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

Siddhant Kumar
Buz Galbraith

January 10th 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 posit that Knightian Uncertainty Models are the most accurate of these models at predicting movements in financial markets. With the understanding that users on Twitter distill information in a unique way that influences the price expectations of market participants, we propose that Twitter can be viewed as a Knightian Uncertainty factor. As such, we hypothesize that a Knightian Uncertainty model augmented with Twitter data will be able to model movements in financial markets more accurately than the original Knightian Uncertainty Model.

2: Data Collection

2.1: Bloomberg Data Collection

The first task of this paper was to collect data which would support qualitative comparisons of financial markets 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, 0, and 1) depending on the relevant factor's effect on that day's financial market movement.

[1] https://twitter.com/elonmusk/status/1256239815256797184?lang=en, Jan 10th, 2022

1									Market Fundamentals																
2	Date	Dow	SP500	NASDAQ	Bonds	Crypto	Exchange	Oil	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
45	10/01/2021	1	1	1	-1	1	1	1			1							1			1				
46	10/04/2021	-1	-1	-1	1		1	1								-1		-1			-1	-1	1		
47	10/05/2021	1	1	1	1	1	-1	1		0											1		1		
48	10/06/2021	1	1	1		1	-1	-1								1		0			1		-1		
49	10/07/2021	1	1	1	1			1								1					0		1	0	
50	10/08/2021		-1	-1	1			1													-1		1		
51	10/11/2021	-1	-1	-1		1	-1	1						0				0			0	-1	1		
52	10/12/2021	-1	-1	-1	-1		-1				-1						-1	-1			0	0		-1	
53	10/13/2021		1	1	-1		1		1	0								-1				0			
54	10/14/2021	1	1	1	-1		-1	1			1										1				
55	10/15/2021	1	1	1	1	1		1			1										1		1		
56	10/18/2021	-1	1	1	1	1	-1			0								0					-1	0	
57	10/19/2021	1	1	1	1	1	1	1		0	1								0		1		1		
58	10/20/2021	1	1	-1	1	1	1	1					1			1					1	0	1		
59	10/21/2021		1	1	1	-1	-1	-1													1		-1		

Figure 1: A crosssection 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 +1 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.

1		Psychology w/ Fundamentals										Pure Psychology			Technical Trading			
2	Date	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	Total Psychology	Trading Days	Momentum	Non-momentum	KU Factors	Title
3	08/02/2021						-1						1	1	1	0	1	Stocks Fall as Yields Dip on Growth Concerns: Markets Wrap
4	08/03/2021												0					Stocks Rise as Earnings Offset Virus, Oil Concerns: Markets Wrap
5	08/04/2021						0 (= -1 + 1)						1	1				U.S. Stocks Slip as Economic Growth Worries Mount: Markets Wrap
6	08/05/2021												0					Stocks Gain as Earnings Outweigh Virus Concerns: Markets Wrap
7	08/06/2021												0				1	Stocks, Treasury Yields Rise on Robust Jobs Report: Markets Wrap
8	08/09/2021						-1						2	1			1	U.S. Stocks Decline Amid Stimulus, Virus Concern: Markets Wrap
9	08/10/2021						-1		1		-1		4	1			1	Tech Drop Exposes Doubts in U.S. Growth Optimism: Markets Wrap
10	08/11/2021			1									2	1				Stocks, Bonds Rise as Taper Debate Takes Spotlight: Markets Wrap
11	08/12/2021												0					Stocks Set Another Record as Trading Range Narrows: Markets Wrap
12	08/13/2021								-1		-1	-1	3	1	1			Stocks Grind to Another Record High; Yields Fall: Markets Wrap
13	08/16/2021												0					S&P 500 Extends Gain, Doubles From Pandemic Bottom: Markets Wra
14	08/17/2021						-1	-1		-1	-1		4	1	0		1	Stocks Suffer Worst Drop in a Month; Dollar Climbs: Markets Wrap
15	08/18/2021												0		-1		1	Stocks Slump, Bonds Waver With Fed Taper in Focus: Markets Wrap
16	08/19/2021							-1			-1		3	1			1	Stocks in Choppy Trading Ahead of Big Options Day: Markets Wrap
17	08/20/2021												1	1				Stock Traders Buy the Dip at End of Jittery Week: Markets Wrap
18	08/23/2021											1	1	1			1	U.S. Stocks Rise Amid Recovery Bets, FDA Approval: Markets Wrap
19	08/24/2021						1			1			2	1	1		1	U.S. Stocks Top Record as Strong Earnings Continue: Markets Wrap
20	08/25/2021										1	1	2	1	1			Stocks Extend Gain as 'Not That Many Bears Left': Markets Wrap
21	08/26/2021									-1	-1	-1	3	1			1	U.S. Stocks Fall on Kabul Blasts, Hawkish Fed View: Markets Wrap
22	08/27/2021												1	1			1	U.S. Stocks Rise on Powell's Dovish Taper Tone: Markets Wrap

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 10th, 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 10th, 2022
[4] https://twitter.com/elonmusknewsorg?lang=en, Jan 10th, 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 10th, 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.

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 the authors 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 three 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 10th, 2022

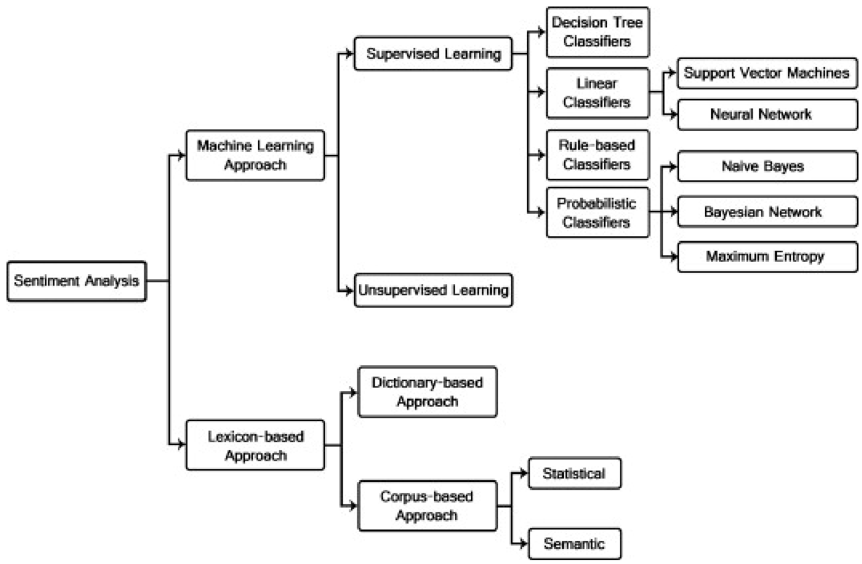


Figure 5: A broad overview of sentiment analysis approaches.

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.

---

[7] https://github.com/cjhutto/vaderSentiment, Jan 10th, 2022
[8] https://cloud.Google.com/natural-language/docs/analyzing-sentiment, Jan 10th, 2022

2.2.7: Twitter Data Examples

Figure 6: Here is 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: Here is 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 Sei	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	@wholemarsb	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	@wholemarsb	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: Here are 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: Macroeconomic Indicator Data Collection

To establish a point of comparison, we needed to collect data on macroeconomic indicator variables over the period of analysis as a measurement of overall movement in financial markets.

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 a number of variables indicating the variation in financial markets. These variables include the S&P 500, Dow Jones Industrial Average, NASDAQ, 10-Year US Treasury Bonds, and Crude Oil. We chose the S&P 500, Dow Jones, NASDAQ as they are large stock indices and capture price movements for high value companies. We chose Treasury Bonds and Crude Oil because they are determining factors in the fixed income and commodities markets. We assume this data to be representative of the true price movement in financial markets. We define the whole period of analysis as Period 0.

[9] https://finance.yahoo.com/quote/%5EGSPC/, Jan 10th, 2022

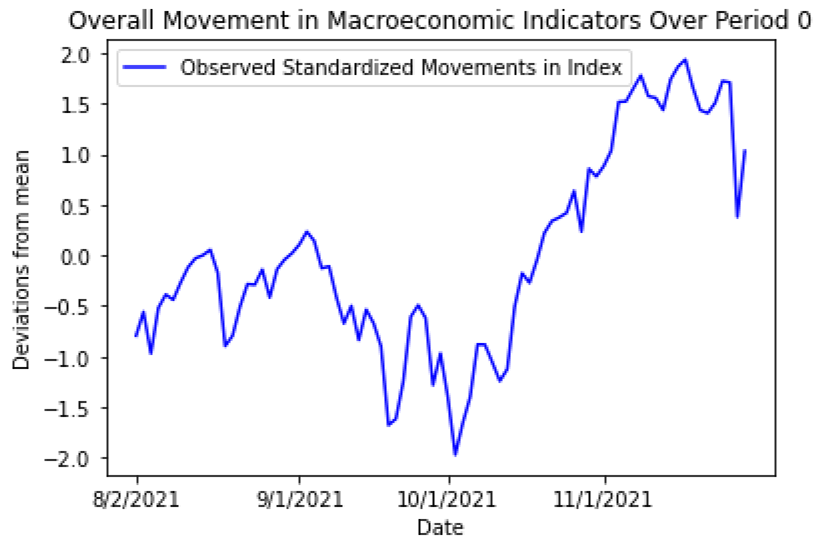


Figure 10: Depicts the standardized sum of movements in our macroeconomic indicator variables over the whole period of analysis.

3: Methodology

3.1: Choosing a Metric of Comparison

In order to compare how well our models fit our macroeconomic indicator variable data, we need an error metric. Thus, our metric of choice was the Mean Squared Error: $MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$. We chose this metric as it is widely used and easily interpretable. Naturally, this comes with the assumption that our Macroeconomic Indicators do represent true price movements in financial markets. Additionally, we do not claim that this quantitative measure offers anything other than a means of comparing the performance of our models.

3.2: Standardization

One key methodological point that came up time and again during analysis was standardization. Naturally, in order to have any hope of comparing or aggregating metrics with different units and scales, standardization is necessary. We used the typical approach to standardization, i.e., $z = \frac{x-\mu}{\sigma}$, in order to understand our data in terms of deviations from the mean. To avoid unnecessary assumptions, however, we viewed all our data as weighed equally. As such, any model that is a sum of multiple variables is in fact a standardized sum of their standardized values. We believe this approach increases comparability, interpretability and further affords consistency in the magnitude of root mean squared errors.

This approach has a key implication when reducing the granularity of our data that must be discussed. Firstly, by assumption, our standardized macroeconomic indicators represent the actual movement in financial markets which we are trying to model. Thus, when reducing granularity, we simply take a local section of this data, which remains standardized over the whole time period of analysis. However, when fitting models to this data, we allow for local standardization, enabling localized KU and Twitter Augmented KU models to be formed. This is theoretically consistent with a Knightian based approach, which posits a reevaluation of model parameters following a KU event.

3.3: Models

It is important to formalize how the term "model" will be used throughout this paper. In the spirit of econometrics, we take a model to mean a parametric function that is trying to fit our data. Thus, while the REH model is a single type of model, there are a potentially infinite number of REH model instances depending on the value of their parameters.

During this paper we constructed a number of model types. 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.

3.4: General Coding Methodology

The coding section of this paper was done by Galbraith using a Jupyter-Hub Notebook and a number of dependences and APIs. All code and the data are openly available on (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

We define the whole period of analysis as Period 0.

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.

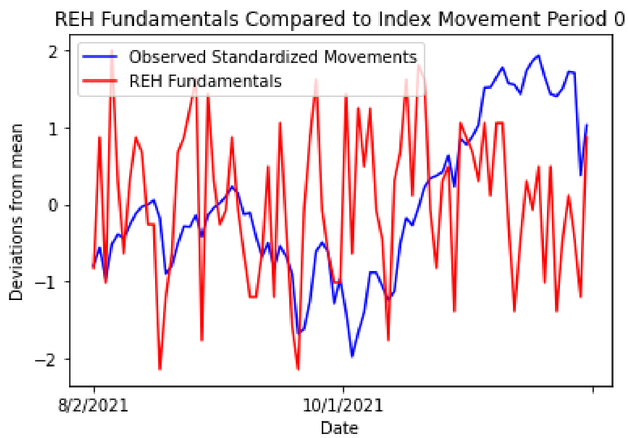


Figure 11: Depicts the REH model plotted against macroeconomic indicator variables over the period of analysis.

Our REH approach seeks to model change in macroeconomic indicators variables solely through fundamental factors present in Bloomberg Market Wraps. 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. Examples of which include psychology around earnings or dividends, and psychology around macroeconomic activity.

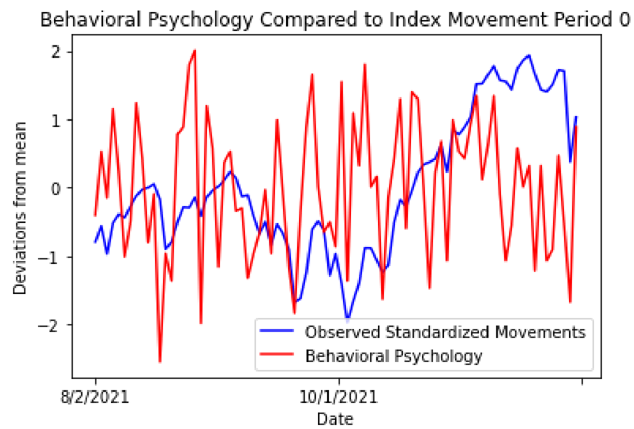


Figure 12: Depicts the Behavioral model plotted against macroeconomic indicator variables over the period of analysis.

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.

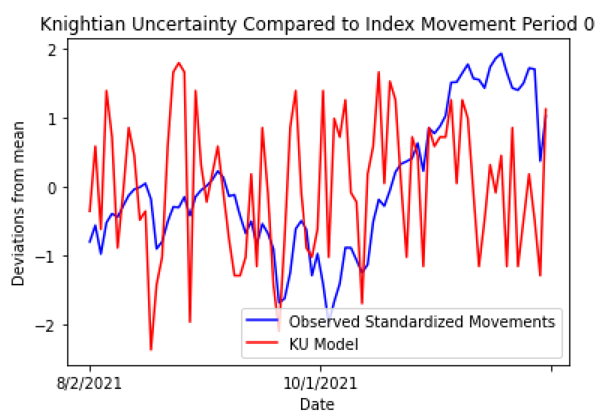


Figure 13: Depicts the KU model plotted against macroeconomic indicator variables over the period of analysis.

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 mean squared error metric to compare how well each model fits the macroeconomic indicator data. Based on figure 14 below, the REH and Behavioral models perform similarly over the entire period of analysis, resulting in mean squared errors near **1.78**. In contrast, the KU model performs slightly better, reporting a mean squared error below **1.70**. 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, such as 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.

Name	Mean Squared Error
REH Fundamentals	1.7784790779565387
Behavioral Psychology	1.7889111415366592
Knightian Uncertainty Model	1.6983281939161667

Figure 14: Depicts the Mean Squared Errors 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 reduces the mean squared error, thus increasing predictive accuracy.

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

Name	Mean Squared Error	Change in MSE from KU Model (%)
REH Fundamentals	1.7784790779565387	4.719399014130032
Behavioral Psychology	1.7889111415366592	5.333653880621135

Knightian Uncertainty Model	1.6983281939161667	-
KU Model + Google Sentiment	1.9234593658100176	13.25604631073821
KU Model + Google Magnitude	1.8322231873917507	7.8839292638036085
KU Model + VADER Sentiment	1.9453757078204936	14.546511963312653
Google Sentiment + Magnitude	1.934292663910785	13.893926441302776
Google Sentiment + VADER	1.9541034270700646	15.06041259104973
Google Magnitude + VADER	1.901475269235555	11.961591172254662
Whole Twitter + KU Model	1.9470959098114438	14.647799923855986

Figure 15 A: Depicts the MSE and each model's corresponding percentage change in error as compared to the unaugmented Knightian Uncertainty model over the whole period of analysis.

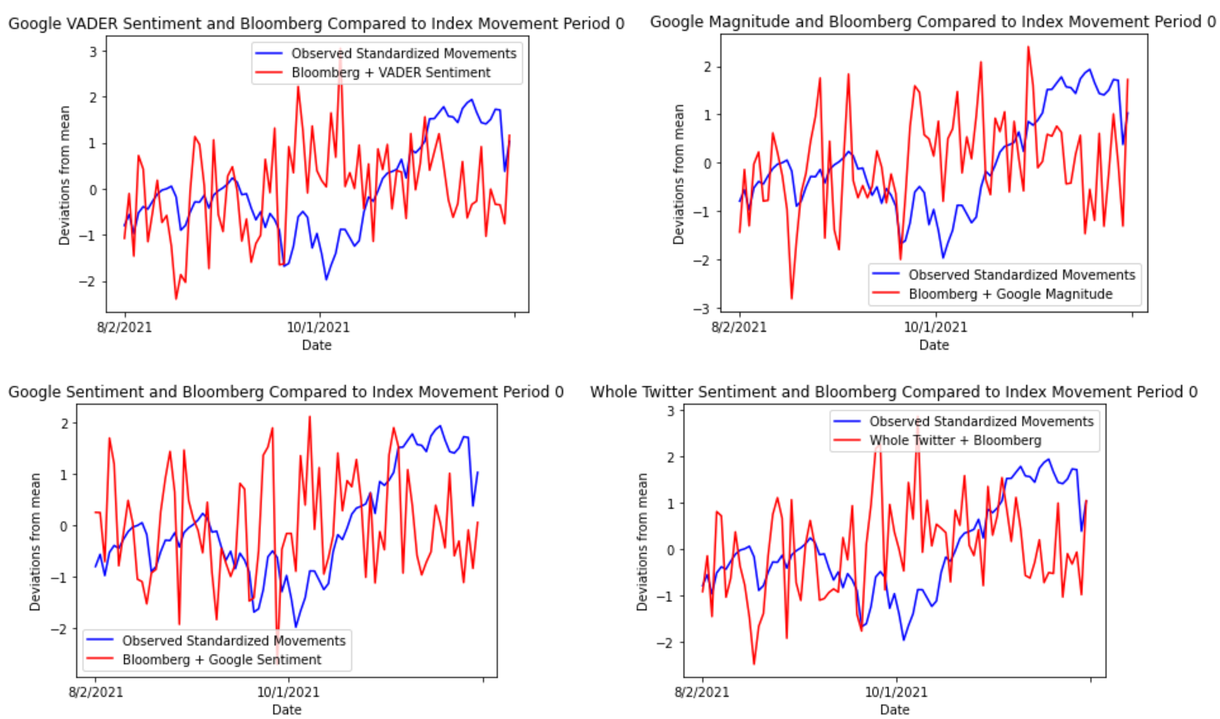


Figure 15 B: Depicts a subsection of the unaugmented models against the macroeconomic indicator variables over the whole period of analysis.

As can be seen from the root mean squared errors in figure 15, 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. 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 changes to the parameters of our model instances. 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 Period 5. Notice first from figure 16 A that overall the month of August is volatile, with a predominantly building trend in financial markets 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.

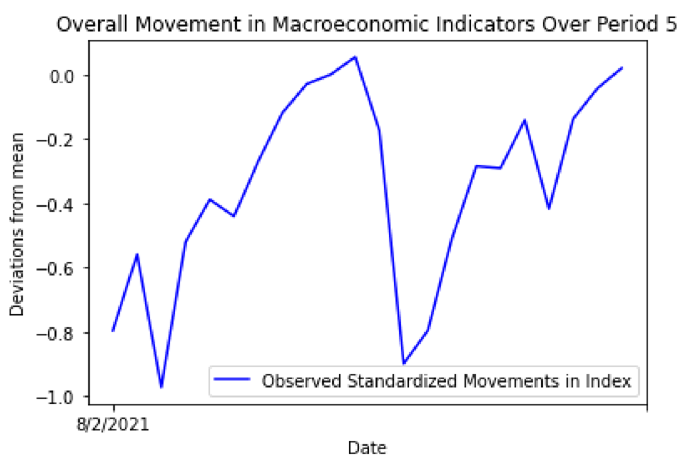


Figure 16 A: Depicts the standardized sum of movements in our macroeconomic indicator variables over the month of August.

Name	Mean Squared Error	Change in MSE from KU Model (%)
Knightian Uncertainty Model	1.411686391917659	-
KU Model + Google Sentiment	1.1400876148618093	- 19.239313958881848
KU Model + Google Magnitude	1.0859187189421375	- 23.07649027720619

KU Model + VADER Sentiment	1.0417494009084445	- 26.20532386847519
Google Sentiment + Magnitude + KU Model	1.1736953527502334	- 16.858633796429423
Google Sentiment + VADER + KU Model	1.1736953527502334	- 21.094529279548357
Google Magnitude + VADER + KU Model	1.0653048755082968	- 24.536718522789727
Whole Twitter + KU Model	1.117375908847874	- 20.848149047465743

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

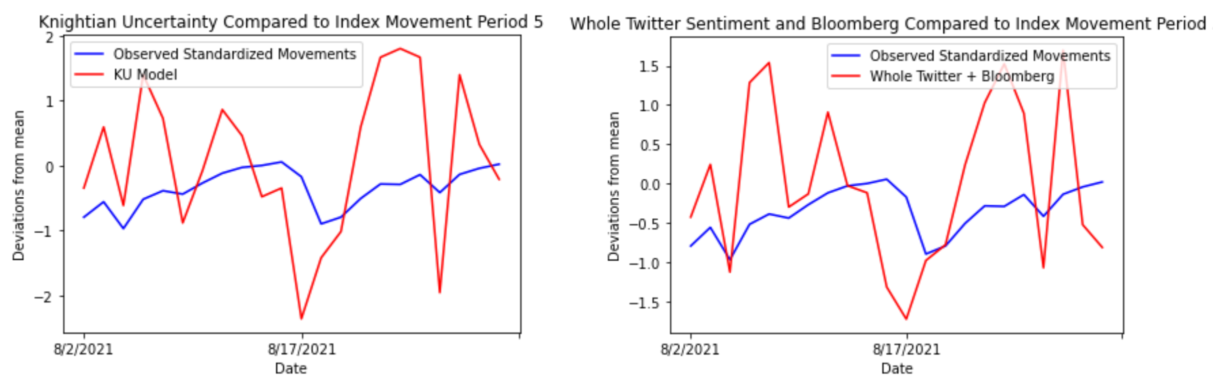


Figure 16 C: Depicts a subsection of the unaugmented models against the macroeconomic indicator variables over the month of August.

As can be seen by looking at figure 16 B, the unaugmented Knightian model has a MSE of **1.41**. Contrasting this result, the KU models augmented with Twitter data perform significantly better. Specifically, the Whole Twitter + KU Model results in a MSE of **1.12**, which equates to an approximately **21%** reduction of error.

5.2.2: September

We define September as Period 6. Looking at figure 17 A, one can see that the first half of the month experienced generally falling movement in financial markets followed by an unstable recovery in the later half of the month.

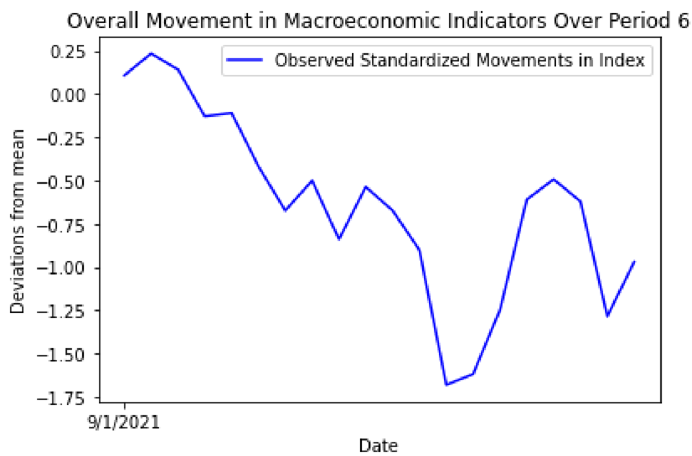


Figure 17 A: Depicts the standardized sum of movements in our macroeconomic indicator variables over the month of September.

Name	**Mean Squared Error**	**Change in MSE from KU Model (%)**
Knightian Uncertainty Model	0.7900157824761473	-
KU Model + Google Sentiment	1.6477803474502188	108.57562393064846
KU Model + Google Magnitude	1.3589370810950767	72.01391557466849
KU Model + VADER Sentiment	1.7557658660997402	122.24440385186246
Google Sentiment + Magnitude + KU Model	1.5463574168650158	95.73753476395953
Google Sentiment + VADER + KU Model	1.6884031716132235	113.71765084505776
Google Magnitude + VADER + KU Model	1.582764950962097	100.34599131693736
Whole Twitter + KU Model	1.6320459425919598	106.58396690210876

Figure 17 B: Depicts the MSE and each model’s corresponding percentage change in error as compared to the unaugmented Knightian Uncertainty model over the month of September.

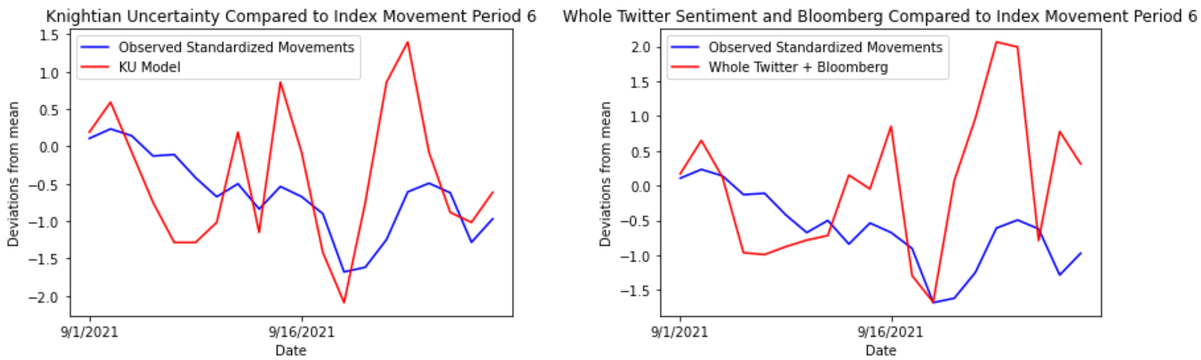


Figure 17 C: Depicts a subsection of the unaugmented models against the macroeconomic indicator variables over the month of September.

In general, our mean squared errors were quite low for this month. Our unaugmented Knightian Uncertainty model had an MSE of around **0.79**, whereas our augmented Knightian Uncertainty models did poorly overall and did not increase predictive accuracy. Specifically, the MSE of the Whole Twitter + Bloomberg model is **1.63**, resulting in an approximate **106.6%** increase in error.

5.2.3: October

We define October as Period 7. We notice from figure 18 A that there is a generally building trend in the movement of financial markets for nearly the entire month of October, with much of the month being notably far away from the whole period mean.

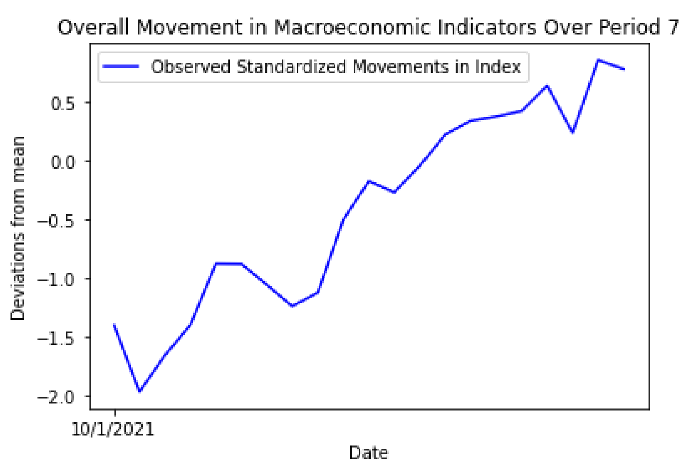


Figure 18 A: Depicts the standardized sum of movements in our macroeconomic indicator variables over the month of October.

Name	Mean Squared Error	Change in MSE from KU Model (%)
Knightian Uncertainty Model	1.9356530470787878	-
KU Model + Google Sentiment	1.898104923365746	- 1.9398168369950335
KU Model + Google Magnitude	1.5119082218857194	- 21.891569144200073
KU Model + VADER Sentiment	2.172679986482359	12.245321534315405
Google Sentiment + Magnitude + KU Model	1.6888822372061507	- 12.748710841803701
Google Sentiment + VADER + KU Model	2.050607428219857	5.938790596514909
Google Magnitude + VADER + KU Model	1.8694079856548274	- 3.4223623662275084
Whole Twitter + KU Model	1.914846454680692	- 1.0749133182465882

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

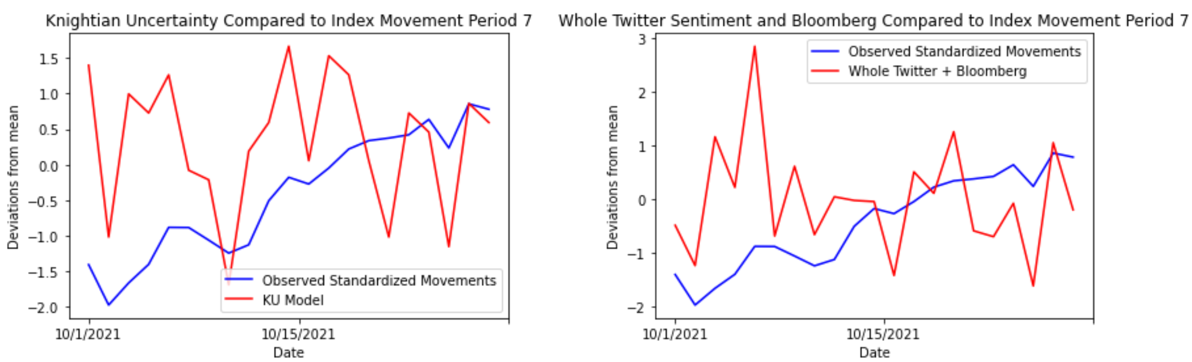


Figure 18 C: Depicts a subsection of the unaugmented models against the macroeconomic indicator variables over the month of October.

The unaugmented Knightian Uncertainty model performed marginally worse this month, having an MSE of approximately **1.94**. In contrast, the augmented Knightian Uncertainty model had an MSE of **1.91**, resulting in a **1.1%** decrease in error.

5.2.4: November

We define November as Period 8. Looking at Figure 19 A, we found that November's financial market movements, while generally building in the first half of the month and falling in the second, are quite volatile. Further, one can note that a substantial portion of the data remains relatively far away from the whole period mean.

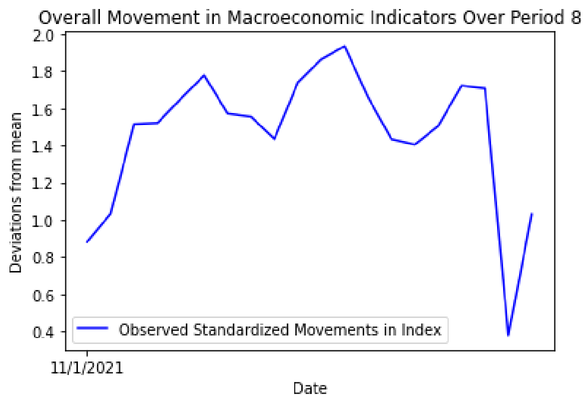
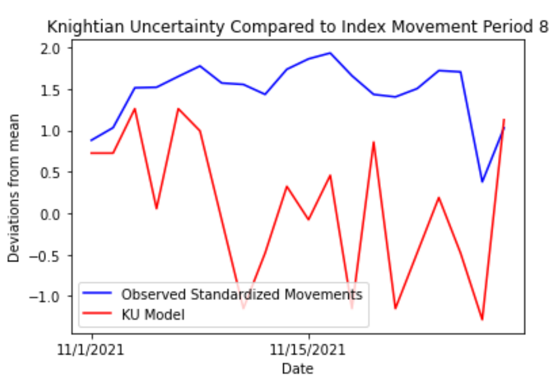


Figure 19 A: Depicts the standardized sum of movements in our macroeconomic indicator variables over the month of November.

Name	Mean Squared Error	Change in MSE from KU Model (%)
Knightian Uncertainty Model	2.6727554917337906	-
KU Model + Google Sentiment	3.3849307965323674	26.645733476225903
KU Model + Google Magnitude	3.2351884708620906	21.04318860695541
KU Model + VADER Sentiment	3.3885249875178056	26.780208589888723
Google Sentiment + Magnitude + KU Model	3.443322339907064	28.830428019190567
Google Sentiment + VADER + KU Model	3.4596412994016505	29.440994887168475
Google Magnitude + VADER + KU Model	3.3620906428809127	25.791178926732172
Whole Twitter + KU Model	3.472410787611595	29.918759809902244

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



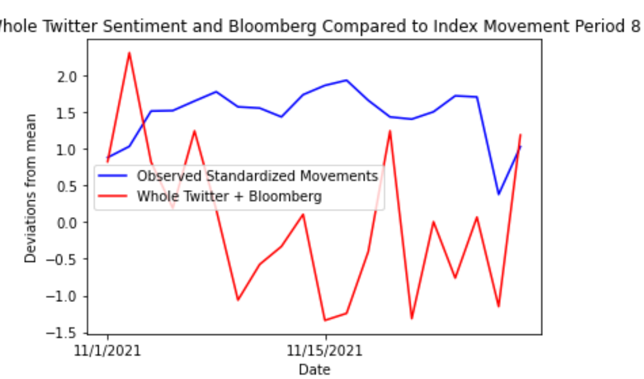


Figure 19 C: Depicts a subsection of the unaugmented models against the macroeconomic indicator variables over the month of November.

This again translates to significantly high root mean squared errors for all models. While the unaugmented KU model has an MSE of **2.67**, the augmented models increase MSE. Specifically, with the Whole Twitter plus Bloomberg model resulting in a MSE of **3.47**, this equates to an approximately **30%** increase in error.

5.2.5: Conclusions

The exercise of reducing granularity on a monthly basis yielded mixed results, with augmented models sometimes performing better and other times worse than the original Knightian Uncertainty model. Specifically, we found that overall model accuracy increased significantly during the month of August, and marginally in October. On the other hand, the augmented models performed significantly worse in the months of September and November. Though it is hard to draw conclusions from this evidence, it is worth noting that our augmented model significantly increases accuracy in August, which is the month closest to the whole period mean. This leads to the assertion that the augmented models increased predictive accuracy near the mean.

5.3: KU Models with Twitter Data by Event

Finally, we constructed localized models around Knightian Uncertainty events in order to further test our hypothesis that reducing granularity would increase model predictive accuracy. 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

We define Event 1 to be Period 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.

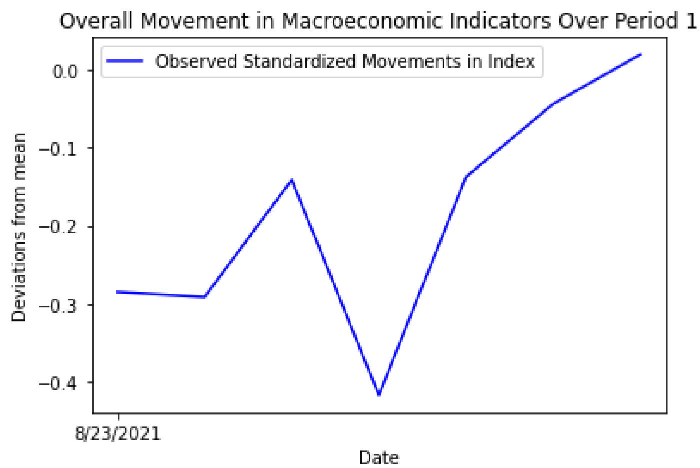


Figure 20 A: Depicts the standardized sum of movements in our macroeconomic indicator variables over Event 1.

Name	**Mean Squared Error**	**Change in MSE from KU Model (%)**
Knightian Uncertainty Model	2.3360142755734405	-
KU Model + Google Sentiment	0.9489704645402465	- 59.37651261539081
KU Model + Google Magnitude	1.109880137817815	- 52.48829814854776
KU Model + VADER Sentiment	1.0624864606816746	- 54.5171246686471
Google Sentiment + Magnitude + KU Model	1.0617586468190918	- 54.548280893598
Google Sentiment + VADER + KU Model	1.0063108279598507	- 56.92188877087134
Google Magnitude + VADER + KU Model	1.1375695464576692	- 51.302971118255655
Whole Twitter + KU Model	1.0886524744814847	- 53.397011060035396

Figure 20 B: Depicts the MSE and each model's corresponding percentage change in error as compared to the unaugmented Knightian Uncertainty model over Event 1.

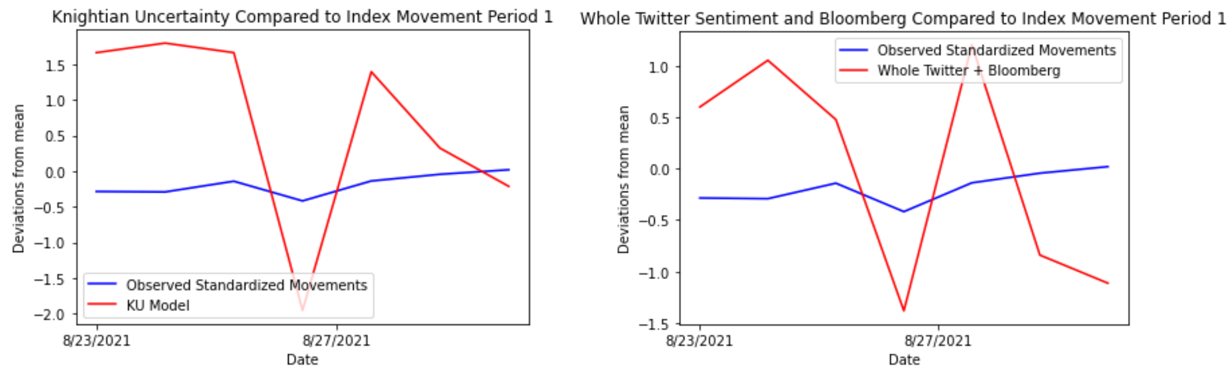


Figure 20 C: Depicts a subsection of the unaugmented models against the macroeconomic indicator variables over Event 1.

As can be seen from 20 B, the unaugmented Knightian Uncertainty model resulted in a MSE of **2.34**, which was significantly higher than the errors produced by the augmented KU models. Specifically, the Whole Twitter plus Bloomberg model had a MSE of **1.09**, resulting in a **53.4%** reduction in error.

5.3.2: Event 2

We define Event 2 to be Period 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.

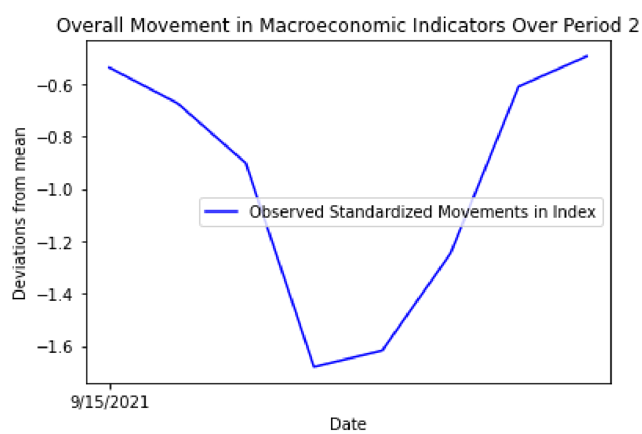


Figure 21 A: Depicts the standardized sum of movements in our macroeconomic indicator variables over Event 2.

Name	Mean Squared Error	Change in MSE from KU Model (%)
Knightian Uncertainty Model	1.5169331784666182	-
KU Model + Google Sentiment	1.568078731565731	3.3716417984088705
KU Model + Google Magnitude	1.6131568058936794	6.343300337350928
KU Model + VADER Sentiment	1.7409470164104839	14.767548177060016
Google Sentiment + Magnitude + KU Model	1.5682921564104295	3.3857112938703837
Google Sentiment + VADER + KU Model	1.6261791853980045	7.201767914511363
Google Magnitude + VADER + KU Model	1.6503439808094822	8.794771202626169
Whole Twitter + KU Model	1.6099540283852423	6.132165295023313

Figure 21 B: Depicts the MSE and each model's corresponding percentage change in error as compared to the unaugmented Knightian Uncertainty model over Event 2.

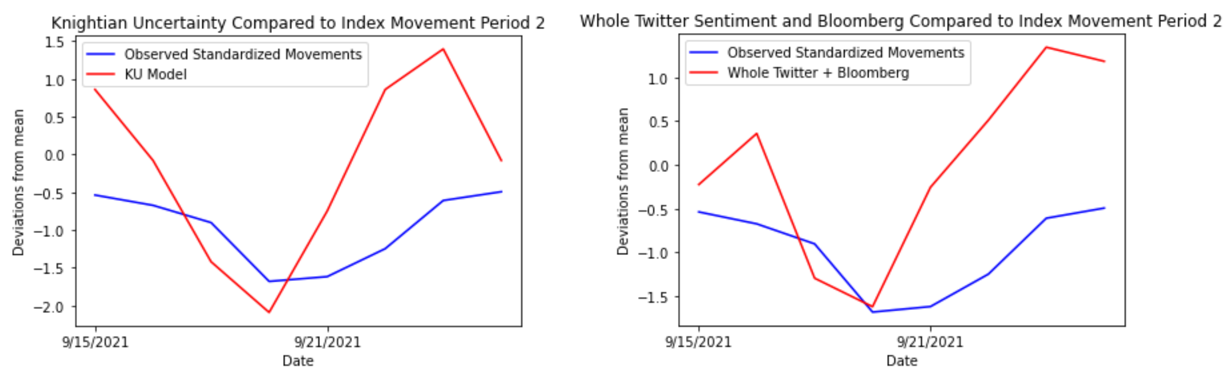


Figure 21 C: Depicts a subsection of the unaugmented models against the macroeconomic indicator variables over Event 2.

As can be seen from figure 21 B, the unaugmented Knightian Uncertainty model resulted in a MSE of **1.52**, which was slightly lower than the errors produced by the augmented KU models. Specifically, the Whole Twitter plus Bloomberg model had a MSE of **1.61**, resulting in a **6.13%** *increase* in error.

5.3.3: Event 3

We define Event 3 to be Period 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.

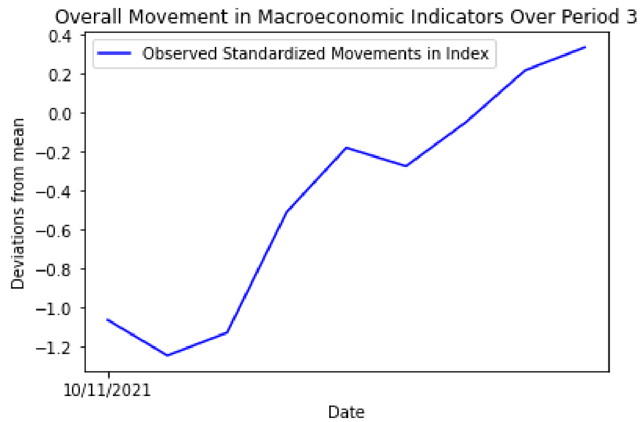
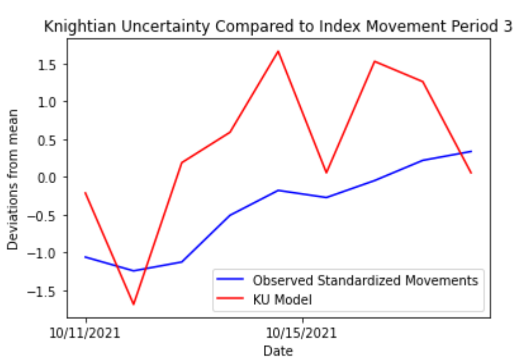


Figure 22 A: Depicts the standardized sum of movements in our macroeconomic indicator variables over Event 3.

Name	Mean Squared Error	Change in MSE from KU Model (%)
Knightian Uncertainty Model	1.2248798001854402	-
KU Model + Google Sentiment	0.749343203443075	- 38.823123433856246
KU Model + Google Magnitude	1.4527226823562438	18.601244149532832
KU Model + VADER Sentiment	1.3472876591160938	9.993458861197789
Google Sentiment + Magnitude + KU Model	0.9611556152756796	- 21.530617524252925
Google Sentiment + VADER + KU Model	0.9639675208064493	- 21.301051690091562
Google Magnitude + VADER + KU Model	1.290989039793047	5.397202206910283
Whole Twitter + KU Model	1.0883218385700428	- 11.14868263765336

Figure 22 B: Depicts the MSE and each model's corresponding percentage change in error as compared to the unaugmented Knightian Uncertainty model over Event 3.



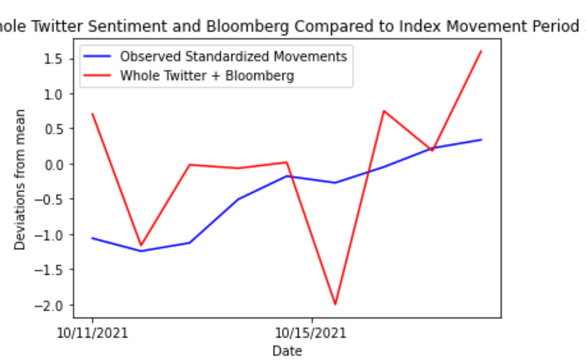


Figure 22 C: Depicts a subsection of the unaugmented models against the macroeconomic indicator variables over Event 3.

As can be seen from figure 22 B, the unaugmented Knightian Uncertainty model resulted in a MSE of **1.22**, which was notbally higher than the errors produced by the augmented KU models. Specifically, the Whole Twitter plus Bloomberg model had a MSE of **1.09**, resulting in a **11.15%** reduction in error.

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.

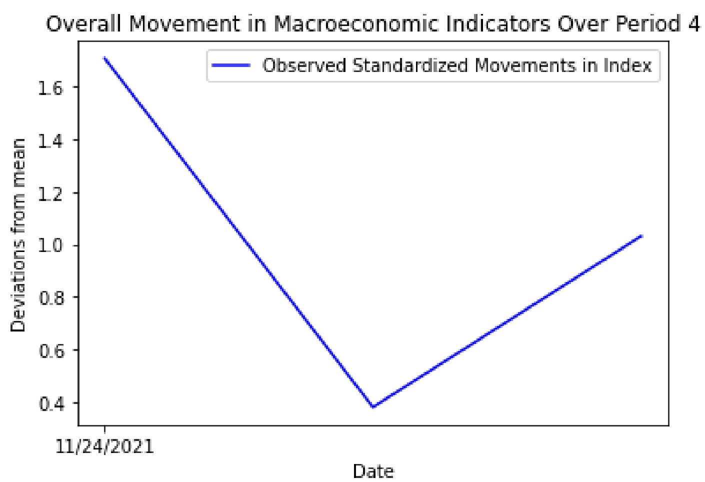


Figure 23 A: Depicts the standardized sum of movements in our macroeconomic indicator variables over Event 4.

Name	Mean Squared Error	Change in MSE from KU Model (%)
Knightian Uncertainty Model	2.5245760209401493	-
KU Model + Google Sentiment	1.5353572371929511	- 39.18356094417843
KU Model + Google Magnitude	1.9209266608228068	- 23.910920293560586
KU Model + VADER Sentiment	2.1651869469063185	- 14.235620993500314
Google Sentiment + Magnitude + KU Model	1.4703861073536002	- 41.757107127792885
Google Sentiment + VADER + KU Model	1.2915719286673133	- 48.84004609271646

Google Magnitude + VADER + KU Model	2.0162381505245337	- 20.135573902278892
Whole Twitter + KU Model	1.6498028574957164	- 34.65029993902375

Figure 23 B: Depicts the MSE and each model's corresponding percentage change in error as compared to the unaugmented Knightian Uncertainty model over Event 4.

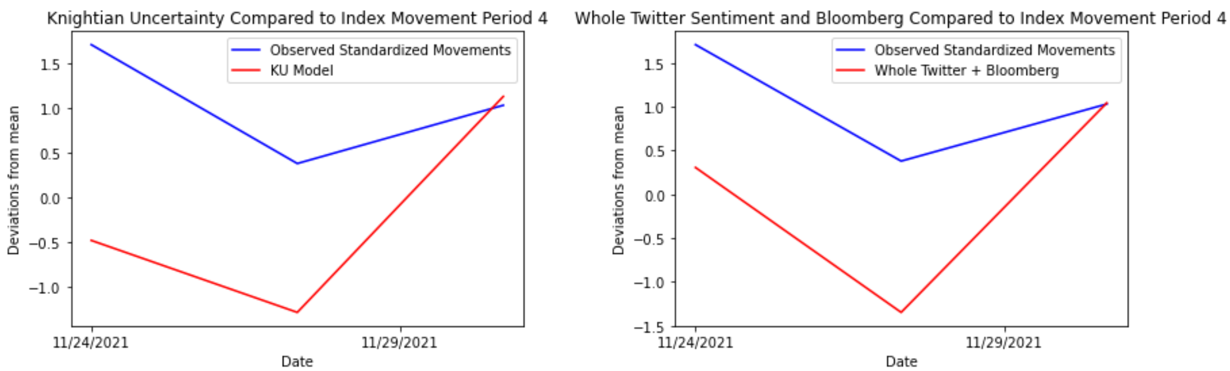


Figure 23 C: Depicts a subsection of the unaugmented models against the macroeconomic indicator variables over Event 4.

As can be seen from figure 23 B, the unaugmented Knightian Uncertainty model resulted in a MSE of **2.52**, which was notably higher than the errors produced by the augmented KU models. Specifically, the Whole Twitter plus Bloomberg model had a MSE of **1.65**, resulting in a **34.65%** reduction in error.

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 encouraging results. Namely, we found that using augmented KU models led to a strong reduction in MSE in three out of four events, which only once resulted in a slight increase in error. In this case the errant event, event two captured the near collapse of the Chinese real estate company Evergrande. 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.

These results lend credence to our previous assertion - that our Twitter Augmented KU models increase predictive accuracy when near the whole period mean. It is clear that the macroeconomic indicator variable data in events one, three and four, while volatile, are *on average* closer to the whole period mean than event two. Specifically, event one is, by selection, close to the mean; event three, while starting well below the mean, generally builds towards it as time goes on; and event four, while starting well above the mean, generally falls towards it for much of the period. In stark contrast, event two starts fairly far away from the mean, falls further below it and only recovers insofar as returning to its initial distance from the mean. Thus, as our assertion would predict, our Twitter Augmented KU models performed better during those events closer to the mean.

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, macro economic indicators chosen, choice of analysis periods and sentiment analysis algorithm used. Further, it is worth noting that the mean squared error is one of many distance metrics, 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. All of these limitations could potentially have been refined and expanded to produce more accurate results.

6.2: Augmented Model Comparison

Name	Number of Times Improved
KU Model + Google Sentiment	5
KU Model + Google Magnitude	4
KU Model + VADER Sentiment	3
Google Sentiment + Magnitude + KU Model	5
Google Sentiment + VADER + KU Model	4
Google Magnitude + VADER + KU Model	4
Whole Twitter + KU Model	5

Figure 24: Describes how many times each model type increased predictive accuracy over the unaugmented Knightian Uncertainty model for all periods of analysis.

As a means of comparing model types, which are detailed in figure 24, we kept track of each instance in which the augmented models resulted in reduced mean squared errors when compared to the original Knightian Uncertainty model. Among these models, our results indicate that the KU model augmented with Google Sentiment, the KU model augmented with both Google Sentiment and Magnitude, and the KU model augmented with Whole Twitter data all increased predictive accuracy in 5 out of 9 total cases. Notably, models augmented primarily with Google Sentiment seemed to more consistently improve performance.

6.3: Conclusion

In attempting to model movement in macroeconomic indicator variables, a Knightian Uncertainty approach to Bloomberg Market Wraps produced on average a lower mean squared error than the REH or Behavioral Models over our entire period of analysis. Further, though attempts to augment this Knightian Uncertainty model with sentiment gathered from influential users on Twitter proved unsuccessful on the level of the entire period of analysis, those attempts with reduced granularity were often fruitful. Specifically, on the monthly level, Twitter Augmented Knightian Uncertainty models were more accurate two out of four months, notably during those months near the whole period mean. Furthermore, model performance was improved to a greater extent by restricting granularity to a narrower period of analysis, i.e., the event level. Here, KU augmented models performed better during three out of four events. These results are in line with our hypothesis. A potential further research question could be whether further augmenting our data with other social media platforms would lead to more conclusive results.

Bibliography

- https://twitter.com/elonmusk/status/1256239815256797184?lang=en, 10th January, 2022
- https://developer.twitter.com/en, 10th January, 2022
- https://mobile.twitter.com/tesla?lang=en, 10th January, 2022
- https://twitter.com/elonmusknewsorg?lang=en, 10th January, 2022
- https://developer.twitter.com/en/docs/twitter-api/fields, 10th January, 2022
- *Sentiment analysis algorithms and applications: A survey*, Medhat, Hassan, Korashy, 2013; https://www.sciencedirect.com/science/article/pii/S2090447914000550, 10th January, 2022
- https://github.com/cjhutto/vaderSentiment, 10th January, 2022
- https://cloud.Google.com/natural-language/docs/analyzing-sentiment, 10th January, 2022
- https://finance.yahoo.com/quote/%5EGSPC/, 10th 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, 10th 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: Fundamentals

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: Psychology 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 25: 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.