Undergraduate Economic Review

Volume 18 | Issue 1 Article 2

2021

The Effect of Epidemiological Investor Sentiment on Financial Market Movements

Ruben A. Silverstone
University of Maryland, College Park, ruben.silverstone@gmail.com

The Effect of Epidemiological Investor Sentiment on Financial Market Movements

Abstract

This paper investigates the effect of public sentiment related to epidemiological crises on financial market movements. The outbreak of COVID-19 provided evidence of the havoc a pandemic can wreak on financial markets. The Ebola outbreak between December 2013 and January 2016 provides the ideal case study to isolate sentiment. Sentiment was quantified with established text processing methods, using news on viral events uncorrelated with other potential causes of market movements and incorporating publisher circulation. I find that epidemiological investor sentiment has a highly statistically significant, current, and non-linear relationship with individual company stock returns when controlling for company-specific fixed effects.

Keywords

financial markets, epidemiology, stock market, ebola, sentiment, text processing

Cover Page Footnote

I would like to thank: Prof. Boragan Aruoba Prof. Nuno Limao Jaehong Choi Jennifer Biaksangi Dr. Paul E. Soto The University of Maryland Honors College Staff

I. Introduction

The outbreak of COVID-19 has acted as a natural experiment that provided ample evidence of the havoc that a pandemic can wreak on financial markets. Beginning with a macro perspective on the overall economy, the outbreak brought the longest economic expansion in United States history to an abrupt halt and caused unemployment to skyrocket from 3.8% to a peak of 14.4%. According to the Brookings Institution, the economy experienced the steepest quarterly decline in economic output in its recorded history, falling 9.1 percent in the second quarter of 2020. Corresponding to the drop in GDP, financial market indices also fell drastically. The S&P 500 index dropped more than 20% in only 16 trading days, the fastest drop into a bear market and nearly twice as fast as the drop precipitating the Great Depression. The Dow Jones industrial average experienced its largest single day drop since Black Monday in 1987. This is certainly not to undermine the ensuing public health crisis and exposure of deep-rooted issues creating inequities among socioeconomic, racial, and ethnic groups; however, these topics require consideration in their own right and are not within the scope of this investigation.

By contrast, the outbreak of *Zaire ebolavirus* (Ebola) that lasted from December 2013 to early 2016 never meaningfully spread beyond West Africa, and therefore never attained the same level of global disruption. Importantly however, from the perspective of those in the early stages of the outbreak, there was no clear indication whether it would remain an epidemic contained within West Africa or evolve into a full-scale pandemic. According to the CDC, poor surveillance and public health infrastructure in rural areas contributed to the rapid spread of the virus to the capitals of Guinea, Liberia, and Sierra Leone. This was the first time that Ebola had spread to densely populated urban areas, and thus provided for an unprecedented chance for transmission. Given that the virus could have spread, coupled with the fact that it did not, provides an ideal case study for isolating the particular vehicle of epidemiological sentiment on the resulting financial market activity. Past research provides the theoretical backing for the intuitive notion that the Ebola outbreak affected financial markets by specifying media coverage and public risk aversion as the means through which it did so. (Ichev and Marinč 2017; Barker and Bacon 2015).

Throughout both crises, public news cycles were dominated by coverage of their respective outbreak, providing a wealth of information regarding current market perceptions. Text processing, or text mining, is the process of deriving unique and actionable information from natural human language. There has been an accumulating body of research that provides a wide range of potential methodologies and data sources for examining financial markets (Hassan et al.

2019; Loughran and McDonald 2011), as well as research that narrows down the list and specifies unique strategies to be used in a viral context (Hassan et al. 2020).

This paper investigates the effect of public sentiment related to epidemiological crises on financial market movements. Sentiment was quantified with established text processing methods, using news on viral events uncorrelated with other potential causes of market movements as data points and incorporating weights based upon publisher circulation. In addition, a novel sentiment dictionary was created by surveying epidemiologists to capture unique variations in the meaning of language in the specific context of epidemiology.

Ebola

A detailed history of the *Zaire ebolavirus* outbreak that occurred between December 2013 and January 2016 is not within the scope of this paper, but a brief overview of the timeline is beneficial. The initial case of the outbreak occurred in December 2013 in Guéckédou, a rural town in the south of Guinea. The virus was confirmed to have spread to neighboring Libera by March 2014 and Sierra Leone by May 2014. By July of 2014, the virus had spread to the capitals of all three countries, providing an unprecedented chance for transmission. The first confirmed, active case outside of Africa occurred in early October 2014 in the United States. The vast majority of cases occurred within these three countries, but in total, seven additional countries had declared cases of the virus: Italy, Mali, Nigeria, Senegal, Spain, the United Kingdom, and the United States.

The WHO declared Ebola outbreak on March 23, 2014, and further declared an "international health emergency" on August 8, 2014. By this time, it had already killed nearly 1,000 people. According to the CDC, the spread of the virus "can be attributed to the unprecedented circulation of EVD into crowded urban areas, increased mobilization across borders, and conflicts between key infection control practices and prevailing cultural and traditional practices in West Africa."

It is also important not to understate the severity of this disease; the average mortality rate across multiple past outbreaks is roughly 50 percent, but some outbreaks have had mortality rates around 90 percent. Fortunately, the virus cannot be transmitted through airborne means, and is spread through contact with infected blood, secretions, organs, or other bodily fluids.

II. Literature Review

Financial markets are typically assumed to accurately price assets and securities, as the collective forces of supply and demand quickly eliminate any advantageous opportunities. To illustrate this point, the efficient market hypothesis states that all publicly available information is immediately incorporated into a security's trading price, making it impossible to consistently outperform the market through security selection or market timing. However, given that traders are human, they are also swayed by a variety of factors that affect feelings and behavior. As an example of previous research demonstrating this, *The Effect of New York City Sports Outcomes on the Stock Market* found that markets experienced abnormally high returns after a professional New York City sports team won a championship, and abnormally low returns after an elimination (Levy, 2015).

This investigation will study the effect of news related to the West African Ebola outbreak that occurred between December 2013 and January 2016 on financial market movements using Python Natural Language Processing (NLP) analysis techniques. As such, it necessitates the examination of three unique subjects: Ebola, financial markets, and text processing. Past research has focused on groupings of two of these subjects – Ebola and financial markets, financial markets and text processing, and Ebola and text processing. The following sections of this paper will chronicle past literature in these respective areas, and highlight key findings, as well as key differences or shortcomings, that are relevant to this current investigation.

Ebola and Financial Markets

The literature discussing the relationship between Ebola and financial markets varies in scope, but most focus on the means through which the outbreak of the virus affected market returns, and the factors that result in differing effects across different market participants.

One notable paper exploring relevant factors in determining the relative effect of Ebola on individual securities is that of Ichev and Marinč (2017). The authors used an event study model to explore whether geographic proximity of information affected US stock prices, finding that its effects were strongest for "the stocks of companies with exposure of their operations to the West African countries and the U.S. and for the events located in the WAC and the U.S." (Ichev and Marinč, 2017). The authors also show that the effects were more pronounced among stocks with, among other variables, "intense media coverage", showing it to be a means for the outbreak to impact financial markets. Importantly, however, the authors did not further explain the vehicle through which that media coverage affected markets. They briefly mentioned the effect of investor sentiment but used the Chicago Board Options Exchange VIX and VXO as a cursory proxy.

LeeAnna Barker and Frank Bacon (2015) extend the findings of the previous paper by using the Ebola outbreak to assess the semi-strong form market efficiently hypothesis through the speed at which Ebola related information is incorporated into market pricing. The authors explain that, "indirect costs of public risk aversion can generate far more economic damage than the direct cost of healthcare outlays and other containment expenditures," (pg. 2) indicating another means for the outbreak to impact financial markets (Barker and Bacon, 2015). Tracking the risk adjusted rate of return on 15 randomly selected airline stocks, the authors found evidence of the semi-strong form of market efficiency. The semi-strong form of efficiency states that current prices found in financial markets have all publicly available information factored in. This is in contrast to the strong form of efficiency, which states that prices have all information, public and private, factored in, a theory generally thought to be less defensible.

These papers provide a theoretical backing for the intuitive notion that the Ebola outbreak affected financial markets by specifying media coverage and public risk aversion as the means through which it did so. Although not directly applicable, the findings of Ramelli and Wagner (2020) support these conclusions using evidence from the COVID-19 pandemic. Their research showed that investor perceptions of companies changed rapidly, "mov[ing] feverishly" (pg. 651-2) between shunning and seeking out companies with different global value chains and international exposures as the situation evolved. Yet despite these fluctuations, investors did exhibit a clear preference for companies with cash holdings, a practice which is typically viewed as unfavorable due to high opportunity costs and potential agency problems. Again, this demonstrates public risk aversion as a means for a viral outbreak to affect financial markets. Although the outbreaks of COVID-19 and Ebola ultimately unfolded very differently and certain factors cannot be generalized to encompass other viral events, insights into investor sentiment can still be applied.

At a fundamental level, the complex relationship between sentiment and financial markets can be explained with theories of behavioral finance, the study of how psychology influences the behavior of market participants. In this regard, Barberis and Thaler (2003) provide an excellent survey of relevant theories which catalogues and justifies traders' deviations from full rationality. Importantly, such theories can also explain how rational traders are unable to correct such mispricing, allowing the results of irrationality to persist.

One notable example is belief perseverance, revealed by Lord, Ross, and Lepper (1979), which suggests that once people form opinions on complex issues, they are likely to approach evidence on the issue in a biased manner. This is primarily because they are likely to accept evidence that supports their beliefs at face value while treating contradicting evidence with critical skepticism, which results in polarization. Additionally, they are less likely to seek out contradicting evidence. Belief perseverance was confirmed to exist in a financial context by Ko and Huang (2012), who concluded that investor perception of information was biased in favor of previously preferred assets.

Another example is availability bias, described by Tversky and Kahneman (1974) as the imperfect judgement of probability due to memories and experiences being imperfectly available for recollection. This bias was confirmed to exist in a financial context by Schraeder (2014), who utilized an overlapping generations equilibrium model to show that differences in the time of experience history leads to heterogeneity in trading among agents.

Financial Markets and Text Processing

Existing literature sets the precedent for the use of a wide variety of text processing methods and data sources when analyzing financial markets.

The most relevant, prominent paper in this area is *Firm-Level Political Risk: Measurement and Effects*, by Hassan et al. (2019). The authors propose a text-based risk measure for quantifying firm exposure to political risk using the proportion of quarterly earnings conference-call transcripts devoted to this issue. Earnings call transcripts are used as they contain direct responses from senior management to questions about the firm. With political risk being such a vague topic, this paper used a library of political texts to discover political bigrams, or two-word combinations that are frequently used in a political context in close proximity to synonyms for “risk” and “uncertainty”. The measure is validated through its intuitive variation across time, various sectors, and with firms taking anticipated action. The results show that firms with higher levels of exposure to a given political risk have a higher likelihood of increasing lobbying in relation to that specific risk. Another important finding from this paper is that the majority of variation was found at the firm level, showing that such text-based risk measures can capture unique risk exposures that are not present at the aggregate level.

Another important, more established paper is that of Loughran and McDonald (2011). The significance of this paper to this investigation is in its emphasis on context. The authors show that previous research methodologies in text analysis tend to misclassify common words in financial text, as many typically negatively connoted words are not considered negative in finance. As such, they created five new tone dictionaries, specifically for financial contexts, and show that it more accurately reflects outcomes in a variety of business actions.

Lastly, while not uniquely relevant to this investigation, Loughran and McDonald (2016) provide an excellent overview of the most common practices in text analysis methods in a financial context. The authors’ focus on the ambiguous elements text analysis, and the situations in which the art of a methodology in question is equally important to the science, is noteworthy. While the full range of the topics discussed is not relevant to this investigation nor within the scope of this review, a few significant examples are existing Word Lists, the Bag-of-Words assumption, TF-IDF, Naïve Bayes, and Latent Dirichlet Allocation.

Text Processing and Ebola

There is considerably less research utilizing text processing methods in the study of viral outbreak affects. Given that there is such a wide variety of these methods, as demonstrated in the previous section, it is important to identify text processing methods and text-based risk measures appropriate for a virus specific context.

Hassan et al. (2020) utilizes a "general text-classification method and identifies the exposure of firms to an outbreak of an epidemic disease by counting the number of times the disease is mentioned in the quarterly earnings conference call that public listed firms host with financial analysts." (pg. 1). The authors validate the use of this method by citing its use in past research to measure firm exposure to other threats such as political risk, Brexit, and Fukushima, as described in the previous section. Again, they justify using earnings call transcripts as a data source in this context as it contains direct responses from senior management to questions about the firm. According to the authors, this reduces the company's ability to manipulate information or craft responses that helps its image.

Hassan et al. also incorporate text-based measures that are more relevant to this investigation. To quantify a company's exposure to disease risk, they employ a metric based upon a count of synonyms for the disease that are within a given proximity for synonyms of "risk", normalized for the length of the earnings transcript. To quantify a company's exposure to disease sentiment, they employ a metric based upon a count of synonyms for the disease combined with the sum of a sentiment score for each word within a given proximity, again normalized for the length of the earnings transcript. It is also essential to understand how the list of synonyms was created. For synonyms for disease, the authors identified the most common synonyms of a disease in question using online resources and newspapers (followed by a "human audit"), for synonyms for "risk", the authors used the Oxford English Dictionary, and for sentiment scores, the authors used a dictionary created by Loughran and McDonald (2011).

Both of these metrics are useful guides for text-based measures in this investigation, as such methodologies can be adapted for the study of Ebola in news articles.

Conclusions and Marginal Contribution

Combining the findings of past research demonstrates the relationship at the intersection of these three unique subjects and provides both the theoretical and methodological framework for this investigation. Literature examining Ebola and financial markets provides the theoretical means through which a viral outbreak affects market returns, literature related to financial markets and text processing provides a wide range of potential methodologies and data sources, and literature related to text processing and Ebola narrows down the list and specifics unique strategies to be used in a viral context. Together, they form a comprehensive backdrop for use of text analysis to study the effect of epidemiological sentiment on financial markets.

This paper will build upon the aforementioned research by quantifying the effects of epidemiological sentiment on financial markets. Findings and methodologies from this investigation could contribute to the body of knowledge informing investor approaches to novel news on both current and imminent pandemics. The text processing method developed using a novel sentiment dictionary created with the help of epidemiologists can also be applied in future research as a tool for forecasting market movements.

III. Approach

Model

This paper uses two models. The first is a time series regression with a broader scope that explores the effect of sentiment on broader market indices based upon the assumption that Ebola had no significant determinant effects on the overall market other than through sentiment. This assumption will be discussed further in the Appendix. The second is a panel regression that explores the effect of sentiment on individual S&P500 companies while controlling for company fixed effects.

The first model regresses an index's rate of return on a given day on the average numerical sentiment score of news articles and abstracts on that day, a single trading day lag of the index's rate of return, and the day of the week. "Sentiment" was quantified with text processing methods, using news on specific viral events uncorrelated with other potential causes of market movements. The metric also incorporates a weight based upon the publisher of any given article, as the circulation provides a proxy for the number of readers and thus the ability of the article's sentiment to influence market movements.

$Returns_t$ = Daily return for a stock market index

$Returns_{t-1}$ = Previous trading day's daily return for the same index

$Sentiment_t$ = Average numerical sentiment score of news articles on that day, taken as the average of title scores and abstract scores

D_i = Dummy variables for day of the week

$$\ln(Returns_t) = \alpha + \beta_1 \ln(Returns_{t-1}) + \beta_2 Sentiment_t + \sum_{i=1}^{4} \theta_i D_{it} + \epsilon_t$$

A lag of the index's previous trading day returns is included to account for momentum in financial markets. Dummy variables for day of the week are included to account for other fixed effects that vary by weekday. The inclusion of both variables is relatively common practice in such regressions.

The second model regresses an individual company's rate of return on a given day on the same market-wide sentiment score, a non-linear sentiment term, a single trading day lag of the company's rate of return, and the day of the week. A fixed-effects panel regression model was used to capture unobserved firm-specific effects that may affect daily returns.

$$Returns_{it} = \alpha + \beta_1 Returns_{it-1} + \beta_2 Sentiment_t + \beta_3 {Sentiment_t}^2 + \sum_{j=1}^{4} \theta_{ij} D_{jt} + \varepsilon_t$$

The second model does not rely on the assumption that Ebola had no significant determinant effects on the overall market other than through sentiment; instead, it controls for the various ways in which Ebola could affect an individual company's cash flows and thus its stock price.

IV. Data

News Data

News data was obtained from ProQuest, a database aggregation of a variety of content formats. Data was obtained by searching the database for content mentioning 'Ebola' between the date range of December 15, 2013 to January 15, 2016. Results were filtered by language, selecting only those in English, and type, selecting only newspaper and magazine articles. This intentionally excludes formats such as conference papers, dissertations, theses, and scholarly journals, as these formats are not typical news sources for the general public. The database provides a wide range of variables for each observation, but the ones relevant to this investigation are the article title, abstract, and publication data. The author(s), place of publication, and publication title were also gathered to form a complete picture of the data used.

In total, 25,709 observations were collected. The data was cleaned by removing stop-words, extremely common words such as 'the', 'is', and 'a' that carry no significant meaning on their own. However, stop-words that act as negations such as ‘don’t’ and ‘wasn’t’ were not removed. HTML syntax was also removed, an unfortunate side effect of digitally collected text data.

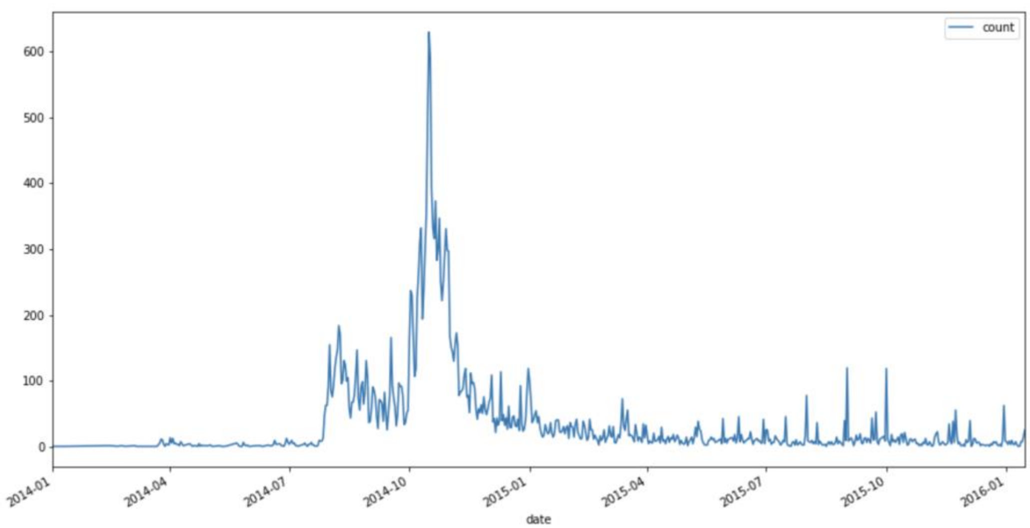
The average length of the clean article titles was 7.197 words, and the average length of the abstracts was 38.497 words. The following table shows a statistical summary of the data length:

Table 1: Statistical Summary: News Data

VARIABLES	(1) Obs	(2) mean	(3) sd	(4) min	(5) max
Article Title Length	25,709	7.20	4.17	1	36
Article Abstract Length	25,709	38.50	32.47	1	1,345

The following graph shows the frequency of article occurrences over time:

Exhibit 1: Count of Articles Over Time



Within the data, 270 article titles and 571 article abstracts had negations present, representing about 1.5% of total data points. These observations have been removed from the data, removing concerns of incorrectly classifying articles due to negations. It should be noted that when a negation appeared in a title, its corresponding abstract was not removed, and when a negation appeared in an abstract, its corresponding title was not removed. This is because their corresponding title and abstract were phrased differently, and thus do not share the concern of having their sentiment misclassified due to the presence of negations. The full list of negations searched for can be found in the Appendix.

Finally, articles that did not have a consistently defined publication date were also dropped.

Epidemiological Sentiment Dictionary

Words for the epidemiological sentiment dictionary were selected by finding the most commonly used words directly used to discuss Ebola. This was achieved by creating 10 word "neighborhoods" around each occurrence of the word 'Ebola', as doing so increases the likelihood that the words are being used to discuss Ebola. These words thereby contribute to epidemiological sentiment, as opposed to sentiment concerning another topic. This methodology was adapted from Hassan et al. (2020). If a word occurred within 5 words of the beginning or end of the text, the neighborhood was shortened accordingly.

The 150 most commonly used words in these "neighborhoods" were selected, excluding words with an obvious sentiment (obvious is defined as having a sentiment classification in either the Harvard IV-4 or Loughran and Mc Donald Financial Sentiment Dictionaries approved for use in an epidemiological context by a human audit), and meaningless words such as locations, numbers, and names. The detailed process for finding the final 150 words can be found in the Appendix. These 150 words were then sent to 10 epidemiologists for classification, with each expert receiving a list of 45 random words such that each word was reviewed exactly 3 times. The expert classifications were then compiled into a dictionary, where each word was classified as possessing the sentiment determined by the majority "vote" of epidemiologists who reviewed it.

A random selection of 8 words from the dictionary can be found below, where a -1 represents negative sentiment and a 1 represents a positive sentiment; the complete dictionary can be found in the Appendix:

Exhibit 2: Sample: Sentiment Dictionary

Word	**Classification (Pos/Neg)**
cases	-1
diagnosis	-1
exposure	-1
incubation	-1
transmission	-1
blood	1
clinical	1
doctor	1

Once the dictionary was created, every cleaned news article title and abstract without negations present was assigned a sentiment score. This was done utilizing the same 10 word "neighborhoods" around each instance of the word 'Ebola' within a piece of text in question and compiling these neighborhoods into a list. For every positive word in that list, a value of 1 is added to the existing sentiment score, and for every negative word in that list, a value of 1 is subtracted from the existing sentiment score. A word is determined to be positive or negative in a three-step process:

1. Evaluate the word using the sentiment dictionary
2. Evaluate the word using the Loughran and Mc Donald Financial Sentiment Dictionary
3. Evaluate the word using the Harvard IV-4 dictionary

The process ends whenever a word is assigned a sentiment. Put differently, a word is only evaluated using the Loughran and Mc Donald Financial Sentiment Dictionary if it does not appear in the sentiment dictionary, and a word is only evaluated using the Harvard IV-4 dictionary if it does not appear in either the sentiment dictionary or the Loughran and Mc Donald Financial Sentiment Dictionary. If a word does not appear in any of the three dictionaries, it does not impact the sum. Once the sum of all sentiments for each word in a neighborhood has been obtained, the final score is normalized by dividing the sum by the total number of words in the list. This process is conducted for every cleaned news article title and abstract without negations present.

The result of this methodology is that every article evaluated solely on the contents of its neighborhoods of words around the word 'Ebola', with no regard for the date at which the article was published. As such, if the same set of words were to appear in similar sentences at the beginning of the crisis and the end of the crisis, they would be classified in the exact same way. The process would not distinguish between these two sentences based upon the prevailing sentiment or perceptions of the crisis at the time of publication.

However, this does not discredit the results of the methodology because it is intended to evaluate articles in an aggregate manner. While the same individual words or sentence may be evaluated the same way in a negative article at the beginning of the crisis and in a positive article at the end of the crisis, the aggregate sentiment of each article will be different due to the remaining differences in their composition. In theory, the negative article would still be classified as negative due to the rest of its neighborhoods containing negative words, and the positive article would still be classified as positive due to the rest of its neighborhoods containing positive words. The single sentence shared by the two hypothetical articles would be outweighed by their differences. Throughout this draft, there are several examples provided of both an article title or abstract along with its classification score which are shown to accord with a human classification. This shows that the methodology can correctly classify articles, including those with multiple sentences.

Dictionary Verification - Classification of Similar Words

The first method for verifying the dictionary classifications is to compare the classifications of similar words that have intuitively similar meanings. This significantly decreases the likelihood of misclassifications as it effectively demonstrates a consensus among an increasing number of experts. There are many examples of this in the dictionary, some of which are shown below:

‘case’ and ‘cases’ → Negative
‘diagnosed’ and ‘diagnosis’ → Negative
‘doctors’, ‘doctor’, ‘dr’, ‘professionals’, ‘researchers’, ‘scientists’ → Positive
‘clinic’ and ‘clinical’ → Positive
‘lab’, ‘facility’, ‘facilities’ → Positive

It is important to note that there were three instances of similar words without the same classifications. These potential misclassifications were resolved using three discrete procedures.

The first procedure was to replace inconsistent sentiment scores of similar words with the same root with the majority classification across those words. If any one expert had more than one “vote” within the broader majority, only one was counted. This was done to normalize influence among experts in the calculation. If the expert’s multiple “votes” were compatible, then that classification was used to calculate the broader majority. If the expert’s multiple “votes” were contradictory, then the majority classification of their “votes” was used to calculate the broader majority. There were no instances of an expert having an even number of contradictory “votes”, thus preventing the computation of a majority. This sentiment dictionary was titled “Root Majority”.

The second procedure was to simply remove all similar words with inconsistent sentiment scores. This sentiment dictionary was titled “Inconsistencies Removed”.

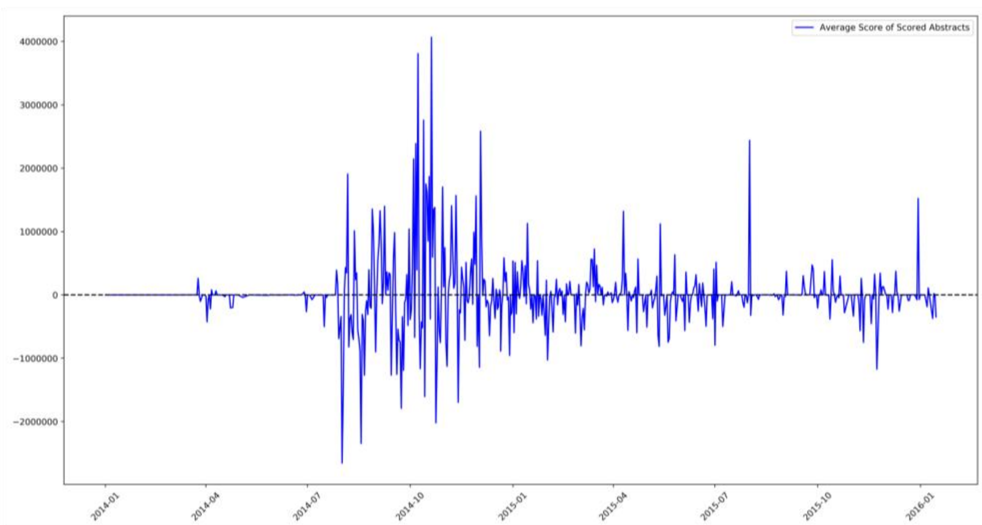
The third procedure was to replace the inconsistent sentiment scores of similar words with a simple average of all classification “votes” across those words. Again, if any one expert had more than one “vote” within the calculation, only one was counted to normalize influence. It should be noted that this resulted in classifications that were not “1” and “-1”, but instead were a value between those bounds. This sentiment dictionary was titled “Inconsistencies Averaged”.

For this investigation, the “Root Majority” dictionary was used as it permits important words to remain in the dictionary, while maintaining its structure for use in future research. All three dictionaries were tested independently using the company model, with similar results. The results of this comparison can be found in the Appendix.

Dictionary Verification – Observable Events

Another method for verifying the dictionary classifications can be obtained after classifying the dataset. A graph of the average score abstracts over time can be found below:

Exhibit 3: Average Score Abstracts Over Time



The measure began as slightly negative but largely negligible throughout early to mid-2014. This reflects a time when cases were increasing in West Africa with poor responses from governments of the affected countries, but a limited response from the international community afforded sparse news coverage. As the issue slowly gains notoriety and news coverage rises, greater levels of negative sentiment are apparent.

Around October 2014 there was a slight increase in cases outside of West Africa, but this was mitigated by the international community responding with more stringent precautions. Most significantly at the end of July, the Liberian president shut down the country's borders, closed schools, and placed selected areas under quarantine. Other examples include the United Kingdom announced temperature screenings at Gatwick and Heathrow airports for arrivals from West Africa on October 7, and the US implemented enhanced Ebola screenings at JFK airport on October 11. This outweighed the relatively insignificant increase in international cases, which is captured by the average score of articles beginning to trend upwards despite volatility due to the erratic nature of public news data and the events unfolding.

Following this, President Obama sent a $6.2 billion emergency funding request to Congress in November 2014, which was approved in December 2014. The text measure is within the positive region during this time. There is no discernable trend as the virus progressed further, as there were a variety of breakthroughs and setbacks, as well as the topic overall simply becoming less newsworthy.

However, while risk perceptions regarding Ebola becoming less relevant may be one explanation for the lack of a meaningful trend in later periods, another explanation may be that the method is failing to capture the sentiment. To ensure this is not the case, a random selection of negatively classified articles and positively classified titles have been gathered from later periods for verification.

Text: "A B.C. nurse practitioner who was being tested for Ebola will be reunited with her family today after health care workers confirmed she does not have the virus."
Publication Date: Jan 2, 2015
Pre-Weighted Score: 0.5 (Positive)
Verification: Positive

Text: "World: Amid uptick in Ebola cases, UN agency cites challenges in reaching affected communities"
Publication Date: Mar 2, 2015
Pre-Weighted Score: -0.222 (Negative)
Verification: Negative

Text: "American doctor cured of Ebola continues his work in Liberia."
Publication Date: Sep 24, 2015
Pre-Weighted Score: 0.429 (Positive)
Verification: Positive

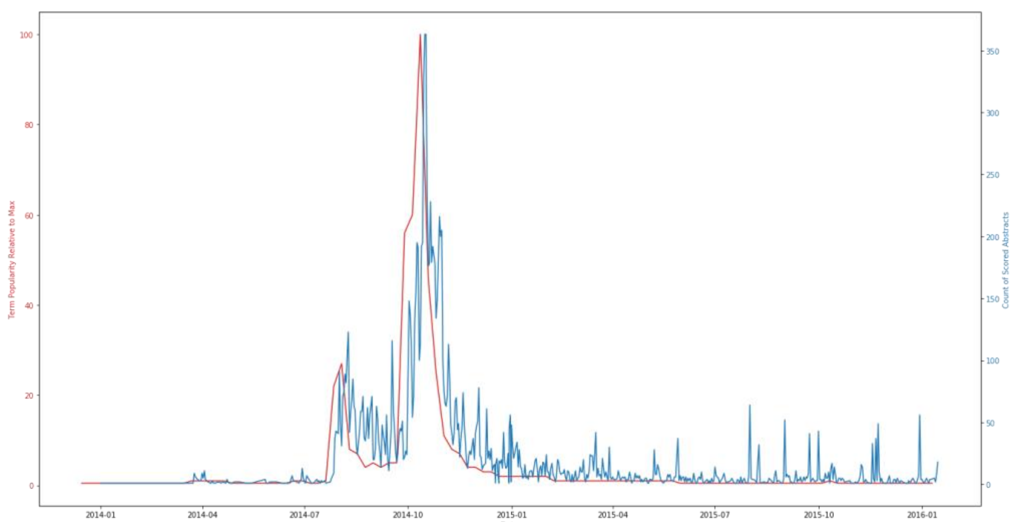
Text: "China will continue to support countries including Liberia in their reconstruction work as Ebola has begun to subside, and is willing to provide help within its ability for Liberia's socioeconomic development as well as the country's work in strengthening its public health system, he said."
Publication Date: Nov 6, 2015

Pre-Weighted Score: 0.273 (Positive)
Verification: Positive

Text: "The world is acutely aware of the fact that the people of Sierra Leone have been victimized by Ebola; the dreaded virus has affected the lives of every man, woman and child in that nation and killed over 4,000 of their countrymen. Ebola has left far too many children without parents, and the stories of grief and suffering amongst everyday people is the real reflection of Ebola's catastrophic impact in Sierra Leone, as well as in neighboring Liberia and Guinea."
Publication Date: Dec 13, 2015
Pre-Weighted Score: -0.227 (Negative)
Verification: Negative

Another observable measure against which to compare the results of the dictionary classifications is the Google Trend data for the search term 'Ebola'. As is to be expected, the search term popularity correlates strongly with the number of news articles ranked on a given day. The Google Trend data graphed against the daily count of scored abstracts is shown below (Google trend data is represented in red, and daily count of scored abstracts in blue):

Exhibit 4: Comparison: Search Term Frequency and Count of Articles Over Time



Given the high degree of correlation between the Google Trend data and the daily count of scored articles, the connection between the Google Trend data and the average score of titles and abstracts over time can be explained similarly. Public interest in searching for information about Ebola spikes around significant events, and the sentiment score of news data tends to respond accordingly. As an illustrating example, the peak of Google Trend interest occurred on October 12, 2014, the day on which the CDC announced a caregiver in the US had tested positive for Ebola. On that day, the average score was reasonably negative. Another significant day occurred on October 8, 2014, when the patient with the first case of Ebola in the US died. On that day, the average score was extremely negative.

Overall, comparing the scored data over time shows that the dictionary classifications result in sentiment scores that respond appropriately to world events and demonstrates appropriate differences in the magnitude of sentiment as well.

Company Level Data

Company level data has been gathered for 40 randomly selected companies in the S&P500 in 2014.

NAICS data used for the dummy variable representing sector fixed effects was obtained from Mergent Intellect.

Data used for the dummy variable representing business model exposure was obtained from company financial statements. For each company, statements from the time period of January 1, 2010 to February 1, 2015 were parsed for mentions of the following words: “Ebola”, “West Africa”, “Guinea”, “Sierra Leone”, and “Liberia”. If there were no mentions of these words during this time period, the company was categorized as having no business model exposure. If there were mentions of these words during this time period, each mention was further analyzed. If the company mentioned a specific business interest in the region or the specific way in which Ebola could affect its profits, then the company was categorized as having business model exposure. Examples include oil companies with offshore interests in the region and airlines or booking companies who would be harmed by a decrease in travel. If the company did not mention a specific business interest in the region or the specific way in which Ebola could affect its profits, then the company was categorized as having no business model exposure. A common example of this is a company mentioning Ebola as simply another entry in a long list of other tail end events such natural disasters and worker strikes in the ‘Risk Factors’ in an effort to be comprehensive. However, this does not indicate that the company in question is any more exposed to Ebola or the affected regions than the average globalized company in the index.

Data for the level variables of asset and revenue exposure were obtained from Bloomberg for the year 2014. If a company had no reported assets reported in

Bloomberg other than Plant, Property, and Equipment, it was assumed that these items constituted all relevant assets.

Data used for the level variable representing workforce exposure was obtained from company 10-K statements. However, it should be noted that there may be a slight degree of measurement error in this variable, as the disclosure of this data is voluntary. Disclosure of the variable occurs because of an exception to the SEC rules on employee pay disclosure. Briefly, companies are required to publish their average worker pay relative to the pay of their CEO; however, if they disclose the percentage of their workforce abroad, it can allow them to report more favorable income values to the public. It is highly likely that any company that does not disclose the percentage of their workforce aboard is doing so because the value is very low, and thus would not benefit their average worker income disclosures. As such, any company that does not disclose a value is highly likely to have little to no workforce exposure. Although these disclosure rules do introduce a small level of measurement error, it is mitigated by the use of tercile dummies, as it is almost certain that a company that does not disclose a value will fall in the lowest tercile.

Financial Market Index and Other Data

Financial market index data was obtained for the same time period as the news article data: December 15, 2013 to January 15, 2016.

Data for the Dow Jones Industrial Average, S&P500, and Russell 2000 was collected from Yahoo Finance.

Data for day of the week was obtained using the Pandas function Timestamp.weekday.

Data for the circulation of the top 100 newspapers magazines in the United States was obtained from the Audit Bureau of Circulation (now AAM) through data aggregator Infoplease. All publishers not within the top 100 of their respective categories were assumed to have a circulation of 5,000, as an estimate of an average between smaller regional publications and minute local publications. This value was estimated as an average of small market newspaper circulations in the United States using data from The Tow Center for Digital Journalism, adjusted downward given the prevalence of local and targeted newspapers found in the data.

Other Data Notes

An issue in the creation of the final dataset arises from the inherent difference in frequency between the news data and the financial market index data; news data occurs on all days of the week, while financial market index data only appears on weekdays. As such, the presence of news articles on weekends would likely impact epidemiological investor sentiment, but any effects would not be captured in the temporarily suspended financial markets.

This investigation has dealt with this issue by scoring the articles appearing on weekends separately, and then incorporating their influence into the next trading day. It would be inappropriate to simply drop these observations, as they likely have a significant impact in epidemiological investor sentiment. Therefore, each the articles appearing on each day of the weekend were scored separately, and the average was taken. The average score for the weekend was then averaged with the score for the next trading day. The intention of this methodology is to include the effect of the articles appearing on weekends on the next trading day, but to give these articles less weight in the average. The reasoning behind the decision that any article occurring on the next trading day is likely to be fresher in the minds of the investing public, and thus are likely to have a greater impact on their overall perceptions of the virus.

V. Results and Discussion

The market model was run on data from the S&P500 index from the time period of December 15, 2013 to January 15, 2016 and from the time period of December 15, 2013 to November 31, 2014. The results from both regressions are shown in the table below:

Table 2: Regression Results: Market Model

VARIABLES	(1) Model 1: Dec 15 2013 to Jan 15 2016	(2) Model 2: Dec 15 2013 to Nov 31 2014
Publisher Weighted Sentiment Score	3.97	7.10
	(7.90)	(6.50)
Lag of Logged Returns	0.03	-0.04
	(0.05)	(0.07)
Observations	472	188
R-squared	0.01	0.04
Weekday Fixed Effects	YES	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In the regression covering the time period of December 15, 2013 to January 15, 2016, the coefficient on the variable of interest possesses the expected sign but was not found to be statistically significant. This indicates that epidemiological investor sentiment did not have a significant effect on overall financial markets. The other variables were also not statistically significant, which is consistent with some past literature using particular market indices. Notably, Heyes et al. (2016) found the lag of returns to be statistically insignificant when studying the effect of air pollution on S&P500 index returns, and Gokmenoglu and Fazlollahi (2015) found the lag of logged returns to be statistically insignificant in the short run when studying the effect of gold and oil prices and volatility on S&P500 index returns. Levy (2015) also found a statistically insignificant effect for the lag of logged returns for the S&P500 index.

In an earlier section, the lack of a meaningful trend later in the time period was discussed. The rough period from December 15, 2013 to November 31, 2014 had noticeable trends that aligned with relevant current events, however later dates had no such trend and were quite volatile due to the lack of a public narrative in the news as the risk of Ebola as a substantial threat essentially left the markets' collective consciousness. To test whether these apparent trends had an effect on the results of the model, the same process was used while restricting the data to only use observations from the time period of December 15, 2013 to November 31, 2014.

The results of this regression are similar to the first; both the variable of interest and the control variables lack statistical significance. This again appears to indicate that epidemiological investor sentiment did not have a significant effect on financial markets, even in time periods with apparent trends in the average score of articles that correspond with relevant events.

There are several reasons the variable of interest may have been found to be insignificant, but it is likely because Ebola had a marginal effect on US markets overall and that market participants were not acutely worried about systematic market issues resulting from its potential spread. Overall, Ebola was simply not a major concern to the markets for much of the crisis.

The company model was run on data from 40 S&P500 companies across a variety of industries from the time period of December 15, 2013 to January 15, 2016. The results from the panel regression are shown in the table below:

Table 3: Regression Results: Company Model

VARIABLES	(1) Base Model
Lag of Logged Company Returns	0.02**
	(0.01)
Publisher Weighted Sentiment Score	0.12***
	(0.02)
(Publisher Weighted Sentiment Score)^2	3.41***
	(0.97)
Observations	18,800
Number of id	40
R-squared	0.00
Weekday Fixed Effects	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The coefficient on the variable of interest possesses the expected sign and is found to be highly statistically significant. A non-linear term of the sentiment variable is also found to be statistically significant, along with the lag of company returns. This indicates that epidemiological investor sentiment has a significant, non-linear relationship with company returns when controlling for company-specific fixed effects. The coefficients on the weekday dummy variables are consistent with Berument and Kiymaz (2001), specifying the lowest returns on Mondays and the highest returns on Wednesdays. All weekday dummy variables were statistically significant at conventional levels except for the Friday variable.

According to the results of the model, a one standard deviation increase in the Publisher Weighted Sentiment Score would, on average, increase a company's rate of return by 0.05 standard deviations. The largest effect on an individual company was an increase in rate of return by 0.078 standard deviations and the smallest effect was increase in rate of return by 0.021 standard deviations. However, as expected, this is primarily because these companies have the least and most volatile returns over the relevant time period, respectively.

To attempt to identify the company specific characteristics that may influence the relationship between sentiment and returns, a variation of the base model was run with the individual introduction of several different exposure variables. The exposure variables used were asset exposure, revenue exposure, business model exposure, and workforce exposure. Asset exposure was a level variable measuring the percentage of a company's assets located abroad. Revenue exposure was a level variable measuring percentage of a company's revenue earned abroad. Business model exposure was a dummy variable measuring a company's business model exposure to viral disruptions and at-risk regions. Finally, workforce exposure was a dummy variable measuring the tercile of a company's percentage of workforce located abroad relative to other companies in the data. The coefficients on the interactions between exposure variables and the sentiment variable can be interpreted as the additional effects from sentiment that the market places upon companies with the corresponding type of exposure. The results from the panel regression are shown in the table below:

Table 4: Regression Results: Company Model with Company Exposure Variables

VARIABLES	(1) Base Model + Company Control: Asset Exposure	(2) Base Model + Company Control: Revenue Exposure	(3) Base Model + Company Control: Workforce Exposure	(4) Base Model + Company Control: Business Model Exposure
Lag of Logged Company Returns	0.02**	0.02**	0.02**	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)
Publisher Weighted Sentiment Score	0.08***	0.11***	0.11***	0.13***
	(0.03)	(0.03)	(0.03)	(0.02)
(Publisher Weighted Sentiment Score)^2	3.41***	3.41***	3.41***	3.41***
	(0.97)	(0.97)	(0.97)	(0.97)
AssetExposure* Publisher Weighted Sentiment Score	0.13			
	(0.10)			
RevenueExposure* Publisher Weighted Sentiment Score		0.11		
		(0.09)		
WorkforceExposure*Publisher Weighted Sentiment Score			0.07	
			(0.07)	
BusinessModelExposure* Publisher Weighted Sentiment Score				-0.02
				(0.07)
Observations	18,800	18,800	18,800	18,800
R-squared	0.00	0.00	0.00	0.00
Number of id	40	40	40	40
Weekday Fixed Effects	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results of this regression indicate that these specific company characteristics do not have a statistically significant effect on the relationship between epidemiological sentiment and a company's returns. The coefficients and statistical significance on the base model variables remain similar.

Given that the fixed effects model controlled for unobserved company specific factors, a variation of the base model was run with the addition of a variable controlling for sector specific factors that may make a given sector more vulnerable to the effects of news about Ebola.

For this model variation, all industries in the panel dataset were classified as either having "Low", "Neutral", or "High" sensitivity to news about Ebola. Low sensitivity industries were those such as Real Estate and Rental and Leasing (NAICS 53) or Information (NAICS 51), which are either domestically focused or possess a product for which demand would not slow if Ebola were to spread. High sensitivity industries were those such as Mining, Quarrying, and Oil and Gas Extraction (NAICS 21) and Transportation and Warehousing (NAICS 48), which exhibit dependance on prevalent natural resources in affected areas, international travel and accommodation, or global shipping. Neutral sensitivity sectors are those that do not fall into either category. The full list of industry classifications can be found in the Appendix.

The dummy variables for Low and High sector sensitivity for a company's NAICS code was interacted with the sentiment variable to represent the effect of news about Ebola on a company given the primary sector they operate within. The results from the panel regression are shown in the table below:

Table 5: Regression Results: Company Model with Industry Fixed Effects

VARIABLES	(1) Base Model + Industry Sensitivity to Ebola
Lag of Logged Company Returns	0.02**
	(0.01)
Publisher Weighted Sentiment Score	0.13***
	(0.03)
(Publisher Weighted Sentiment Score)^2	3.41***
	(0.97)
LOW Sensitivity NAICS Code* Publisher Weighted Sentiment Score	-0.03
	(0.05)
HIGH Sensitivity NAICS Code* Publisher Weighted Sentiment Score	0.01
	(0.06)
Observations	18,800
Number of id	40
R-squared	0.00
Weekday Fixed Effects	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results of this regression indicate that a company's primary sector does not have a statistically significant effect on the relationship between epidemiological sentiment and a company's returns. The coefficients and statistical significance on the base model variables remain similar.

A final variation was run to test the timing of the relationship between epidemiological sentiment and market returns, or whether investors require time to incorporate sentiment into their actions. The results of this regression found lags of sentiment to be statistically insignificant, thus indicating that its relationship with company returns is current.

The findings from these regression models indicate that epidemiological investor sentiment has a statistically significant, current, non-linear effect on a company's stock returns when controlling for company-specific fixed effects. However, the characteristics of asset exposure, revenue exposure, workforce exposure, and business model exposure do not appear to influence the effect of sentiment. The significance of sentiment is not found at the overall market level.

VII. Conclusion

This paper has explored the effect of epidemiological investor sentiment resulting from the outbreak of the Ebola virus from December 15, 2013 to January 15, 2016. A novel sentiment dictionary of meaningful words commonly used in conjunction with the word Ebola was created by surveying epidemiologists. Sentiment then was quantified with a text processing method established by Hassan et al. (2020), using news on viral events uncorrelated with other potential causes of market movements as data points and incorporating weights based upon publisher circulation.

Results indicate that epidemiological investor sentiment did not have a statistically significant effect on financial markets when examining broad market indices. This is likely due to the marginal effect Ebola had on US markets overall, and that market participants were not acutely worried about systematic failures from it spreading. However, epidemiological investor sentiment was found to have a highly statistically significant, current, and non-linear effect on a company's stock returns when controlling for company-specific fixed effects. This is consistent with the work of Hassan et al. (2019), who found that text-based risk measures can capture unique risk exposures that are not present at the aggregate level.

As for further extensions, it is likely that significant findings can be achieved by examining heterogenous effects between positive and negative sentiment and exploring other text-based measures such as the use of significant phrases or groups of words called "bi-grams" that have been utilized in Hassan et al. (2020). Once additional findings have been obtained, the question of whether this body of research can be generalized in a predictive manner for other viral events can be more conclusively addressed. The intention of this paper is for the specialized sentiment dictionaries to provide future researchers with novel tools to continue work on this increasingly relevant issue.

References

"2014-2016 Ebola Outbreak in West Africa," March 8, 2019. https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html.

Barberis, Nicholas, and Richard Thaler. "A Survey of Behavioral Finance." NBER Working Papers from National Bureau of Economic Research, Inc, 2003. https://doi.org/10.3386/w9222.

Barker, L., and F. Bacon. "THE EBOLA OUTBREAK: A TEST OF MARKET EFFICIENCY." In Academy of Accounting and Financial Studies 20 (2015): 2–6.

Bauer, Lauren, Kristen E. Broady, Wendy Edelberg, and Jimmy O'Donnell. "Ten Facts about COVID-19 and the U.S. Economy," September 18, 2020. https://www.brookings.edu/research/ten-facts-about-covid-19-and-the-u-s-economy/.

Berument, Hakan, and Halil Kiymaz. "The Day of the Week Effect on Stock Market Volatility." Journal of Economics and Finance 25, no. 2 (2001): 181–93. https://doi.org/10.1007/bf02744521.

Davis, Rebecca. "Ebola Epidemic 2014: Timeline," October 15, 2014. https://www.theguardian.com/world/2014/oct/15/ebola-epidemic-2014-timeline.

"Ebola Virus Disease." Accessed December 3, 2020. https://www.afro.who.int/health-topics/ebola-virus-disease.

Gokmenoglu, Korhan K., and Negar Fazlollahi. "The Interactions among Gold, Oil, and Stock Market: Evidence from S&P500." Procedia Economics and Finance 25 (2015): 478–88. https://doi.org/10.1016/s2212-5671(15)00760-1.

Hassan, Tarek Alexander, Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. "Firm-level Exposure to Epidemic Diseases: COVID-19, SARS, and H1N1." (June 1, 2020). http://dx.doi.org/10.2139/ssrn.3566530

Hassan, Tarek Alexander, Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. "Firm-Level Political Risk: Measurement and Effects." The

Quarterly Journal of Economics 134, no. 4 (November 2019): 2135–2202. https://doi.org/https://doi.org/10.1093/qje/qjz021.

Heyes, Anthony, Matthew Neidell, and Soodeh Saberian. “The Effect of Air Pollution on Investor Behavior: Evidence from the S&P 500,” 2016. https://doi.org/10.3386/w22753.

Ichev, Riste and Marinc, Matej. “Stock Prices and Geographic Proximity of Information: Evidence from the Ebola Outbreak.” International Review of Financial Analysis, Forthcoming (December 13, 2017). https://ssrn.com/abstract=3087093

Kates, Jennifer, Josh Michaud, Adam Wexler, and Allison Valentine. “The U.S. Response to Ebola: Status of the FY2015 Emergency Ebola Appropriation,” March 13, 2019. https://www.kff.org/global-health-policy/issue-brief/the-u-s-response-to-ebola-status-of-the-fy2015-emergency-ebola-appropriation/.

Ko, K. Jeremy, and Zhijian (James) Huang. “Persistence of Beliefs in an Investment Experiment.” *Quarterly Journal of Finance* 02, no. 01 (2012): 1250005. https://doi.org/10.1142/s201013921250005x.

Kochhar, Rakesh. “Unemployment Rose Higher in Three Months of COVID-19 than It Did in Two Years of the Great Recession,” August 26, 2020. https://www.pewresearch.org/fact-tank/2020/06/11/unemployment-rose-higher-in-three-months-of-covid-19-than-it-did-in-two-years-of-the-great-recession/.

Levy, Nir. “The Effect of New York City Sports Outcomes on the Stock Market.” Undergraduate Economic Review 12, no. 1 (2015).

Lord, Charles G., Lee Ross, and Mark R. Lepper. “Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence.” *Journal of Personality and Social Psychology* 37, no. 11 (1979): 2098–2109. https://doi.org/10.1037/0022-3514.37.11.2098.

Loughran, Tim, and Bill McDonald. “Textual Analysis in Accounting and Finance: A Survey.” Journal of Accounting Research 54, no. 4 (September 2016): 1187–1230. https://doi.org/https://doi-org.proxy-um.researchport.umd.edu/10.1111/1475-679X.12123.

Loughran, Tim, and Bill McDonald. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." The Journal of Finance 66, no. 1 (2011): 35-65.

Menton, Jessica, and Nathan Bomey. "Dow Endures Worst Day since 'Black Monday'; S&P 500 Enters Bear Market as Coronavirus Spreads Economic Gloom," March 12, 2020. https://www.usatoday.com/story/money/2020/03/12/dow-plunges-trump-speech-fails-quell-coronavirus-fears/5029964002/.

Ramelli, Stefano, and Alexander F Wagner. "Feverish Stock Price Reactions to COVID-19*." The Review of Corporate Finance Studies 9, no. 3 (2020): 622–55. https://doi.org/10.1093/rcfs/cfaa012.

Schraeder, Stefanie. "Information Processing and Non-Bayesian Learning in Financial Markets." Swiss Finance Institute Research Paper No. 14-14 (2020)

Steelfisher, Gillian K., Robert J. Blendon, and Narayani Lasala-Blanco. "Ebola in the United States — Public Reactions and Implications." New England Journal of Medicine 373, no. 9 (2015): 789–91. https://doi.org/10.1056/nejmp1506290.

Tversky, A. and D. Kahneman. "Judgment under uncertainty: Heuristics and biases." Science 185, no. 4157 (1974): 1124-31. https://doi.org/10.1126/science.185.4157.1124

Appendix

Item 1

List of negations that had appearances removed from the news data before analysis.

Note: punctuation has been removed to be consistent with the format of the cleaned data.

"no", "not", "none", "nothing", "neither", "never", "doesnt", "isnt", "wasnt", "shouldnt", "wouldnt", "couldnt", "wont", "cant", "dont"

Item 2

Detailed Process to Obtain 150 Words for Dictionary

1. Iterate over all cleaned news article titles and abstracts without negations present for occurrences of the word ‘Ebola’. Whenever an instance is found, capture the 10-word neighborhood around that instance (5 words before the instance, and 5 words after the instance). If the instance occurs within 5 words of the beginning or end of the text, the neighborhood is shortened accordingly.
2. Place every captured neighborhood into a list, such that each title and abstract has an individual list containing the neighborhoods associated with it.
3. Join all individual lists into a master list that contains all neighborhoods within the entire dataset.
4. Select the 150 most commonly appearing words within the master list.
5. From the initial list, remove all words with an obvious sentiment. This is defined as having a sentiment classification in either the Harvard IV-4 or Loughran and Mc Donald Financial Sentiment Dictionaries approved for use in an epidemiological context by a human audit. Additionally, remove all meaningless words such as names, locations, and times.
6. For every obvious and meaningless word removed, replace that word with the next most commonly appearing word in the master list (e.g., When the first obvious word was removed, it was replaced with the 151^{st} most commonly appearing word within the master list).
7. Repeat Steps 5-6 until the final list has no obvious or meaningless words.

Item 3 - Complete sentiment dictionary (Root Majority Version).

admitted	-1	**fear**	-1	**monkeys**	-1	**screenings**	1
affected	-1	**fears**	-1	**nih**	1	**severe**	-1
aid	1	**federal**	1	**officials**	1	**spread**	-1
airport	1	**fight**	1	**organisation**	1	**spreading**	-1
airports	-1	**flight**	-1	**passengers**	-1	**spreads**	-1
alert	-1	**flu**	-1	**patient**	-1	**staff**	1
battle	-1	**fluids**	-1	**patients**	-1	**team**	1
beds	-1	**foreign**	-1	**political**	-1	**test**	1
blood	1	**free**	1	**precautions**	1	**testing**	1
bodily	-1	**funding**	1	**prepared**	1	**tests**	1
body	-1	**gear**	1	**preparedness**	1	**toll**	-1
case	-1	**government**	1	**preparing**	1	**transmission**	-1
cases	-1	**guidelines**	1	**president**	-1	**transmitted**	-1
cdc	1	**handle**	1	**prevent**	1	**travel**	-1
centers	1	**handling**	1	**prevention**	1	**travelers**	-1
clinic	1	**healthcare**	1	**procedures**	1	**traveling**	1
clinical	1	**hemorrhagic**	-1	**professionals**	1	**treat**	1
contact	1	**home**	1	**protocol**	1	**treated**	1
control	1	**hospital**	1	**protocols**	1	**treating**	1
critical	-1	**hospitals**	1	**public**	1	**treatment**	1
deadliest	-1	**incubation**	-1	**quarantine**	-1	**treatments**	1
diagnosed	-1	**infectious**	-1	**quarantined**	-1	**trial**	1
diagnosis	-1	**international**	-1	**quarantines**	1	**trials**	1
doctor	1	**isolated**	-1	**readiness**	1	**troops**	-1
doctors	1	**isolation**	-1	**recovered**	1	**un**	1
dr	1	**issues**	-1	**remains**	-1	**united**	1
drug	1	**lab**	1	**research**	1	**vaccine**	1
drugs	1	**leaders**	1	**researchers**	1	**vaccines**	1
economic	1	**life**	1	**response**	1	**viral**	-1
effort	1	**malaria**	-1	**restrictions**	-1	**virus**	-1
efforts	1	**media**	-1	**returned**	-1	**viruses**	-1
evacuated	-1	**medical**	1	**returning**	-1	**volunteers**	1
evd	-1	**medicine**	1	**risk**	-1	**worker**	1
experimental	1	**milestone**	1	**risks**	-1	**workers**	1
exposure	-1	**military**	-1	**samples**	1	**working**	1
facilities	1	**monitor**	1	**scientists**	1	**world**	-1
facility	1	**monitored**	1	**screen**	1		
family	1	**monitoring**	1	**screening**	1		

Item 4

Comparison of Methodologies to Rectify Dictionary Inconsistencies

VARIABLES	(1) Dictionary Determination 1: Root Majority	(2) Dictionary Determination 2: Inconsistencies Removed	(3) Dictionary Determination 3: Inconsistencies Averaged
Lag of Logged Company Returns	0.02**	0.02**	0.02**
	(0.01)	(0.01)	(0.01)
Publisher Weighted Sentiment Score	0.12***	0.12***	0.12***
	(0.02)	(0.02)	(0.02)
(Publisher Weighted Sentiment Score)^2	3.41***	3.95***	3.11***
	(0.97)	(1.09)	(1.12)
Observations	18,800	18,800	18,800
R-squared	0.00	0.00	0.00
Number of id	40	40	40
Weekday Fixed Effects	YES	YES	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Item 5

List of All Industry Classifications for Sensitivity to Ebola

High Sensitivity to Ebola:

- Mining, Quarrying, and Oil and Gas Extraction (21)
- Paper, Petroleum, Chemical, Plactics, Minerals Manufacturing (32)
- Transportation and Warehousing (48)
- Accommodation and Food Services (72)

Neutral Sensitivity to Ebola:

- Food, Beverage, Textile, Apparel Manufacturing (31)
- Metal, Machinery, Computer and Electronic, Transportation, Furniture, Misc. Manufacturing (33)
- Finance and Insurance (52)

- Administrative and Support and Waste Management and Remediation Services (56)

Low Sensitivity to Ebola:

- Wholesale Trade (42)
- Retail Trade (44)
- Retail Trade (45)
- Information (51)
- Real Estate and Rental and Leasing (53)

Item 6

Results of Company Base Model, including Lags of Sentiment

VARIABLES	(1) Base Model + Lags of Sentiment
Lag of Logged Company Returns	0.02*
	(0.01)
Publisher Weighted Sentiment Score	0.12***
	(0.02)
(Publisher Weighted Sentiment Score)^2	3.45***
	(0.97)
Lag 1: Publisher Weighted Sentiment Score	-0.03
	(0.03)
Lag 2: Publisher Weighted Sentiment Score	-0.05**
	(0.02)
Observations	18,720
Number of id	40
R-squared	0.00
Weekday Fixed Effects	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Item 7
Discussion of Market Model Assumptions

When considering the robustness of the market model findings and whether they can be generalized to study other viral events, it is important to address the assumption of exogeneity. Concerns originate from the fact that stock prices are meant to be the net present value of a firm's future cash flows. If a pandemic were to disrupt markets and reduce those cash flows, then decreases in the price of a stock or market index in response to news may not be due to epidemiological investor sentiment, but rather rational calculations of decreased future profitability. For variation in market indices to be attributed to epidemiological investor sentiment, it must be shown that the Ebola virus did not result in other determinant effects on US companies, the exclusion of which would result in omitted variable bias and the incorrect attribution of their effects to sentiment.

It is simple to show that Ebola had no other significant determinant effects on US companies other than through sentiment. The countries of Guinea, Liberia, and Sierra Leone were by far the hardest hit by the virus, together accounting for more than 99.8% of all confirmed cases. In 2014, trade flows with these countries made up only 0.021% of US exports and 0.008% of US imports. In 2016, when the epidemic was declared to be over, trade flows with these countries made up 0.021% of US exports and 0.005% of US imports. In fact, only four companies reporting to the SEC mentioned the names of these countries within the relevant time period; Chevron, Royal Caribbean, and Firestone mentioned Liberia, and Anadarko mentioned Sierra Leone. This lack of significant trade and business interests causes a virus contained within these countries to be of little concern for the future of US companies.

However, this is not enough. As stated earlier, stock prices are theoretically derived from future cash flows, and thus an expectation of the virus spreading and disrupting the global value chains of US companies would be reflected in current prices. For the Ebola outbreak to be truly exogenous to market returns, it must also be shown that market participants did not rationally calculate their risk perceptions based upon the objective spread risk of the virus to the extent it could affect US business fundamentals. This is a much more difficult task, requiring evidence that shows the evolving perceptions of various market participants.

For this, we begin with the perception of influential investors. These perceptions are difficult to quantify, and as such, their views are best explained in these investors' own words:

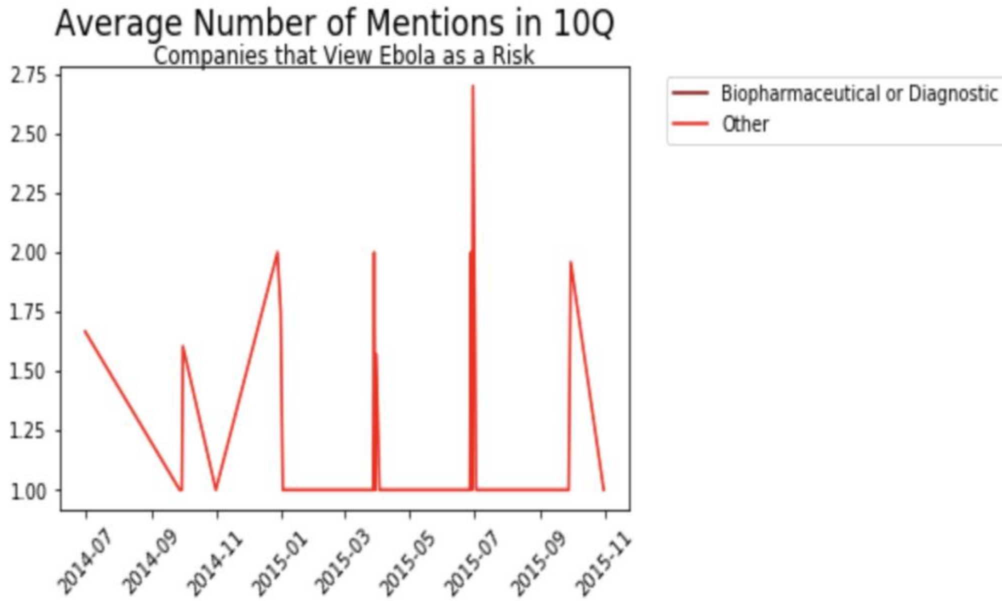
> "The Ebola scare is not having an impact on many businesses at this point. For that to happen, the deadly virus would have to spread much more rapidly in the

> West. That means earnings -- the fundamental component of stock prices -- aren't being hit."
> - Matt Egan, Lead Writer, CNN Business
>
> "The stock market is now driven by emotion rather than fundamentals. The former rapidly switches from greed to fear on Ebola news,"
> - Ed Yardeni, President of Investment Advisory, Yardeni Research
>
> "The market is not trading off of fundamental news, it is trading off worries, whether it's the importing of weak economic growth or Ebola,"
> - Mark Luschini, Chief Investment Strategist, Janney Montgomery Scott

This is certainly insufficient for proving that markets did not believe Ebola would spread, but when combined with quantifiable perceptions of other market participants, provides for a robust argument.

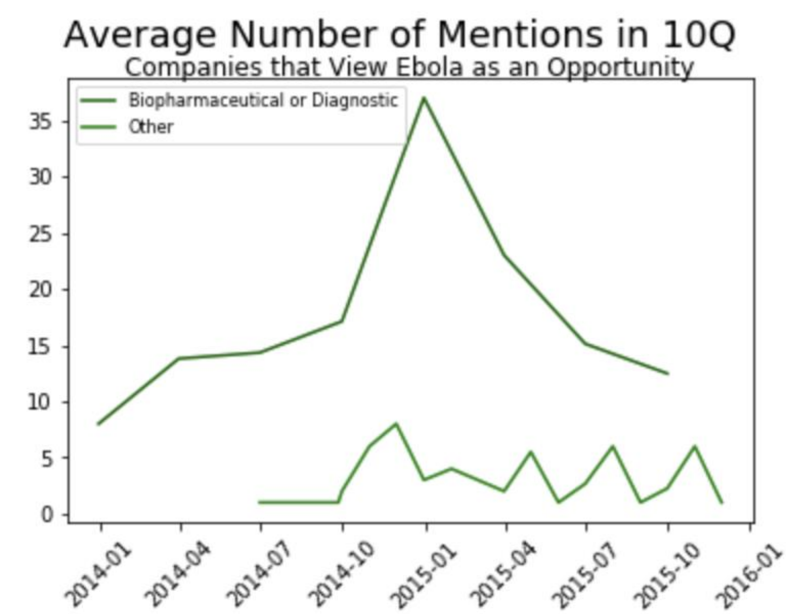
Moving on to major companies, their perceptions are easier to quantify. Any company meeting certain thresholds and other requirements are tasked by the SEC to publish what they believe to be the major risks to their continuing operations to the public. Searching the SEC database for company 10Q reports reveals that there are 109 unique companies that mentioned the word 'Ebola' across 259 different reports from the period of December 15, 2013 to January 15, 2016. This data was gathered, and each report was categorized as either describing Ebola as a risk or an opportunity upon which the company could capitalize. 134 of the reports viewed Ebola as a risk, and 125 viewed it as an opportunity. The majority of companies who viewed Ebola as a risk mentioned the term 'Ebola' only once in their report, and in general, it was simply stated in a long list of other tail-end events such as "earthquakes" and "tsunamis" that were included in an effort to be comprehensive. A graph of the average number of mentions from companies that view Ebola as a risk can be seen below:

Appendix Exhibit 1:



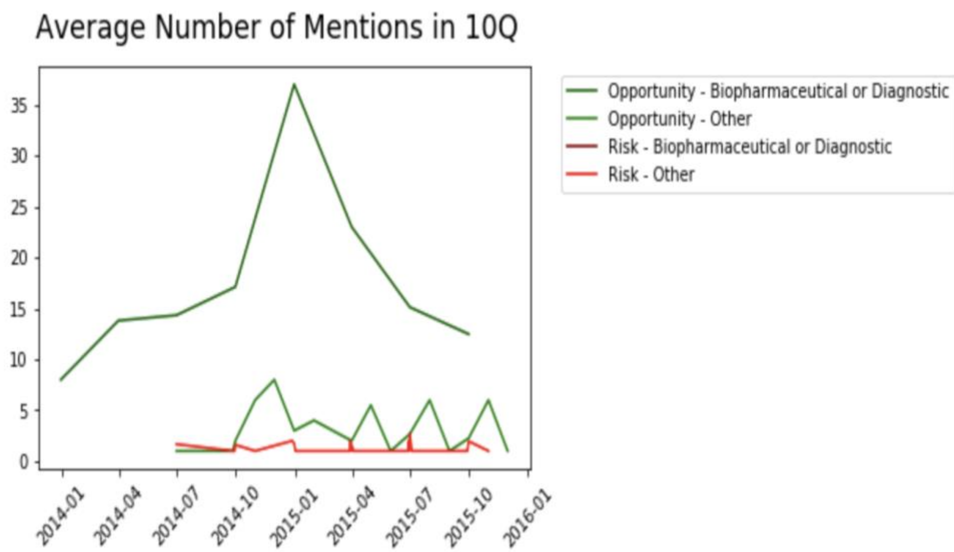
The vast majority of companies who viewed Ebola as an opportunity were either in the Biopharmaceutical or Diagnostic industries. A graph of the average number of mentions from companies that view Ebola as an opportunity can be seen below:

Appendix Exhibit 2:



Plotting the average number of mentions on the same graph reveals that companies viewing Ebola as an opportunity tended to mention it far more times than those companies viewing it as a risk. The graph can be seen below:

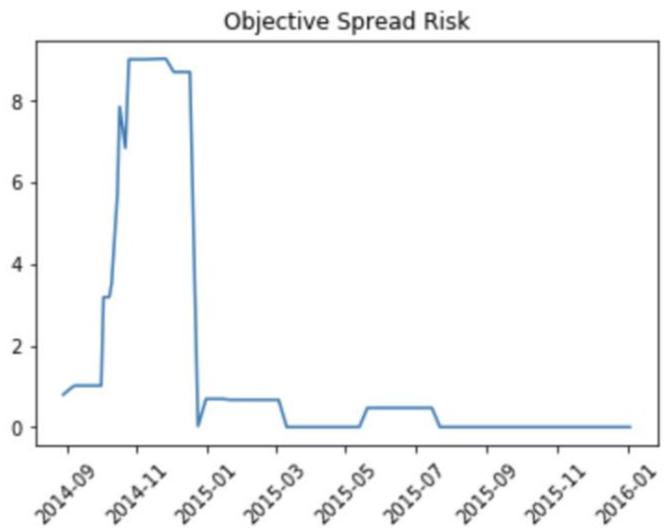
Appendix Exhibit 3:



This demonstrates that companies mentioning Ebola as an opportunity devote far more time to discussing it, and thus that they consider it far more important to their future operations. Companies that mention Ebola as a risk typically only do it in passing.

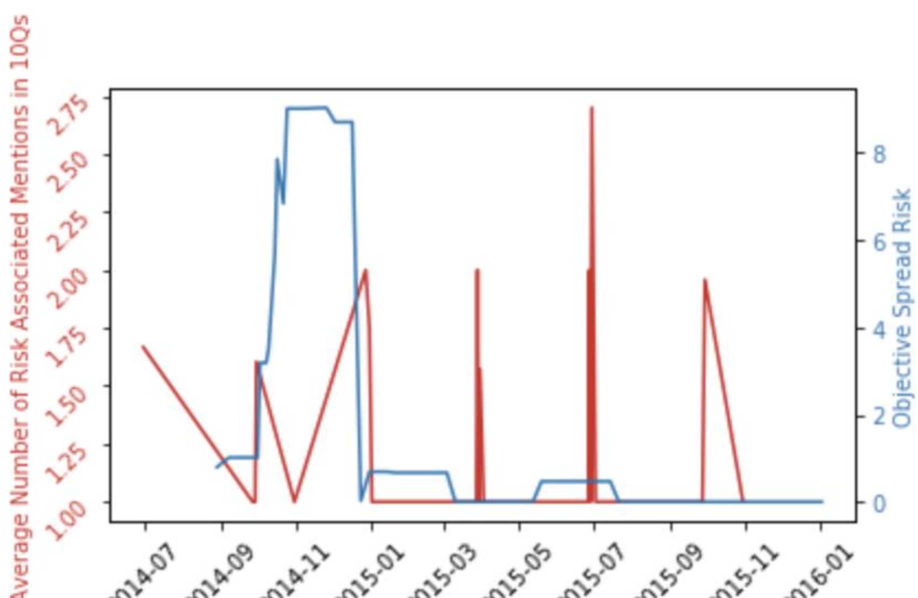
The exogeneity of Ebola to market returns is further demonstrated by comparing epidemiological risk perceptions of major companies against the objective spread risk. Objective spread risk was calculated as a custom metric derived from Ebola situation reports published by the World Health Organization. Roughly every week of the outbreak, the WHO published a report detailing the progression of the virus within Guinea, Liberia, and Sierra Leone, as well as "Countries with an Initial Case or Cases, or with Localized Transmission". The number of cases within these additional countries approximately represents the probability of the virus spreading into new regions. The number of confirmed cases was multiplied by the value of global exports of a country as a means of weighting based upon the extent to which the country is connected to global markets. The custom metric was then created by summing this product for every country in this category with confirmed cases on each day of a situation report. The graph of objective spread risk is shown below:

Appendix Exhibit 4:



The graph comparing the objective spread risk of Ebola to epidemiological risk perceptions of major companies demonstrated by the average number of risk associated mentions of the term 'Ebola' in 10Qs is shown below:

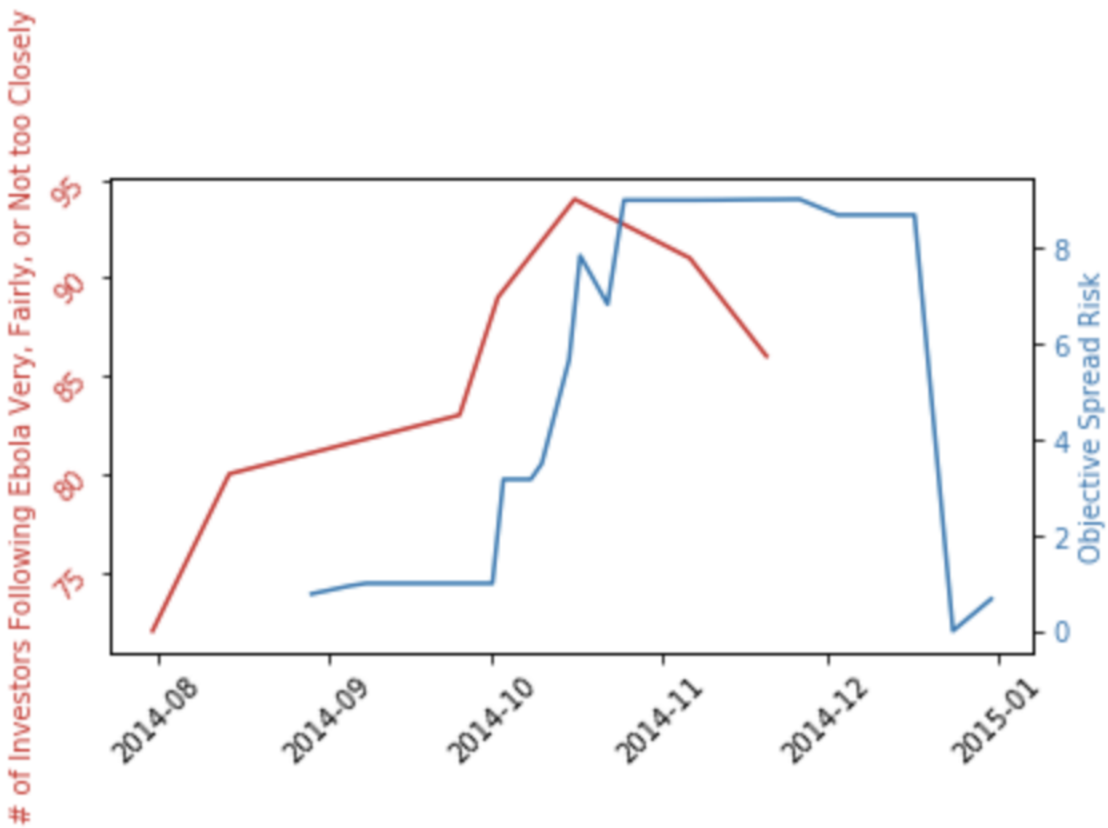
Appendix Exhibit 5:
Average Number of Risk Related Mentions in Company 10Qs and Objective Spread Risk, Compared



There is little positive correlation between the two variables, with a significant dip in risk associated mentions at the peak of objective spread risk, and significant peaks in risk associated mentions with not corresponding rise in spread risk. The lack of correlation shows that company perception of epidemiological risk is not rational in the sense that it was not produced by objective spread risk, demonstrating the exogeneity of Ebola to market expectations.

The last market participant to discuss is the average investor, whose epidemiological risk perceptions can be quantified through a survey conducted by Pew Research Center over the period of July 7, 2014 to November 20, 2014. In this survey, participants were asked a total of 7 times throughout the period how closely they were following the outbreak of the Ebola virus in Africa. The graph of the proportion of respondents following the outbreak of Ebola compared with the measure of objective spread risk is shown below:

Appendix Exhibit 6:
Investor Epidemiological Risk Perceptions and Objective Spread Risk, Compared



Although there is some positive correlation between the positive variables, it is not strong; investor perception of epidemiological risk begins rising while the objective spread risk remains flat, peaks before objective spread risk peaks, and then drops rapidly while objective spread risk remains high. In addition, there is no economic logic for the movements of investor perception of epidemiological risk responding before objective spread risk rises. The relationship between which variable responds first should be reversed, further weakening any interpretation of a positive correlation. Again, the lack of meaningful correlation shows that investor perception of epidemiological risk is not rational in the sense that it was produced by objective spread risk, demonstrating the exogeneity of Ebola to market expectations.

The combination of evidence demonstrating that market participants either did not expect Ebola to spread to the extent that it would affect US business fundamentals and that epidemiological risk perceptions were not based upon the virus' objective spread risk justifies the assumption that the Ebola virus itself was exogenous to market returns.

The incorrect attribution of omitted variable effects to sentiment is less of a concern for the company model, as the ways in which Ebola could directly affect a firm other than through sentiment are controlled for. While the relevant characteristics do not appear to be the ones tested in this investigation, the model controls for the unobserved factors.