2021 - 10/21/2021. We chose this event as it was a technical non-momentum factor that had a clear effect on financial markets.

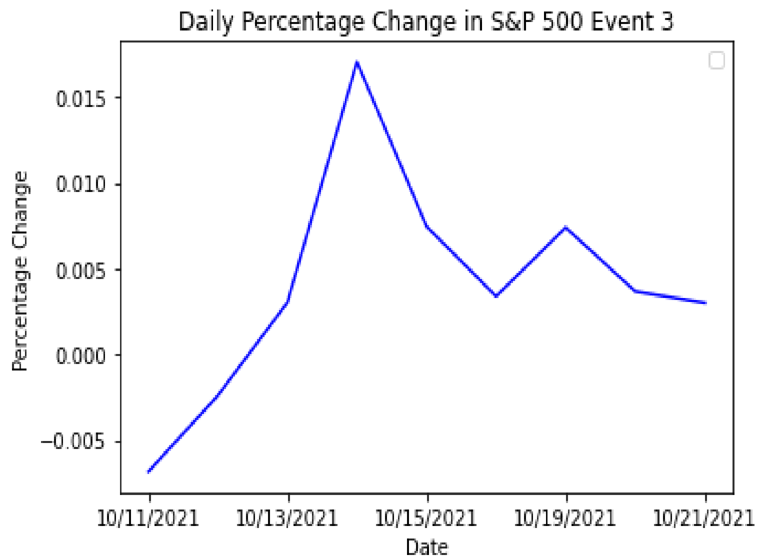


Figure 22 A: Depicts the percentage change in S&P 500 over Event 3.

Model	Spearman Correlation Coefficient	Change in Speakman from KU Model (%)
REH Fundamentals	0.541269362057162	0.20666857398920357
Behavioral Psychology	0.5078887213738058	-5.973180189353963
Knightian Uncertainty Model	0.5401530354813733	-
KU Model + Google Sentiment	-0.6193317483934955	-214.6585703885862
KU Model + Google Magnitude	0.07908002085668829	-85.35970073995533
KU Model + VADER Sentiment	-0.5931689881875611	-209.8149874616443
KU Model + Google Sentiment + Google Magnitude	-0.6594282488889528	-222.0817445376909

KU Model + Google Sentiment + VADER	-0.2623903409968338	-148.5770372026137
KU Model + VADER Sentiment + Google Magnitude	-0.6177411485616793	-214.364098317278
KU Model + VADER Sentiment + Google Magnitude + Google Sentiment	-0.23132088582880467	-142.8250644972598

Figure 22 B: Depicts each model's SCC and their corresponding percentage changes in correlation as compared to the unaugmented Knightian Uncertainty model over Event 3.

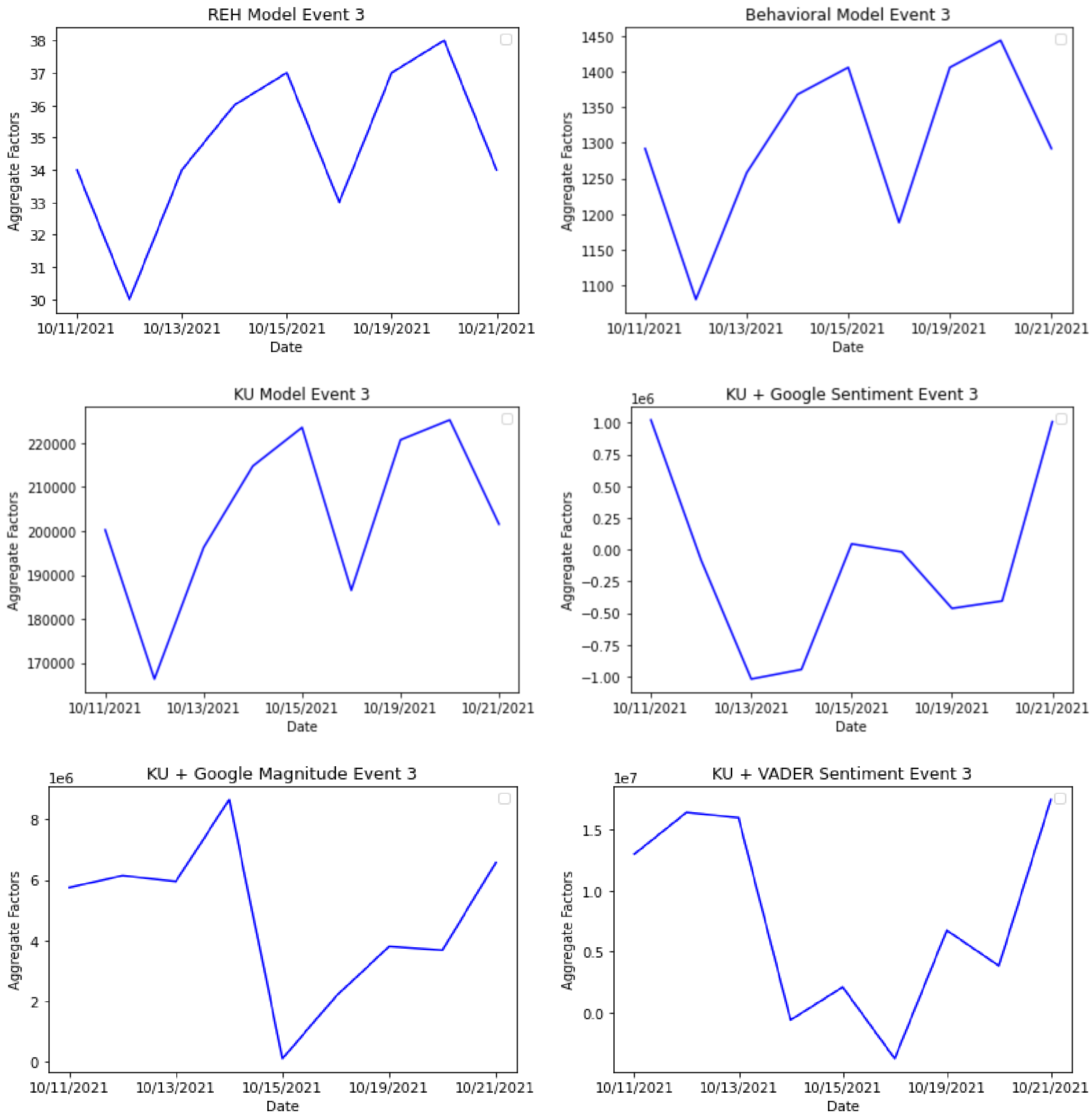


Figure 22 C: Depicts a subsection of the unaugmented models against the S&P 500 over Event 3.

As can be seen from figure 22 B, the unaugmented Knightian Uncertainty model resulted in a SCC of **0.51**, which was notably higher than the SCC produced by the augmented KU models. Specifically, the Bloomberg + Google Sentiment model had a SCC of **-0.62**, resulting in a **215%** reduction in correlation. Again this result is consistent with what one would expect looking at the graphs. This compares unfavorably to the Behavioral and REH model which both perform within **6%** of the **KU** model.

5.3.4: Event 4

We define Event 4 to be Period 4. For the month of November we chose the emergence of the Omicron variant of COVID-19, analyzing the period of 11/24/2021 - 11/29/2021. We chose this period as it was a highly disruptive Pandemic Knightian Uncertainty factor, which caused dramatic movements in financial markets.

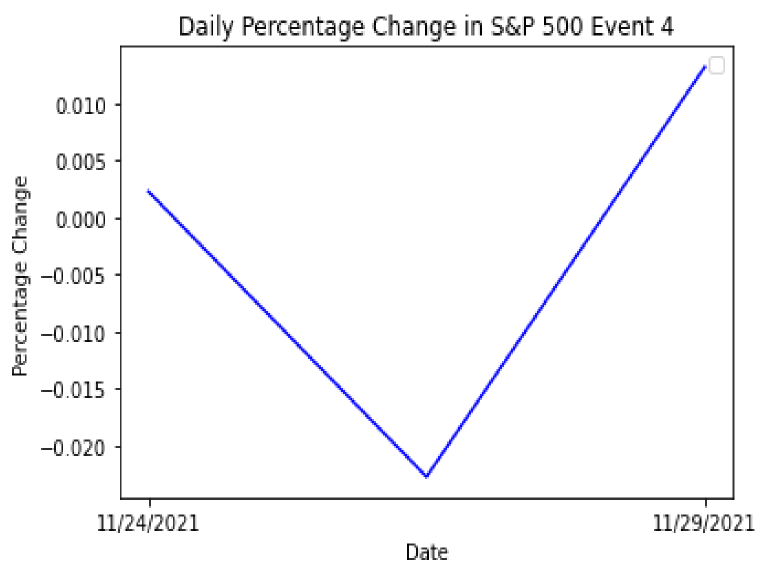
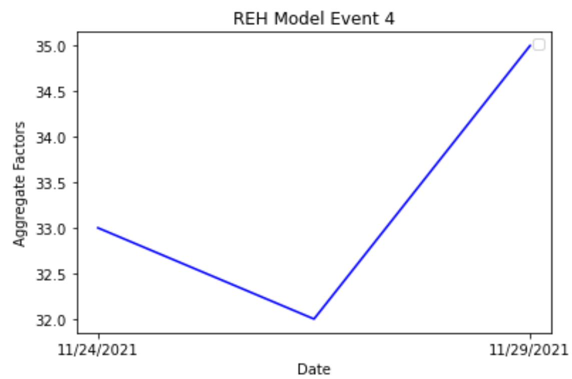


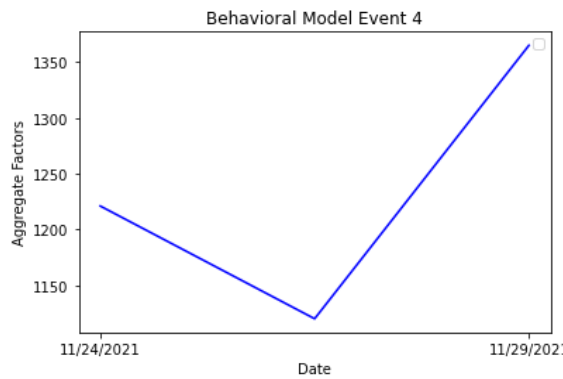
Figure 23 A: Depicts the Percentage change in S&P 500 over Event 4.

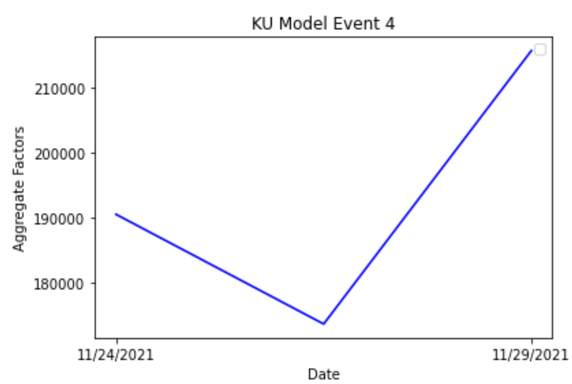
Model	Spearman Correlation Coefficient	Change in Speakman from KU Model (%)
REH Fundamentals	0.9158600229022802	-2.9659711276106413
Behavioral Psychology	0.9479656973215205	0.4355780836889686
Knightian Uncertainty Model	0.9438544740904646	-
KU Model + Google Sentiment	-0.6054252330762946	-164.143917277047
KU Model + Google Magnitude	0.9722791510994033	3.0115529235934204

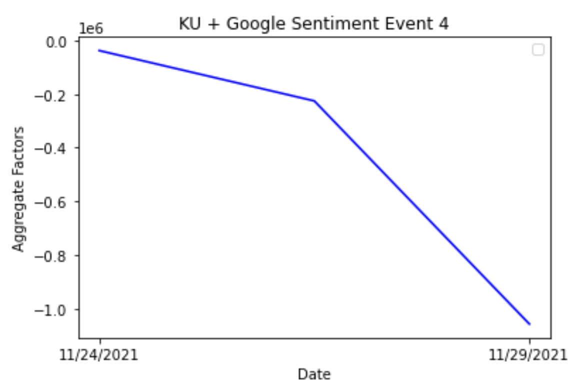
KU Model + VADER Sentiment	0.6649102681523181	-29.553730325530125
KU Model + Google Sentiment + Google Magnitude	-0.6799784159854569	-172.04271788198184
KU Model + Google Sentiment + VADER	-0.6365077041694457	-167.4370595936211
KU Model + VADER Sentiment + Google Magnitude	0.8491616803655834	-10.032562892296609
KU Model + VADER Sentiment + Google Magnitude + Google Sentiment	-0.691194475220867	-173.2310429409077

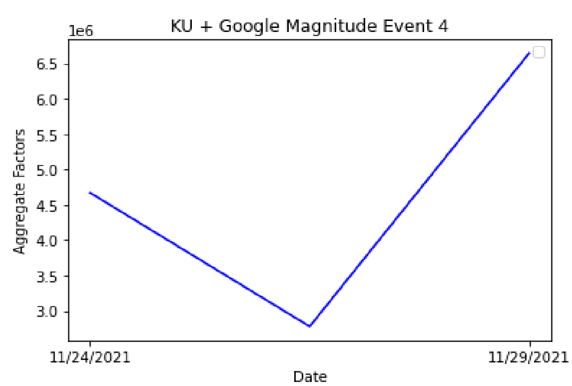
Figure 23 B: Depicts each model's SCC and their corresponding percentage changes in correlation as compared to the unaugmented Knightian Uncertainty model over Event 4.











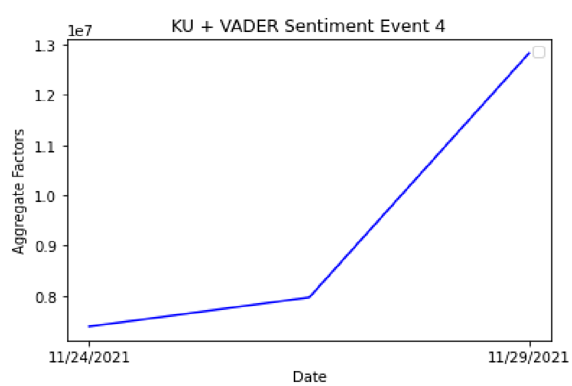


Figure 23 C: Depicts a subsection of the unaugmented models against the S&P 500 over Event 4.

As can be seen from figure 23 B, the unaugmented Knightian Uncertainty model resulted in a SCC of **0.54**, which was notably higher than the errors produced by the augmented KU models. Specifically, the Bloomberg + Google Sentiment results in a SCC of **-0.61** resulting in a **215%** reduction in SCC. In this case both the REH and Behavioral models performed much more similarly to the KU model, resulting in only small changes in SCC.

5.3.5: Conclusion

In testing our hypothesis - that reducing granularity by restricting analysis to local KU events increases model predictive accuracy, we found generally discouraging results. Namely, we found that using augmented KU models led to a large reduction in SCC in all events. It is worth noting in the case of events that SCC is a rank ordering system and so events with over a short time that is with few observations may have similar ranks, for spurious reasons.

6: Analysis and Conclusions

6.1: Limitations

As in any undertaking there were certain technical and time limitations that we faced in the process of writing this paper. Among these limitations were the number and/or subset of Twitter users chosen, terms used in tweet selection, Stock Market index chosen, choice of analysis periods and sentiment analysis algorithm used. Further, it is worth noting that the SCC is one of a number of measures of rank correlation, and not necessarily the best suited for this task. Additionally, the scoring of Bloomberg Market Wraps is highly subjective, only using three discrete values to represent highly complex market factors. Furthermore, the manner in which we weighted the Twitter Sentiment Analysis data in the augmented KU models might have been done in an overly simplistic manner. Refining the formula should lead to more accurate results, allowing for better interpretation. All of these limitations could potentially have been refined and expanded to produce more accurate results.

6.2: Conclusion

In attempting to model movement in the S&P 500, a Knightian Uncertainty approach to Bloomberg Market Wraps produced on average a higher SCC than the REH or Behavioral Models over our entire period of analysis. Despite this, attempts to augment this Knightian Uncertainty model with sentiment gathered from influential users on Twitter proved unsuccessful on multiple levels of analysis including the overall period, month by month and by event. Specifically we found that in every case analyzed Twitter data decreased our SCC. This leads one to a surprising potential conclusion, that models are made notably worse by the inclusion of sentiment gathered from Twitter. These results are in direct opposition to the common claim that "retail investors" are having their forecasts influenced by online influencers as opposed to traditional media outlets. This is a question that could certainly be explored further, and a potential further research question could be whether further augmenting Bloomberg data with other social media platforms would lead to different results.

Bibliography

- https://twitter.com/elonmusk/status/1256239815256797184?lang=en, 14th January, 2022
- https://developer.twitter.com/en, 14th January, 2022
- https://mobile.twitter.com/tesla?lang=en, 14th January, 2022
- https://twitter.com/elonmusknewsorg?lang=en, 14th January, 2022
- https://developer.twitter.com/en/docs/twitter-api/fields, 14th January, 2022
- *Sentiment analysis algorithms and applications: A survey*, Medhat, Hassan, Korashy, 2013; https://www.sciencedirect.com/science/article/pii/S2090447914000550, 14th January, 2022
- https://github.com/cjhutto/vaderSentiment, 14th January, 2022
- https://cloud.Google.com/natural-language/docs/analyzing-sentiment, 14th January, 2022
- https://finance.yahoo.com/quote/%5EGSPC/, 14th January, 2022
- *Lecture Notes*, Risk & Fluctuations in Financial Markets, Frydman, 2021
- *Expectations Concordance and Stock Market Volatility: Knightian Uncertainty in the Year of the Pandemic*, Frydman, Mangee, October 2021
- *Bloomberg Market Wraps*, Bloomberg Terminal, 14th January, 2022

Python Packages

- Van Rossum, G. (2020). The Python Library Reference, release 3.8.2. Python Software Foundation.
- McKinney, W., & others. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56).
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2
- J. D. Hunter, "Matplotlib: A 2D Graphics Environment", Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007.
- Json Python Package
- Yfinance Python Package
- Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.
- Google.Cloud Python Package

Appendix

7: Bloomberg Scoring Factors

7.1: Fundamental Factors

Benchmark Valuation,Central Bank Communication, Company Variables, Currency Markets, Dividends or Earnings, Financial Institutions, Geopolitical Issues, Government or Fiscal, Housing, Inflation Rates, Interest Rates, International Trade, Macroeconomic Activity, Supply Chain, Oil/Energy, Rest of World, Sales

7.2: Behavioral Factors

Psych w/ Funds, Psych w/ Benchmark Valuation, Psych w/ Central Bank, Psych w/ Company Variables, Psych w/ Currency Markets, Psych w/ Dividends or Earnings, Psych w/ Financial Institutions, Psych w/ Geopolitical Issues, Psych w/ Government or Fiscal, Psych w/ Housing, Psych w/ Inflation Rates, Psych w/ Interest Rates, Psych w/ International Trade, Psych w/ Macroeconomic Activity, Psych w/ Supply Chain, Psych w/ Oil/Energy, Psych w/ Rest of World, Psych w/ Covid, Pure Psychology

7.3: Technical Factors

Momentum, Non-momentum

7.4: Knightian Uncertainty Factors

Bankruptcy, Executive turnover/health, Legal or accounting issues, Firm added to index, Executive salary issues, Assets, Reorganization/spinoffs/partnerships, Liabilities/Debt, Share issuances, IPOs, Business spending, Mergers and acquisitions, Accident/recall/data breach, Price target announcements, New products/production processes, Insider trading, Credit ratings, Resource discovery, exploration/drilling, Labor layoff/strike/union, CEO/CFO comments, Purchase/sale of large stake, Stock split, Share buyback, Exchange rates, FX Intervention, Gap from benchmark levels, Overvalued, Undervalued, Trade agreements, Trade negotiations, Tariffs, Quotas, Subsidies, Foreclosures, Home prices/Sales, Real estate prices, Commercial prices, Fiscal policy, Comments by officials, Stimulus plan, Industry regulation, Taxes and rules on CEO bonuses, Government shutdown, Impeachment, Military spending, Austerity measures, Credit worthiness/ratings, Bailout or nationalization of banks, Healthcare issues, Political elections, Political conflicts, instability or corruption, Congressional testimony, Financial reform, Cabinet changes, Leverage or credit issues, Liquidity Issues, Credit card defaults, Credit ratings, Capital funding, Monetary policy,

Comments by officials/books, Macroprudential policy, Commodity prices,
Armed conflicts/embargo, Border control/policy/immigration, Travel bans/transportation,
Nuclear testing, Terrorism, Natural disasters, COVID cases/deaths, COVID vaccination,
Climate change/pollution, Crime/assasination/shooting, Protests, Civil unrest,
Analyst comments w/ funds, Analyst price targets, Analyst Ratings

8: Twitter Users and Terms

8.1: Twitter Users

User	Region	Occupation
Elon Musk	North America	Tesla CEO
Bernie Sanders	North America	Liberal Politician
Barack Obama	North America	Liberal Politician
Alexandria Ocasio-Cortez	North America	Liberal Politician
Joe Biden	North America	Current President
Narendra Modi	Asia	Current Prime Minister
Howie Hawkins	North America	Green Party Candidate
Micheal Bloomberg	North America	Founder of Bloomberg
Anderson Cooper	North America	Journalist for CNN
Cagil M. Kasapoglu	Middle East	BBC Reporter focused on Turkey
Racheal Maddow	North America	Reporter for MSNBC
Jack Taper	North America	Reporter for CNN
Tucker Carlson	North America	Reporter for Fox
Ben Shapiro	North America	Conservative Political Commentator
Ezra Klien	North America	Founder of VOX
Qiao Collective	Asia	Chinese Focused Media Collective
Hua Chunying	Asia	Chinese Assistant Minister of Foreign Affairs,

Jack Dorsey	North America	Twitter CEO
Nathan Law	Asia	Exiled Hong Kong Politician
Tsai Ing-wen	Asia	Prime Minister Of Taiwan
Phoebe Bridges	North America	American Musician
Paul Kerry	Asia	Copy Editor of the Korean Herald
David M. Friedman	Middle East	Former US Ambassador to Israel
Arsen Ostrovsky	Middle East	Israli Human Rights Lawyer
US Consulate Munich	Europe	US Consulate in Munich
Tom Morello	North America	American Musician
NCT K	Aisa	Korean Pop Group
Keith Bradsher	Asia	Writer for the Economist focused on Asia
Tom Nuttall	Europe	Economist Bureau Chief for Berlin
Gregg Carlstrom	Middle East	Economist Correspondent in the Middle East
Joseph Stinziano	Asia	Samsung Executive
Greg Thome	Asia	Toyota Executive
Megan Baldino	Europe	BP Executive
Robin Tickle	Europe	Nestle Executive
Rob Shwerwin	Europe	Shell Executive

Figure 24: Depicts all Twitter users included in our sample.

8.2: Twitter Terms Used for Scraping

all the companies on the S&P 500, and their tickers, interest rate, central bank, the fed, fed chairman, inflation, prices, deflation, conflict, war, trade war, boom, bust, crash, election, impeach, coup d'etat, insurrection, revolution, corruption, treaty, agreement, bubble, over valued, legalization, under valued, bull, bear, ipo, quantitative easing, taper, monopoly, oligopoly, duopoly, anti-competitive, regulation, breaking up, legislative change, bond, yield curve inversion, devalued, market, long, short, crypto, exchange rate, bitcoin, btc, etherium,

eth, dogecoin, doge, covid, delta, omicron, border close, trade, jobs, supply chain, higher, fire, unemployment, benchmark, oil prices, globalization, depression, earning ratio,
price to book ratio, price to earnings, growth ratio, yield, dividend, alpha, beta, r squared, standard deviation, sharpe ratio, 10k, 10q, revenue, net income, earnings per share, earnings, return on equity, market distributors, leveraging ratio, debt to equity, value at risk, run on banks, panic, uncertainty, money supply, reserve requirement, speculation, speculative,
speculative attack, federal funds rate, announcement, jerome powell,
international monetary fund, imo, nato, the un, the eu, trade bloc, communism, capitalism, socialism, neoliberalism, anarchy, democracy, republicanism, dictatorship, dictator, authoritarian, monarchy, oligarchy, federalism, theocracy, leader, aristocracy, bourgeoisie, proletariat, workers, owners, rich, poor