Does Perception Matter in Asset Pricing? Modeling Volatility Jumps and Returns Using Twitter-Based Sentiment Indices*

Anthony Sanford†

May 18, 2018

Abstract

Do consumer and investor perceptions matter in asset pricing? I find that it is possible to forecast high-frequency stock returns and volatility jumps using consumer and investor sentiment indicators. Using tweets that I scraped from Twitter, I perform textual analysis to construct daily sentiment indices. While other scholars have relied on third-party companies like Stocktwits to complete these tasks, doing so reduces transparency and limits the potential for customization. The sentiment indices I constructed are numerical scores, not dichotomous variables, which allows me to control for sentiment strength (e.g., good vs. great) and not just positive/negative overall feelings. Results indicate that sentiment indices can not only be used to obtain out-of-sample forecasts of daily returns, but can also forecast volatility jumps. Using a simple Markov-switching framework, I find that, as overall sentiments shift from positive to negative (or vice versa), volatility jumps occur.

Keywords: Behavioral finance, volatility jumps, Twitter, social media, stock returns.

JEL Code: G40, G41, G14, G17

*I am grateful for helpful comments from Eric Zivot, Joannie Tremblay-Boire, Pushpak Sarkar, Thomas Fillebeen, and seminar participants at the University of Washington, Amazon, and CIREQ. Financial support for this research was provided by the Fonds de recherche du Québec – Société et culture (FRQSC). Computing resources were provided by the University of Washington's Center for Studies in Demography and Ecology (CSDE). All errors are my own.

†Department of Economics, University of Washington. Email: sanfoan@uw.edu

1 Introduction

Over the last several years, a growing number of news articles have claimed that investors use social media to help them guide their investment decisions (Huang, 2015; Openshaw, 2013; Kollmeyer, 2016; Borzykowski, 2016). In general, investors are quick to dismiss disinformation. As such, if more investors are turning to social media to make investment decisions, it must be the case that social media contains some valuable information. This article, focusing on Twitter specifically, asks: do consumer and investor perceptions matter in asset pricing? In his now seminal 1988 presidential address, Roll (1988) showed argues news, as defined by the Dow-Jones News Retrieval System, had little to no relationship with stock returns. Although news itself may not impact stock returns directly, the way we feel, react, interpret, and/or perceive new sources of information may in fact affect stock returns (Boudoukh et al., 2013). For instance, Baker and Wurgler (2006) (and to some extent, Cen et al. (2013)) find that sentiments affect a cross-section of asset prices. Other scholars have demonstrated that individual opinions posted on Twitter can predict firm earnings and announcements (Bartov et al., 2015) and that investor sentiments from stock-related message boards can forecast Amazon stock returns (Das and Chen, 2007). This article presents a somewhat different argument: my position is that tweets can actually alter investors' information set and thus provide us with an out-of-sample forecast of a stock's returns. As such, this work expands the literature mapping linkages between the sentiment consensus of consumers/investors and firm asset returns (Hribar and McInnis (2012); Da et al. (2013)).

I obtain "perceptions" by converting tweets scraped from Twitter into quantifiable signals about investor sentiments towards a product or company. Assume, for example, that consumers are deciding whether to buy the most recent smartphone from some company. If the smartphone has potential issues, such as bugs or annoying features, others' perceptions about the product should impact consumers' decision to purchase it. For example, when Apple decided to remove the headphone socket from its phones, consumers were outraged. Yet, as they began adopting the new smartphone, consumers started to report that the change to the phone was not a deal breaker. Sales of Apple's new phone started more slowly than for previous releases, but they picked up and were eventually in line with expectations (Dunn, 2017). Although this is merely one example, it shows how social media can be a source of information when deciding whether or not to adopt a new product or technology. In a more recent example, Bloomberg (Vasquez, 2018) reported that a tweet by Kylie Jenner had caused a 6.1% decline in the stock of Snapchat after she claimed to no longer be using the smartphone application. The company's stock plummeted from a close of $18.64 on the day before Jenner's tweet to a close of $16.32 a few days later (the day of the tweet, the stock closed at $17.51). More than a month later, Snapchat stocks had still not recovered fully from the Kylie Jenner tweet. Although this example is anecdotal, it suggests that Twitter user sentiments (especially from "influential" users) may have a significant effect on asset prices, irrespective of the fundamental valuation of the firm. This phenomenon is precisely what this article explores.

Since Twitter is primarily driven by information that is not necessarily based on fundamentals,[1] asset pricing theory would suggest that consumer sentiments would merely contribute to price noise, if they contribute at all (Roll, 1988).[2] As this paper will show, this should be reflected in our ability to forecast these short-term returns fluctuations if we assume that these fluctuations are driven by short-term fears or other human emotions.[3] Currently, very little scholarly work attempts to model short-term return fluctuations using high-frequency financial market data. Most research argues that noise is something that cannot be estimated (Cochrane, 2009; Campbell and Hentschel, 1992). I argue, in this paper, that short-term fluctuations in asset prices can be modeled. Figure 1 shows our current ability (based on the three best/most used

[1] In fact, I would argue that the majority of tweets that are made by the general public (not investment professionals) with respect to a specific company are based on spur of the moment feelings, not firm fundamentals.

[2] What is meant by noise is a stock's deviations from the fundamentally valued price.

[3] Note that throughout this paper, short-term fluctuations and noise are used interchangeably.

models) to forecast asset prices or returns at various time horizons (Sanford, 2017):

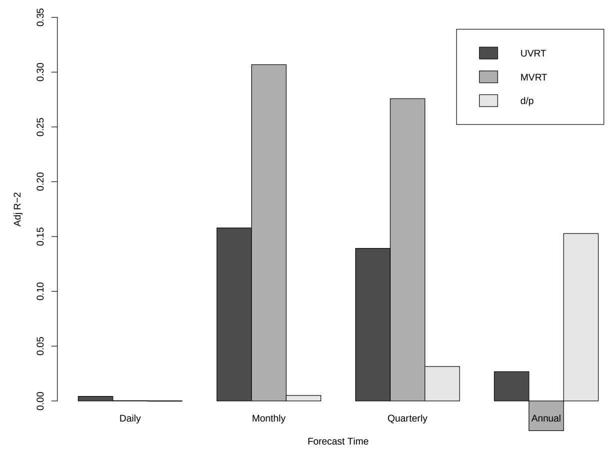

Figure 1: Adjusted R^2 - Forecasting Models

The x-axis on figure 1 represents the forecast horizon and the y-axis represents the adjusted R^2 for the forecast using different models. The first bar in this figure represents the forecast regression using the recently developed Recovery Theorem (Ross, 2015); the second bar represents the multivariate Recovery Theorem (Sanford, 2017); and the final bar uses a dividend-price ratio (Cochrane, 2008). Clearly, we do quite well at forecasting asset prices at horizons ranging from one month to a few years. However, very few models are successful at explaining short-term (say, one-day) variations in asset prices or returns. This is where sentiments from social media come into play: they can help explain these short-term variations. More specifically, I explore the proportion of short-term noise that can be explained using sentiments from social media.[4]

My findings indicate that, contrary to the literature's expectations, it is possible to forecast high-frequency stock returns and volatility jumps using consumer and investor sentiment indicators. Using tweets that I scraped from Twitter, I perform textual analysis to construct daily sentiment indices. While other scholars have relied on third-party companies like Stocktwits to complete these tasks, doing so reduces transparency

[4]For the purposes of this paper, short-term noise is is equivalent to short-term returns and therefore, the two are used interchangeably – simply put, daily returns are often viewed as noise.

and limits the potential for customization. The sentiment indices I constructed are numerical scores, not dichotomous variables, which allows me to control for sentiment strength (e.g., good vs. great) and not just positive/negative overall feelings. Results show that sentiment indices can not only be used to obtain out-of-sample forecasts of daily returns, but can also forecast volatility jumps. Using a Markov-switching framework, I find that, as overall sentiments shift from positive to negative, volatility jumps occur.

The rest of the paper proceeds as follows: section 2 defines and derives the model of short-term stock returns; section 3 introduces a Markov-switching framework for modeling volatility jumps; section 4 provides on overview of the data used in this paper; section 5 presents the empirical results; and section 6 concludes the paper.

2 Stock return modeling

This article tests whether stock returns can be attributed, at least in part, to a behavioral component. In the era of social media, other people's perceptions of a company and/or of its products have become important decision drivers for investors. The model presented in this section is one that is akin to the early factor models of Fama and French (1993, 2012) and are therefore quite simple from an econometric standpoint.

Stock returns To test whether sentiments can explain short-term asset returns, I employ a simple out-of-sample linear model, as follows:

$$r_t = \alpha + \beta\omega_{t-1} + v_t \tag{1}$$

where r_t is the current period's return and ω_{t-1} is the previous period's sentiment. Note that in regression 3, the sentiment variable, ω, appears as a single number. In reality, if we had a sentiment score distribution that ranged from -3 to $+3$, the sentiment variable would reflect the tweet count for each score in that range. Hence, the regression equation would look more like this:

$$r_t = \alpha + \beta_1\omega(-3)_{t-1} + \beta_2\omega(-2)_{t-1} + \beta_3\omega(-1)_{t-1} + \beta_4 \\ \omega(+1)_{t-1} + \beta_7\omega(+2)_{t-1} + \beta_6\omega(+3)_{t-1} + v_t \tag{2}$$

where $\omega(-1)$, for example, represents the number of times words associated with a sentiment score of -1 appeared in the sample. Similarly, one would expect that, as the amount of information being fed into the market at any given time to increase, so would the volatility of asset prices. This will be formally tested, again using a simple out-of-sample volatility regression as follows:

$$\Delta\sigma_t = \alpha + \beta\omega_{t-1} + v_t \tag{3}$$

where $\Delta\sigma_t$ is the change in a stock's volatility at time t and ω_{t-1} is the previous period's sentiment.

3 Volatility jumps

In section 2, I presented a model by which stock returns can be explained using Twitter sentiments. In addition, I suggest that a Markov-switching model will allow us to explain volatility jumps in asset returns (Hamilton, 1989). By definition, as the amount of short-term fluctuations increases in the market, so does volatility. As such, if a sentiment distribution helps to explain changes in a stock's return fluctuations, it should also explain changes in the stock's volatility. More interesting, however, is whether changes in sentiments can help explain drastic changes in volatility states. I test this proposition by assuming that volatility has two possible states (in the empirical section, I will also test a Markov-Switching model with three possible states): a low state and a high state. As the noise increases, so should volatility. Ultimately, this will show up in the volatility as volatility jumps. As the general sentiment about a given company or product changes from "positive" to "negative" (or vice versa), we would expect volatility jumps to occur. I assume a Markov-switching model with two possible volatility states described by the following set of equations:

$$\sigma_t = \begin{cases} \alpha_0 + \beta z_{t-1} + \epsilon_t, & s_t = 0 \\ \alpha_0 + \alpha_1 + \beta z_{t-1} + \epsilon_t, & s_t = 1 \end{cases} \tag{4}$$

where σ is the volatility of stock returns, z is the skewness of the sentiment distribution from previous period $t-1$, and ϵ_t and β are i.i.d. random variables with mean zero and variance equal to σ_ϵ^2. This model introduces a system where volatility is a linear model with two different intercepts: α_0 and $\alpha_0 + \alpha_1$. In other words, the "jump" in this system is based on the introduction of an additional intercept in the process (α_1). As a simple example, let us assume that coefficient β is equal to one, that we have a skewness value equal to 0.2, and that, in initial state s_0, α_0 is equal to zero. In other words, we have a system where the volatility is constant and equal to 0.2. In this switching model, when the state switches from state zero to state one, if we assume that the additional intercept term is equal to 0.1, we would now have a new volatility level equal to 0.3. This is the idea behind this Markov-switching model. I will assume that the transitions are governed by a first-order Markov process defined as follows:

$$\begin{aligned} Prob(s_t = 1|s_{t-1} = 1) &= p_{1,1} \\ Prob(s_t = 0|s_{t-1} = 1) &= p_{0,1} \\ Prob(s_t = 1|s_{t-1} = 0) &= p_{1,0} \\ Prob(s_t = 0|s_{t-1} = 0) &= p_{0,0} \end{aligned} \tag{5}$$

which can be represented in matrix form as:

$$P = \begin{bmatrix} p_{0,0} & p_{0,1} \\ p_{1,0} & p_{1,1} \end{bmatrix}$$

which represent the probabilities of switching between the two states of the model. For example, assuming that we were in state s_0 in the previous period, the probability of transitioning to state s_1 is equal to $p_{0,1}$. The question then is whether we can forecast accurately when the model will switch between states and, perhaps more interestingly, what is the probability of switching between states. As mentioned above, the switch will depend in large part on the skewness of the sentiment distribution derived in this paper. In other words, we should be able to re-write the probabilities as functions of sentiment skewness:

$$\begin{aligned} Prob(s_t = 1|s_{t-1} = 1, z_{t-1}) &= p_{1,1} \\ Prob(s_t = 0|s_{t-1} = 1, z_{t-1}) &= p_{0,1} \\ Prob(s_t = 1|s_{t-1} = 0, z_{t-1}) &= p_{1,0} \\ Prob(s_t = 0|s_{t-1} = 0, z_{t-1}) &= p_{0,0} \end{aligned} \tag{6}$$

These probabilities are precisely what we estimate in section 5.3. This simple example can easily be generalized to an N-State Markov-Switching model. For example, section 5.3.2 will extend the simple two-state model presented here to a three-state model.

4 Data

4.1 Sample selection

The sample used in this paper is from June 2009 (when I started scraping Twitter) to December 2009. The analysis stops on the last day of 2009 because it became impossible to scrape the entire population of tweets from Twitter (more below in section 4.2).

Why Apple? During the sample period, people tweeted about Apple approximately 600,000 times. By comparison, according to the website gigatweet.com, as of 17 March 2018, companies like United Airlines, Microsoft, Google, and Samsung have a similar number of tweets today to Apple back in 2009. Hence, it could be argued that, we could use Apple as a proxy for the Twitter exposure that the average company receives today. Furthermore, at the time, Apple was releasing what was then a relatively new product: the iPhone. People were still susceptible to others' impressions of the product. Because barriers to switch between technology products are high (because of the current Apple architecture – e.g. Apple Airplay ensures that only Apple products can be used together), one could argue that, at least today, overall impressions of products is diminished. Analyzing a company in the early stages of product adoption is helpful to test the premise posited in this paper.

I obtained return data for Apple stocks from the Wharton Research Data Services (WRDS) database for June–December 2009. Price data were collected from the Center for Research in Security Prices (CRSP) dataset.

4.2 Twitter data

Using Twitter data, I determine the sentiment "value" of tweets by associating certain words with a score. One of the most significant advantages of Twitter is the length of statements. Because of the 140-character limit, users must make precise statements. This means that, in general, we do not need to worry about things like double negatives. For example, we would not expect someone on Twitter to write: "I did not not want to go to the store." The limited length removes a lot of subjectivity based on grammar. It allows me to use the tweets at face value instead of wondering about ulterior meanings.

I scraped Twitter data directly from twitter.com. According to the World Wide Web Foundation (2018), tweets grew from about 2.5 million per day in 2009 to about 35 million per day in 2010. Twitter limits the number of live tweets that can be scraped from its servers. The tweeting levels reached in 2010 made scraping the entire population of tweets impossible beyond the last day of 2009. Why is this important? By 2010, the scraping limits imposed by Twitter meant that one's scraped database might miss influential tweets that had a significant impact on the stock market. In other words, if I was scraping tweets about Snapchat in early 2018, because of Twitter's scraping restrictions, it is not guaranteed that my algorithm would have captured the Kylie Jenner tweet that single-handedly affected Snapchat's stock price. As such, I have limited this analysis to the time period in which no tweets were missing (i.e. we could capture the entire tweet population): from June 2009 to December 2009. During this period, there were 600,000 tweets about Apple. Each tweet in the database contains the date and time of the tweet, its location, and the text of the tweet. To test the hypothesis that stock prices are affected by investor/consumer sentiments in the short-term, I subset tweets by day.

To produce the sentiment indices, I use two word dictionaries (see section 4.3). I match words from each tweet to their score in the two dictionaries and aggregate scores for each day. For example, if, on day X, 125 words with a score of +3 are used in tweets about Apple, then the +3 bin for that specific day will have a count of 125. The scores associated with each tweet represents the word with the maximum score in that tweet. In other words, online a single word in each tweet is used to determine that tweets sentiment score – this means that if a tweet has the words "good" and "great," since "great" has the highest score, that word will be used for that entire tweets sentiment score.

The sentiment indices are used to forecast daily fluctuations in asset prices. I aggregate all sentiment scores for a given day and compare them to the next-day stock price for the asset. Sentiment scores obtained on days where trading does not occur (such as weekends and holidays) are aggregated until trading occurs. For example, weekend tweets about Apple are summed together as a single measure (Friday to Sunday) that is used as the sentiment for the next trading day, Monday. As another example, if trading did not occur on a Tuesday because of a holiday, the tweets would be aggregated with the Monday tweets as the sentiments predicting stock price for Wednesday.

To isolate sentiments about specific companies, some tweets had to be dropped from the sample. For example, if one is constructing a sentiment index for Apple, someone's tweet about the great time they had at the apple orchard on a particular day cannot be included. As such, it is essential to filter the tweets to make sure that only tweets related to the company or its products remain in the sample. To accomplish this, I adopted a two-step triage system. The first step involves listing possible word combinations that should not be part of the final tweet sample. In our example, that would mean removing any tweet with the word *orchard* preceded or followed by the word *apple*. The second step involves examining each flagged tweet to ensure that it was not related to the company or its products. The second step was only possible because the subset of tweets that were rejected by the sorting algorithm was quite small (a few thousand tweets). In a larger sample, this entire process would need to be automated.

A final verification involved creating a subsample of one week and reading all tweets to verify their relevance to Apple, the company. The weekly subsample was selected randomly. All tweets in the subsample were indeed about Apple, which led me to conclude that the subsetting was working properly. To ensure that over-deletion did not occur, the deleted tweets were read individually. A minimal number (less than 0.1%) of tweets were brought back into the sample after being deleted.

4.3 Dictionaries

To associate words with scores, I use two separate word sentiment dictionaries (Loughran and McDonald, 2011; Nielsen, 2011). The two dictionaries 1) provide a robustness check on each other, and 2) attribute scores differently to words. Each dictionary is described in turn below.

Loughran and McDonald (2011) word dictionary The first dictionary (Loughran and McDonald, 2011) is a dictionary of mostly business words. For example, in the world of finance, we often say that we are "going long" when buying a stock. In normal parlance, going long does not mean much (at least not in terms of defining a sentiment). Yet, if a tweet states that investors should go long on Apple, it reflects a positive sentiment toward Apple products. The Loughran and McDonald (2011) dictionary defines words as either positive or negative (binary scores). For example, it would give a score of $+1$ to the word *good* and a score of -1 to the word *bad*. Similarly, the dictionary gives a score of $+1$ to the word *excellent* and a score of -1 to the word *horrible*.

The dictionary combines words from the EDGAR 10-X fillings with base words from the English dictionary. Only those words that appear at least 100 times in the 10-X filings and that can be identified as actual words are added to the dictionary. The dictionary is updated almost every year to incorporate words that are in vogue at that moment in the business world. Words are divided into seven categories: negative, positive, uncertain, litigious, constraining, superfluous, and interesting. In this article, I use only words that fall into the negative or positive categories because these are the words that are clearly associated with sentiments. The rest of the dictionary is somewhat arbitrary. In total, the dictionary contains over $85,000$ words. Once we subset the word list to only include words that are positive or negative, we are left with a dictionary of about $2,700$ words. As an example, the dictionary identifies *boom* as a positive word and *bankrupt* as a negative word. Interested readers should visit the website to obtain more information on the dictionary compilations and the various dictionaries available.[5]

Nielsen (2011) word dictionary Instead of proposing a positive-negative dichotomy, the second dictionary (Nielsen, 2011) distinguishes the relative strength of words. Scores range from negative three to positive three. Using this dictionary, I am able to demonstrate that forecast results are wildly different when we account for the fact that *excellent* is stronger than *good*. For example, I conducted a textual analysis to match the words for every Apple-related tweet on 3 August 2009 (see figure 2). A score is recorded for each tweet containing a word in the Nielsen (2011) dictionary. This information can be aggregated into a distribution such as the one shown in figure 2. The sentiment distribution allows us to: 1) control for the relative strengths of certain words, and 2) determine if certain characteristics of the distribution have a stronger impact on prices than others. More specifically, certain groupings of words or moments of the distribution may have different impacts on asset prices.

[5] `https://www3.nd.edu/~mcdonald/Word_Lists.html`

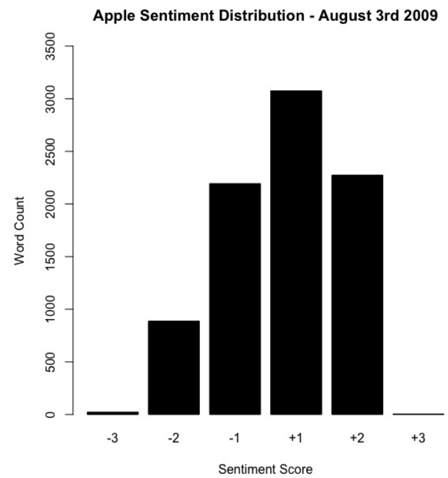

Figure 2: Apple Sentiment Distribution - August 3rd 2009

Figure 2 gives an example of what the sentiment index of Apple looks like on August 3rd 2009. On this day, the distribution seems to be quite normal, although one could argue that the distribution does appear to have a slight negative skew (longer tail on the negative side of the distribution).

One shortcoming of the Nielsen (2011) dictionary is that Twitter users may consciously choose to use words with fewer characters. In the example above, *excellent* was stronger than *good*. However, *excellent* has five more characters than *good*. Even if a user truly wanted to use the word *excellent*, they may decide to use the word *good* to ensure that they can say everything they want to say within the 140-character limit imposed by Twitter. Furthermore, the Nielsen (2011) dictionary was not constructed to associate sentiment scores to business terminology. For example, it does not recognize *long* and *short* as words that have any sentimental value. This is why using both the Loughran and McDonald (2011) dictionary (which codes sentiments for finance words) and the Nielsen (2011) dictionary (which codes a range of sentiments) is so important.

One shortcoming of both dictionaries is that they cannot identify "abbreviations/ shortcuts." For example, while *great* would receive a positive sentiment score, *gr8* is not scored. Although it is difficult to know what abbreviations are in vogue at any point in time, it would be worthwhile to consider these devices in future research.

5 Results

5.1 Descriptive statistics

The sample period for this article is from June to December 2009, which corresponds to 138 trading days. In this section, I describe Apple data for the sample period (prices, returns, etc.) and provide summary statistics for the sentiment indices.

Table 1 shows the number of observations (138 days), minimum, first quartile, median, mean, third quartile, maximum, and number of missing observations for each variable. The descriptive statistics are not surprising or out of the ordinary. For example, the average return during the period of interest for Apple was about 0.32% per day while the maximum daily return was about 6.8%.

Variable	**n**	**Min**	$\mathbf{q_1}$	$\widetilde{\mathbf{x}}$	$\bar{\mathbf{x}}$	$\mathbf{q_3}$	**Max**	**#NA**
Prices	138	19.14	23.26	26	25.20385	27.96	30.44714	0
Log Prices	138	2.95178	3.147	3.258	3.2183	3.331	3.415992	0
Return	138	-0.04410595	-0.00714	0.002689	0.003194133	0.01367	0.06787332	0
SD of Return	138	0.01279945	0.01478	0.01609	0.0166411	0.01865	0.02632063	0

Table 1: Descriptive Statistics DVs

Table 2 provides summary statistics for the sentiment analysis based on the Nielsen dictionary. The first six rows represent the six categories of sentiments, from "very very negative" (*VV Negative* or -3) to "very very positive" (*VV Positive* or $+3$). Interestingly, on average, 14.45 words are considered *VV Negative* daily compared to 2.20 *VV Positive* words. As with most review websites, there seem to be more people complaining about Apple than praising it (or perhaps people generally use stronger language when denouncing companies/products). Thus, we can see a a much larger number of observations at the *VV Negative* category when compared to the *VV Positive* category, on average. This, however, only seems to be the case at the extreme ends of the distribution. In the middle, there are a larger number of positive sentiment scores counts (toward the *V Positive* and *Positive* categories).

Variable	n	Min	q_1	$\widetilde{x}$	$\bar{x}$	q_3	Max	#NA
VV Negative	138	0	8	15	14.44928	20	38	0
V Negative	138	26	212	350.5	402.9638	592	1288	0
Negative	138	54	409.8	750	850.1522	1260	2262	0
Positive	138	75	689.8	1307	1472.725	2162	4236	0
V Positive	138	56	528.5	969	1110.812	1673	3478	0
VV Positive	138	0	1	2	2.195652	3	11	0
Total	138	211	1827	3400	3853.297	5766	11292	0
$\frac{\text{Positive}}{\text{Negative}}$	138	1.335907	1.945	2.059	2.085759	2.185	3.235915	0
Mean	138	35.16667	304.5	566.8	642.2162	961.1	1882	0
SD	138	28.67296	264.8	494	552.8232	823.9	1610.637	0
Kurt	138	-2.954419	-1.928	-1.67	-1.435617	-1.287	1.621007	0
Skew	138	-0.06308114	0.1219	0.1691	0.2309631	0.286	0.9515991	0

Table 2: Descriptive Statistics IVs (Nielsen dictionary)

In table 2, the seventh row (under the six possible sentiment strengths) is the total count for the sentiment distribution: the number of times words were associated with any sentiment category for each day. The variable ranges from 211 to $11,292$, which indicates that the extent to which people discussed Apple on Twitter varied widely depending on the day. The variables mean, SD, kurt, and skew represent the mean, standard deviation, kurtosis, and skewness of the daily distribution of the sentiment index. The skewness variable will be used later as a summary measure of the overall sentiment about Apple in the economy. When the skew shifts toward the left, it means that the overall sentiment regarding Apple has become more negative, and vice versa. This is important because it allows us to capture the nuances of word selection when characterizing a company or product on Twitter.

Table 3 presents the same variables as table 2. However, the sentiment categories are presented as proportions of the total number of tweets with flagged sentiment scores on any given day. I use this as a robustness check to ensure that regression results are not simply a function of the level variables.

Variable	n	Min	q_1	$\widetilde{x}$	$\bar{x}$	q_3	Max	#NA
VV Negative	138	0	0.002716	0.003677	0.003844734	0.00492	0.009205426	0
V Negative	138	0.07124682	0.0962	0.1052	0.1038737	0.111	0.1470924	0
Negative	138	0.159601	0.2066	0.2168	0.21826	0.229	0.3269877	0
Positive	138	0.2829947	0.3735	0.3818	0.3906763	0.4006	0.5137157	0
V Positive	138	0.2066116	0.2681	0.2839	0.2827466	0.2971	0.3362885	0
VV Positive	138	0	0.0001577	0.0004389	0.0005986797	0.0009185	0.002991027	0
Mean	138	0.1666667	0.1667	0.1667	0.1666667	0.1667	0.1666667	0
SD	138	0.1358272	0.141	0.1433	0.1451027	0.1467	0.178436	0
Kurtosis	138	-2.954419	-1.928	-1.67	-1.435617	-1.287	1.621007	0
Skewness	138	-0.06308114	0.1219	0.1691	0.2309631	0.286	0.9515991	0

Table 3: Descriptive Statistics Ratio IVs

Finally, table 4 provides summary statistics for the sentiment analysis based on the Loughran and McDonald dictionary. This dictionary focuses on words that are regularly used in business. As such, it is used frequently in the textual analysis literature in business (Kuhnen and Niessen, 2012; Garcia, 2013; Loughran and McDonald, 2014; Li et al., 2014). Here, we are including words that may not have been captured by the other (Nielsen) sentiment dictionary, which provides a kind of robustness check. However, I cannot conduct as many in-depth tests as with the Nielsen index because this dictionary does not assess the strength of terms. The first four rows of table 4 present the sentiment measures. The first two are the positive and negative tweet words matching the dictionary, and the second two are proportions of the total number of sentiment words on any given day. The ratio-of-positive-to-negative variable is the number of words that were flagged as positive (numerator) compared to the number of words that were flagged as negative (denominator). Finally, the weighted ratio variable is the ratio of positive to negative multiplied by the total number of tweets expressing a sentiment. This variable captures both the skew from positive to negative (a ratio that is less than one implies more negative tweets that day), and the "importance" of a given day. For example, a day with a thousand tweets might be more important than a day with only ten. The next section discusses the characteristics observed from time series graphs.

Variable	n	Min	q_1	$\widetilde{x}$	$\bar{x}$	q_3	Max	#NA
Positive Terms	138	19	194	425	474.3188	718.2	1181	0
Negative Terms	138	16	153	352	384.8333	552.5	1331	0
Proportion Good to Total	138	0.1985294	0.4249	0.4475	0.4510679	0.4776	0.7018425	0
Proportion Bad to Total	138	0.2981575	0.5224	0.5525	0.5489321	0.5751	0.8014706	0
Ratio Bad to Good	138	0.424821	1.123	1.237	1.279647	1.361	4.037037	0
Total	138	35	359.5	781	859.1522	1271	2292	0
Weighted Ratio	138	37.6129	452.4	939.1	1082.909	1675	3023.036	0

Table 4: Descriptive Statistics IVs (Loughran and McDonald dictionary)

5.2 Time-series characteristics

This section plots the time-series figures for the dependent variables used in the out-of-sample forecast regressions. All data are for Apple stock from June to December 2009. In each group of figures below, the left is the time series plot and the right is the autocorrelation plot. The autocorrelation plot verifies that the series is covariance stationary, which is important in a forecast regression. The first two sets of figures (figures 3–6) are used in the forecast regressions while the last set of figures (figures 7–8) is used for the Markov-switching volatility jump forecasting tests.

Figures 3 and 4 are the daily returns plot and autocorrelation function (ACF) plot for Apple stock, respectively. Visual inspection of the figures suggests the series is at least approximately covariance stationary and ergodic. In a forecasting exercise, a process should be covariance stationary and ergodic because this ensures that our indicators are explaining variations in the returns, and not only explaining the trend of the returns, for example (Hamilton, 1994).

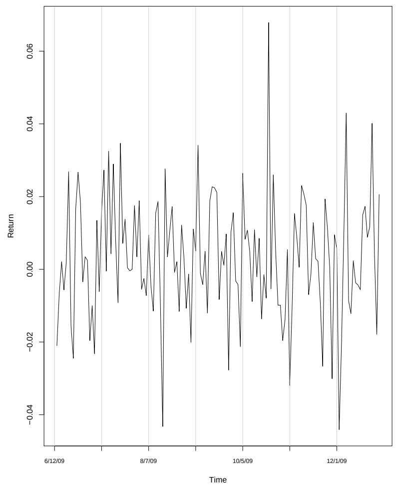

Figure 3: Apple Returns

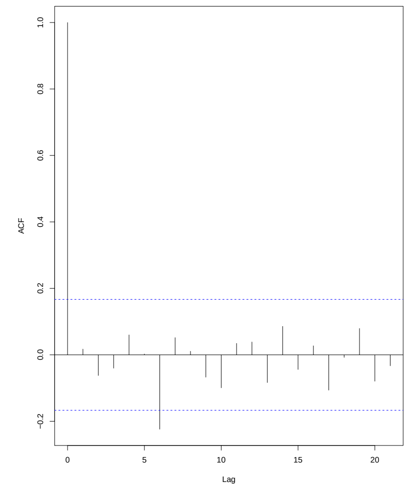

Figure 4: ACF Plot of Returns

Figures 5 and 6 illustrate that the change in daily volatility also appears to be approximately covariance stationary and ergodic. Other than a few spikes in the change in volatility, both the mean and variance are fairly constant. There are no significant autoregressive lags.

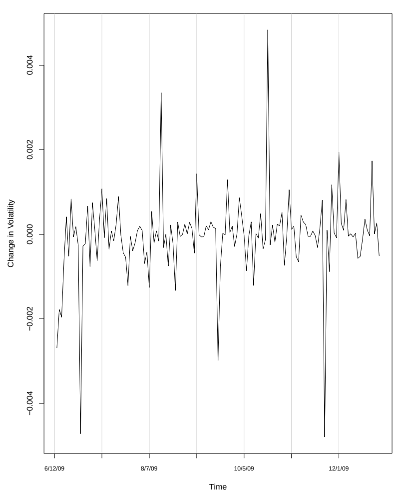

Figure 5: Change in Volatility

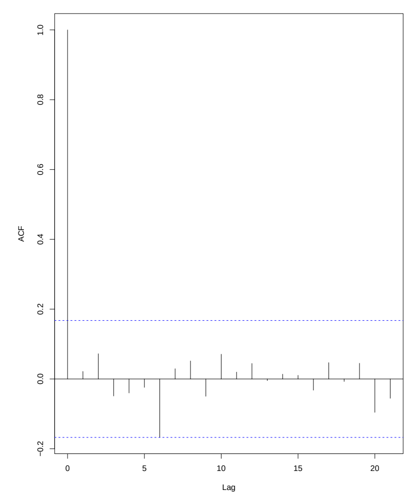

Figure 6: ACF Plot of Change in Volatility

Unlike the previous figures, the volatility time series seen in figures 7 and 8, as expected, does not seem to be covariance stationary and ergodic. The purpose of the volatility analysis is to model the jumps in figure 7 using the skewness of the sentiment distribution. As such, we do not want to transform the series because these jumps are exactly what this article attempts to model. They are the long vertical lines that go up or down significantly. They represent a significant change in volatility from one day to the next. They also appeared in figure 5 as the spikes in the time series figure. The objective below will be to determine if change in skewness of the index can forecast these volatility jumps (for a more in-depth discussion of volatility jumps, see Eraker et al. (2003); Eraker (2004); Kou (2002)).

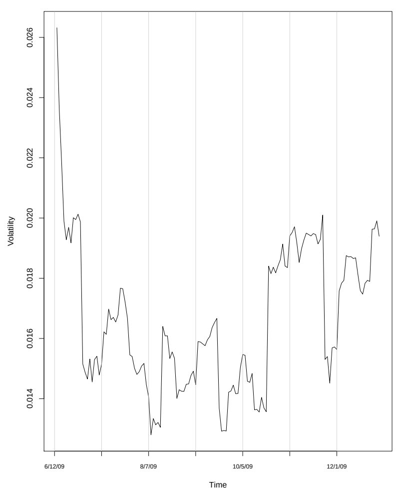

Figure 7: Volatility

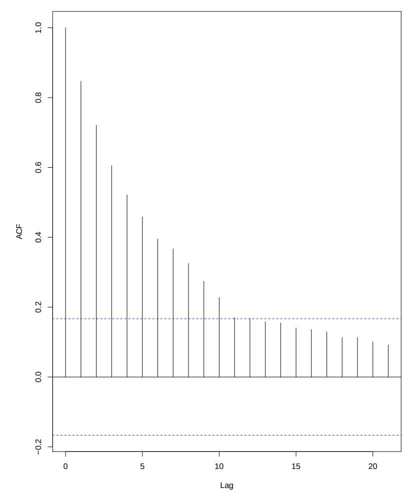

Figure 8: ACF Plot of Volatility

There are theoretical reasons to believe that sentiments indices would impact both returns and volatility jumps. Several articles (including French et al., 1987; Turner et al., 1989; Campbell and Hentschel, 1992; Bekaert and Wu, 2000) have documented "volatility feedback" effects, where periods of low returns are contemporaneously associated with higher volatility. Hence, we should expect that certain indicators, such as sentiments, would be able to forecast both volatility jumps and an asset's returns. The following section presents empirical results.

5.3 Markov-switching results

In this section, I test whether a Markov-Switching model will allow us to explain volatility jumps (or large, instantaneous volatility changes) in asset returns. Let us begin with an ordinary least squares (OLS) regression as follows:

$$\sigma_t = \alpha + \beta skew_{t-1} + \epsilon \tag{7}$$

where σ_t is the daily volatility of Apple stock, α is the regression intercept, β is the regression coefficient, $skew_{t-1}$ is the previous period's skewness value (obtained from the daily skewness of the sentiment index), and ϵ is the error. This ordinary linear regression is used because the relationship between the skewness of the sentiments index (the independent variable) and the volatility is a linear one. Once we have determined this relationship, we can then assume two or three states and the Markov-Switching model should allow us to obtain 1) a different set of parameters for each state and 2) a better fit overall (Goldfeld and Quandt, 1973; Hamilton, 1989; Kim, 1994).

5.3.1 Two State Model

We start the Markov-Switching method by first presenting results for when we are simply assuming two possible states. As previously mentioned, we first assume a linear relationship between the dependent and the independent variables. For this setting, the regression produces the following results:

	Model 1
(Intercept)	0.01574^{***}
	(0.00031)
$skew_{t-1}$	0.00375^{***}
	(0.00106)
R^2	0.0390
Num. obs.	138

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Table 5: Apple tweets June–December 2009 (≈600,000 tweets)

Now, we can adopt a two-state Markov-switching specification to get the following state-dependent regressions results:

	Regime 1	Regime 2
(Intercept)	0.0181^{***}	0.0147^{***}
	(0.0002)	(0.0002)
$skew_{t-1}$	0.0046^{***}	0.0016^{*}
	(0.0007)	(0.0007)
R^2	0.4435	0.0539
Num. obs.	138	138

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Table 6: Apple tweets June–December 2009 (≈600,000 tweets)

Based on the results in table 6, it is clear that the best fit occurs when we are in regime one. The skew variable does not seem to do a very good job at specifying the regime 2 based model. This is the first clue that perhaps a three state model should be used. This will be tested in the next subsection.

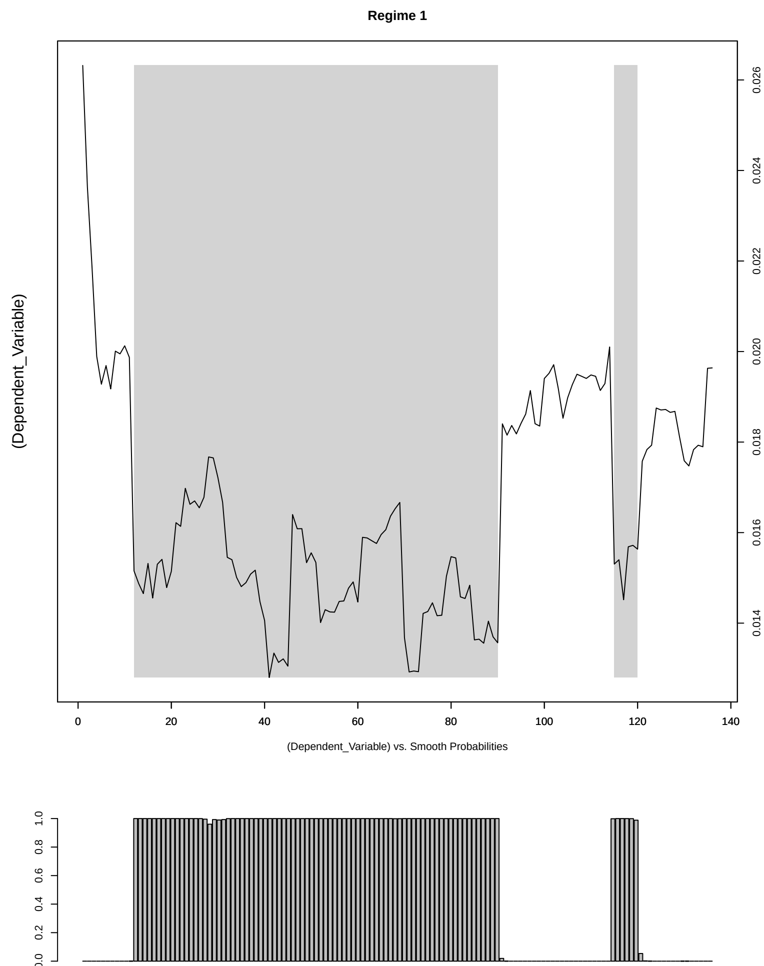

Figure 9: Apple Volatility

Figure 9 shows the fit of the model where we include the one-period lagged skewness of the sentiment distribution. We see that the model captures the two different volatility levels quite well (a high volatility state and a low volatility state). The grayed-out area corresponds to the low-volatility regime. This is the first volatility state in the model.

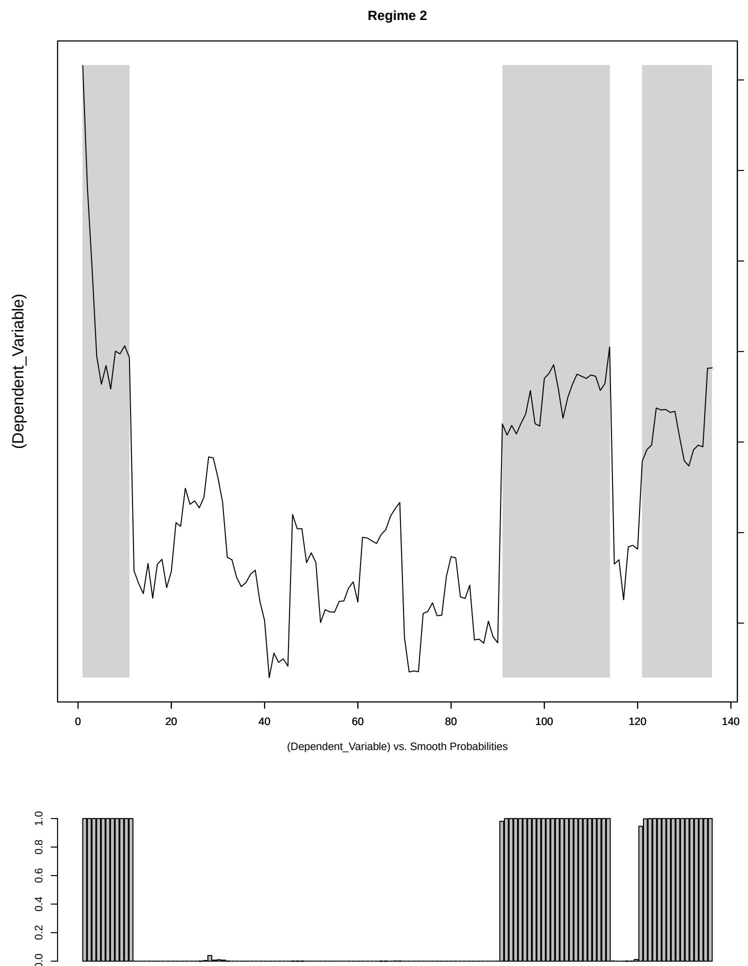

Figure 10: Apple Volatility

The second volatility state is high volatility, which can be seen in figure 10. If we define the high-volatility state as H and the low-volatility state as L, we can rewrite the transition probability matrix as follows:

$$P = \begin{bmatrix} p_{L,L} & p_{L,H} \\ p_{H,L} & p_{H,H} \end{bmatrix}$$

where, for example, $p_{L,L}$ represents the probability that the volatility will transition from a current low-volatility state to a future low-volatility state. In other words, $p_{L,L}$

is the probability that the volatility will remain the same at some point in the future.

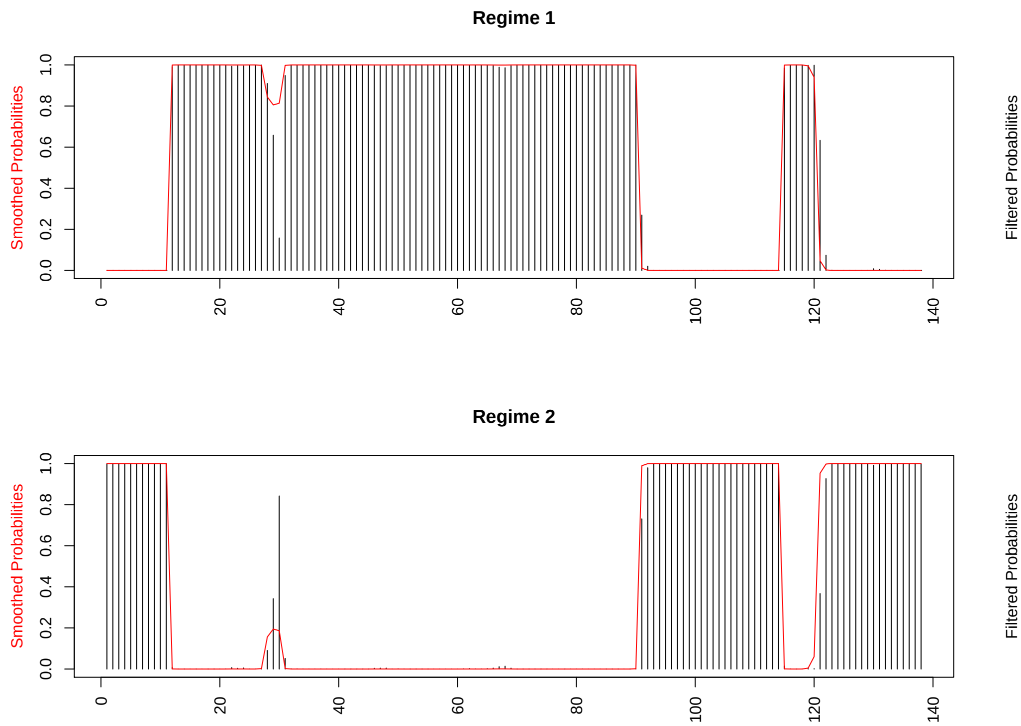

Figure 11: Smoothed and Filtered Probabilities

Figure 11 shows the smoothed and filtered probabilities for regime one and two based on the Markov-switching specification. The resulting transition probability matrix is:

$$P = \begin{bmatrix} p_{L,L} & p_{L,H} \\ p_{H,L} & p_{H,H} \end{bmatrix} = \begin{bmatrix} 0.9598 & 0.0241 \\ 0.0402 & 0.9759 \end{bmatrix}$$

which illustrates that the initial states are highly persistent. Intuitively, these transition probabilities indicate that there is a very small probability of switching from one state to another in any given period. The three-state Markov-switching model presented in the next section will provide us with more interesting transition dynamics.

5.3.2 Three State Model

Next, we introduce the Markov-Switching method by assuming three possible states. As previously mentioned, we first assume a linear relationship between the dependent and the independent variables. For this setting, the regression produces the following results:

	Model 1
(Intercept)	0.01551^{***}
	(0.000335)
$skew_{t-1}$	0.00236^{*}
	(0.00114)
$skew_{t-2}$	0.00233^{*}
	(0.00116)
R^2	0.0934
Num. obs.	138

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Table 7: Apple tweets June–December 2009 (≈600,000 tweets)

Clearly, the results in table 7 leave much to be desired. The natural question is whether or not we can improve these results by assuming a regime-switching framework. This implies, as previously mentioned, that we will be looking at the various coefficients for the relationship between the independent and dependent variables by assuming that there can be three different regression coefficients. Adopting a Markov-switching specification we get the following results:

	Regime 1	Regime 2	Regime 3
(Intercept)	0.0180^{***}	0.0150^{***}	0.0134^{***}
	(0.003)	(0.0003)	(0.0003)
$skew_{t-1}$	0.0027^{***}	0.0018^{*}	0.0012^{*}
	(0.0008)	(0.0007)	(0.0006)
$skew_{t-2}$	0.0020^{***}	0.0020^{***}	0.0020^{***}
	(0.0004)	(0.0004)	(0.0004)
R^2	0.3185	0.3747	0.4375
Num. obs.	138	138	138

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Table 8: Apple tweets June–December 2009 (≈600,000 tweets)

where based on the results in table 8, we can think of the different states as a high volatility state, a current volatility state, and a low volatility state. The goal is to determine the coefficient that corresponds to the relationship whenever the volatility switches to one of these states. Clearly, the results here suggest that a three state model does a much better job at explaining the variation and/or the "switches" in volatility for Apple stock.

Figures 12 to 14 graphs the volatility of Apple's stock compared to the smoothed probabilities. Intuitively, the greyed area represents the different volatility states defined by the three-state Markov-Switching model.

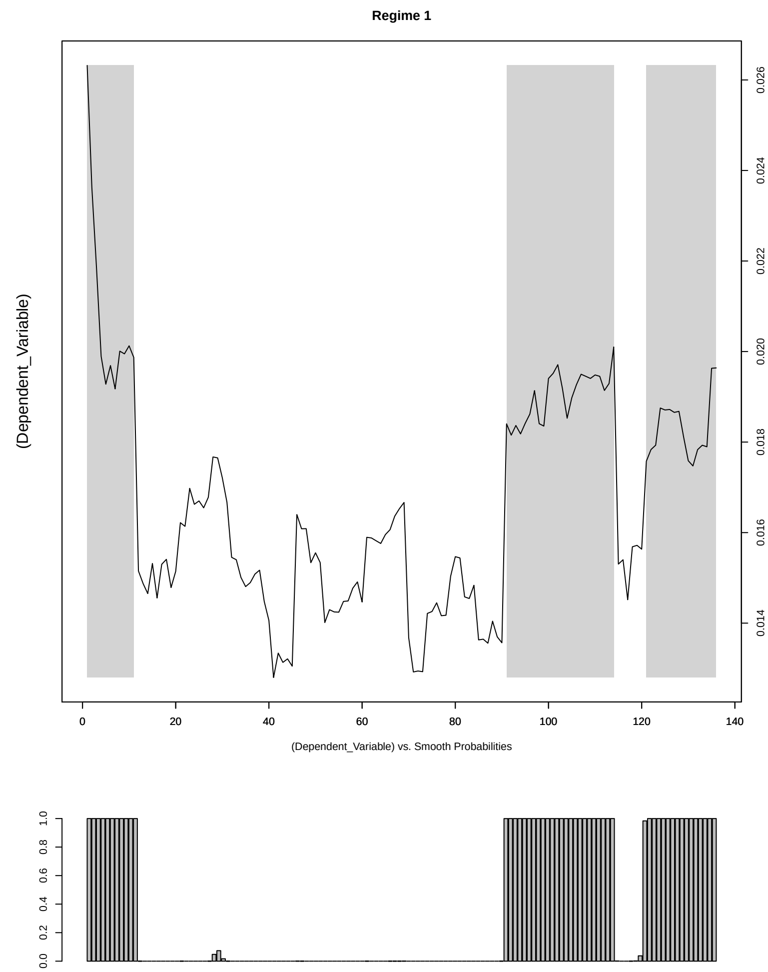

Figure 12: Apple Volatility

We can think of regime one from figure 12 as the high volatility state.

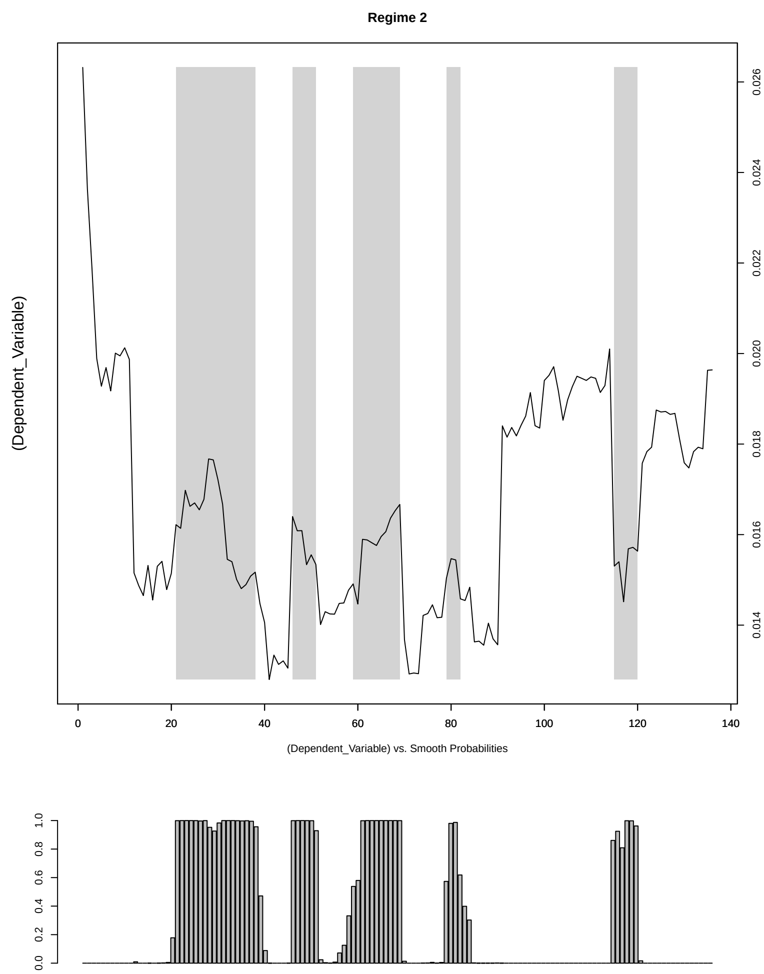

Figure 13: Apple Volatility

We can think of regime one from figure 13 as the high volatility state.

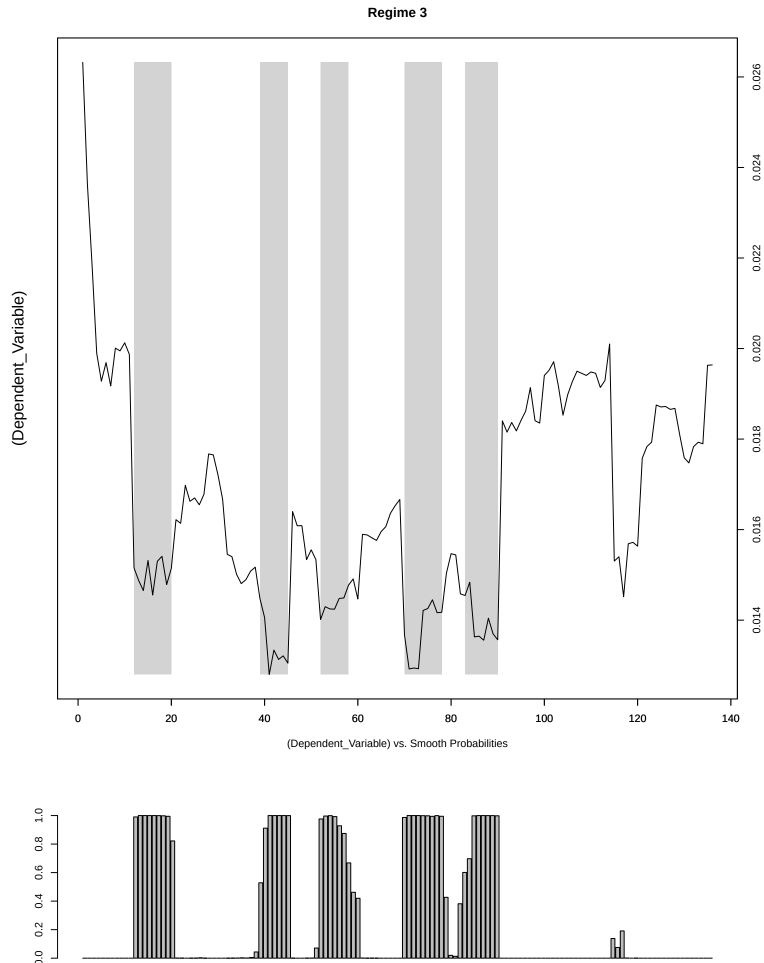

Figure 14: Apple Volatility

We can think of regime one from figure 14 as the high volatility state. If we define the high-volatility state as H, the medium (or constant) volatility state as M, and the low-volatility state as L, we can rewrite the transition probability matrix as follows:

$$P = \begin{bmatrix} p_{L,L} & p_{L,M} & p_{L,H} \\ p_{M,L} & p_{M,M} & p_{M,H} \\ p_{H,L} & p_{H,M} & p_{H,H} \end{bmatrix}$$

where, for example, $p_{L,L}$ represents, as before, the probability that the volatility will transition from a current low-volatility state to a future low-volatility state. In other words, $p_{L,L}$ is the probability that the volatility will remain the same at some point in the future.

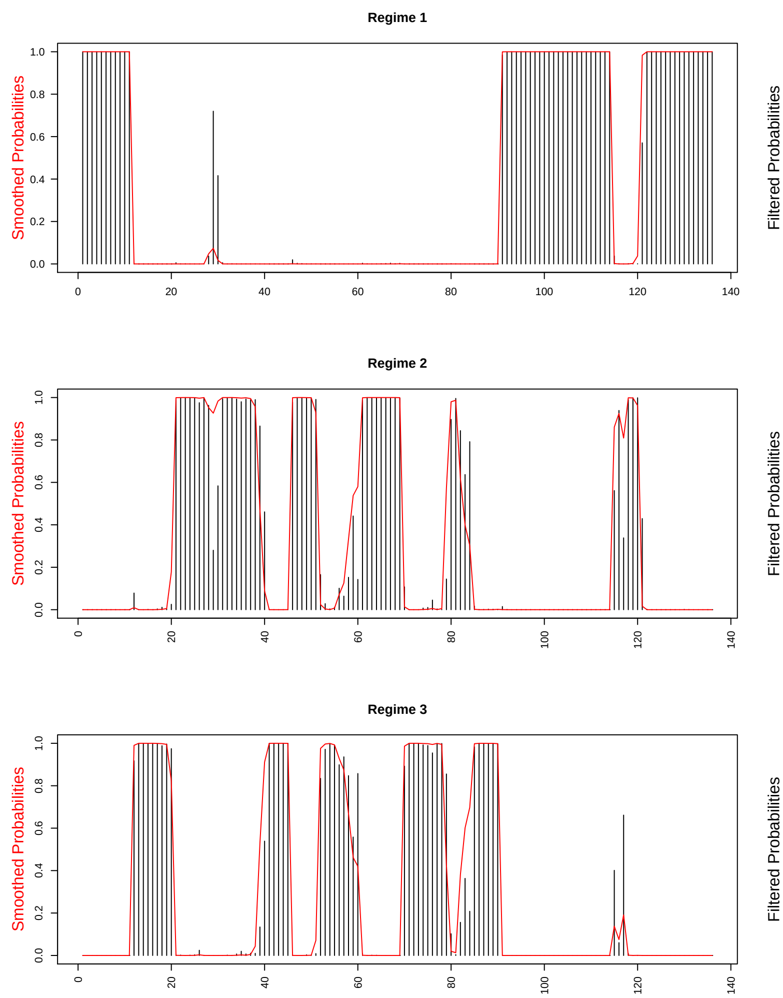

Figure 15: Smoothed and Filtered Probabilities

Figure 15 shows the smoothed and filtered probabilities for regimes one, two, and three. Again, the model does seem to perform significantly better at explaining the various levels of volatility when compared to the two-state Markov-switching model.

$$P = \begin{bmatrix} p_{L,L} & p_{L,M} & p_{L,H} \\ p_{M,L} & p_{M,M} & p_{M,H} \\ p_{H,L} & p_{H,M} & p_{H,H} \end{bmatrix} = \begin{bmatrix} 0.9594 & 0.02123 & 0.0281 \\ 0.0211 & 0.8806 & 0.1057 \\ 0.0195 & 0.0981 & 0.8662 \end{bmatrix}$$

Much like the two-state Markov-switching model, the model does indicate that the states are still highly persistent. However, for example, the transition probability $p_{M,M}$ indicates that there is a much larger chance of transitioning to another state when compared to any of the other states in the two-state model presented earlier.

5.4 Forecast

In this section, I present the results for the out-of-sample (OOS) regressions for return and volatility of Apple stock. To be clear, OOS forecasts mean that at time zero, a user of the model would obtain a forecast for time t and then these results are compared to the realized return for the same period. All OOS regressions conducted in this section are, unless otherwise noted, for a one-day period.

5.4.1 Results – Nielsen word dictionary

The first set of results are for the OOS regressions using the Nielsen (2011) word dictionary. The regression specification (table 9) is the following:

$$r_t = \alpha + \sum_{i=1}^{6} \beta_{t-1,i} z_{t-1,i} + \epsilon_t \tag{8}$$

where r_t is the return at time t, α is the regression intercept, $\beta_{t-1,i}$ is the regression coefficient on the independent variable with a one-day lag, $z_{t-1,i}$ is the previous day's sentiment for the various sentiment scores i, and ϵ_t is the regression error term.

	Model 1	Model 2	Model 3
(Intercept)	-0.0540	-0.0516^{**}	-0.1576^{***}
	(0.0403)	(0.0191)	(0.0409)
Proportion VV Negative		$1.6742^{\dagger}$	0.2430
		(0.8926)	(0.8891)
Proportion V Negative		0.1427	0.2304^{*}
		(0.1160)	(0.1134)
Proportion Negative		0.1544^{*}	0.1699^{**}
		(0.0612)	(0.0615)
Proportion Positive	0.1167^{*}		0.1579^{**}
	(0.0583)		(0.0505)
Proportion V Positive	0.0286		0.1316^{*}
	(0.0754)		(0.0646)
Proportion VV Positive	6.1016^{*}		
	(2.3933)		
R^2	0.0815	0.0761	0.1225
Num. obs.	138	138	138

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Table 9: Apple Out-of-Sample Return Regressions

In table 9, the variables are calculated as the number of observations in a given sentiment category divided by the total number of observations on that day. For example, if there were 1,000 sentiment words on any given day and 100 were in the *V Negative* category, the *V Negative* variable would get an observation of 0.1 on that specific day. The OOS forecast for this model specification is quite good with an R^2 of about 0.123. Since the variables in the models presented are proportion variables, this means that one of the variables must be omitted from the model. Thus, the coefficients are in relation to that reference variable. Model 3 in table 9 has the reference variable as *VV Positive*. This level of OOS forecasting at the daily interval compares to the forecasting power of models for medium- (say monthly) and long-term (say yearly) OOS forecasts (Cochrane, 2009). Note that all coefficients in table 9 are positive. This is because these coefficients are with respect to the reference category variable.

The regression specification for table 10 is the following:

$$\Delta\sigma_t = \alpha + \sum_{i=1}^{6} \beta_{t-1,i} z_{t-1,i} + \epsilon_t \tag{9}$$

where $\Delta\sigma_t$ is the change in volatility at time t, α is the regression intercept, $\beta_{t-1,i}$ is the regression coefficient on the independent variable with a one-day lag, $z_{t-1,i}$ is the previous day's sentiment for the various sentiment scores, i, and finally ϵ_t is the regression error term.

	Model 1	Model 2	Model 3
(Intercept)	$-0.0003^{\dagger}$	-0.0001	-0.0001
	(0.0002)	(0.0002)	(0.0002)
VV Negative	0.1838		0.9056
	(1.7266)		(1.7128)
V Negative	-0.2706^{*}		0.0543
	(0.1332)		(0.1750)
Negative	0.1530^{*}		0.1946^{**}
	(0.0598)		(0.0707)
Positive		-0.1814^{**}	-0.2582^{***}
		(0.0539)	(0.0682)
V Positive		0.2289^{***}	0.1503^{*}
		(0.0660)	(0.0705)
VV Positive		10.2847^{*}	10.1482^{*}
		(4.5090)	(4.5306)
R^2	0.0589	0.1057	0.1599
Num. obs.	137	137	137

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Table 10: Apple Out-of-Sample Change in Volatility Regressions (scaled IV variables)

In table 10, the key indicators are the raw sentiment word counts. It is important to note that these independent variables are different than those from table 9 in that they are levels instead of being proportions. This is because the volatility is, in effect, a measure of noise. Noise will be introduced based on the number of tweets we observe on Twitter. This is why the level of the independent variable is used here instead of the proportions. Here, the coefficients for the independent variables have been scaled by a magnitude of $100,000$ to see the full effect of the sentiments on changes in volatility. So the coefficients are going to correspond to a $100,000$ change in the independent variables. In other words, one hundred thousand *Negative* tweets would be expected to have a 0.1530 effect on the daily change in the standard deviation of returns. As was the case before, the OOS forecast for changes in volatility for Apple is quite good with an R^2 of about 0.16 in the model 3 specification. The coefficients increase significantly along with the "level" of the variable. In other words, as the words become more positive, the effect on the change of the volatility also increases, as we would expect. Sentiment indices do not impact changes in volatility in a clear direction. One would expect that extreme sentiments should lead to larger swings in volatility, which is exactly what is observed in the results. But whether a slightly positive or slightly negative sentiment should have a positive or negative effect on the change in the volatility is not so clear. The effects for variables outside of the VV categories are almost zero, perhaps as it should be.

The regression specification for table 11 is the following:

$$\Delta\sigma_t = \alpha + \sum_{i=1}^{4} \beta_{t-1,i} z_{t-1,i} + \epsilon_t \tag{10}$$

where $\Delta\sigma_t$ is the change in volatility at time t, α is the regression intercept, $\beta_{t-1,i}$ is the regression coefficient on the independent variable with a one-day lag, $z_{t-1,i}$ is the previous day's sentiment for the various sentiment scores, i, and ϵ_t is the regression error term.

	Model 1	Model 2	Model 3
(Intercept)	$-0.0003^{\dagger}$	-0.0001	-0.0001
	(0.0002)	(0.0002)	(0.0002)
VV and V Negative	-0.2571^{*}		0.0199
	(0.1219)		(0.1711)
Negative	0.1525^{*}		0.2031^{**}
	(0.0596)		(0.0710)
Positive		-0.1431^{**}	-0.2126^{**}
		(0.0519)	(0.0656)
VV and V Positive		0.1977^{**}	$0.1250^{\dagger}$
		(0.0655)	(0.0694)
R^2	0.0584	0.0720	0.1282
Num. obs.	137	137	137

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Table 11: Apple Out-of-Sample Change in Volatility Regressions (scaled IV variables)

In table 11, we collapse the most extreme sentiment categories (VV and V) into a single category as a robustness check because the observations in the VV and V categories are limited. As was the case before, the sentiment levels still explain a significant share of changes in the volatility of Apple's returns. The OOS R^2 for model 3 is about 0.13.

5.4.2 Results – Loughran and McDonald word dictionary

The second set of results is for the OOS regressions using the Loughran and McDonald (2011) word dictionary. The regression specification for table 13 is the following:

$$r_t = \alpha + \sum_{i=1}^{k} \beta_i \left(\frac{Good}{Total} \right)_{t-i} + \epsilon_t \tag{11}$$

where r_t is the return at time t, α is the regression intercept, β_i is the regression coefficient on the independent variable at lag k, the ratio $\left(\frac{Good}{Total}\right)_{t-i}$ is the previous period's count of good sentiment words to bad sentiment words, and finally ϵ_t is the regression error term.

	Model 1	Model 2
(Intercept)	-0.0053	-0.0026
	(0.0110)	(0.0158)
$\left(\frac{Good}{Total}\right)_{t-1}$	-0.0194	0.0198
	(0.0241)	(0.0243)
$\left(\frac{Good}{Total}\right)_{t-2}$		-0.0062
		(0.0244)
R^2	0.0390	0.0424
Num. obs.	133	132

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Table 12: Apple Out-of-Sample Return Regressions (scaled independent variable)

One important thing to note about table 13 is that the coefficients are for sentiment scores for every 1,000 tweets. The coefficients in this regression specification are consistent when we add an additional lag into the model specification (model 2). These results show that the OOS forecasts are clearly dependent on the dictionary specification. Here, clearly, the sentiments obtained using the Loughran and McDonald (2011) word dictionary are unable to capture the returns of Apple during the period used in this paper. In the appendix, I propose to use a weighted measure of good to bad sentiments.[6]

6 Conclusion

In this paper, I performed a textual analysis of tweets to obtain sentiment indices that explained stock returns and volatility jumps for Apple. There are two key contributions of this paper: first, I find that by using sentiment indices rather than sentiments obtained from dichotomous variables, I am able to forecast daily fluctuations in stock retruns. Second, I find that by adopting a very simple Markov-switching model, it is possible to use the skewness of the sentiment index distribution to better predict volatility jumps. In other words, when overall sentiment towards a company or its

[6]This ratio of good to bad is shown because it allows us to reflect the skewness of the sentiment about Apple in the market.

products shift, we can expect a volatility jump to occur. Hence, I would argue that consumer and investor perceptions (or sentiments) does indeed matter in asset pricing and that firms should consider investing in making sure that their social media presence is viewed in a relatively positive light. Firms would also greatly benefit from carefully considering what consumers and investors are saying about them on social media because these can, as shown in this paper, serve as proxies for their future stock returns.

One important implication of this piece concerns open access of social media data. Without a doubt, such data should be used carefully to preserve the privacy of social media users. However, as I have shown in this paper, it could allow us to further our understanding of the linkages between economic systems and the systems' participants. The literature has long argued that stock market noise and volatility jumps could not be explained using traditional datasets. And yet, here we are. Social media information could potentially help us explain phenomenon that have yet to be understood in economics more generally, and in financial markets more specifically. But, as of today, if we were to attempt to collect the population of tweets over any significant amount of time, it would cost millions of dollars and take researchers several years because of the artificial barriers imposed by social media companies (limits on requests, for example). Although social media platforms own these data, there should be realistic pathways through which researchers can access data without having to incur unrealistically prohibitive time and cost constraints.

References

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4):1645–1680.

Bartov, E., Faurel, L., and Mohanram, P. (2015). Can twitter help predict firm-level earnings and stock returns? *The Accounting Review*.

Bekaert, G. and Wu, G. (2000). Asymmetric volatility and risk in equity markets. *The review of financial studies*, 13(1):1–42.

Borzykowski, B. (2016). How investors are using social media to make money. Online. [Online; accessed 23-March-2018].

Boudoukh, J., Feldman, R., Kogan, S., and Richardson, M. (2013). Which news moves stock prices? a textual analysis. Technical report, National Bureau of Economic Research.

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of financial Economics*, 31(3):281–318.

Cen, L., Lu, H., and Yang, L. (2013). Investor sentiment, disagreement, and the breadth–return relationship. *Management Science*, 59(5):1076–1091.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4):1533–1575.

Cochrane, J. H. (2009). *Asset Pricing:(Revised Edition)*. Princeton university press.

Da, Z., Liu, Q., and Schaumburg, E. (2013). A closer look at the short-term return reversal. *Management science*, 60(3):658–674.

Das, S. R. and Chen, M. Y. (2007). Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management science*, 53(9):1375–1388.

Dunn, J. (2017). It's been a year since smartphones started killing the headphone jack, and it's still inconvenient. [Online; accessed 18-March-2018].

Eraker, B. (2004). Do stock prices and volatility jump? reconciling evidence from spot and option prices. *The Journal of Finance*, 59(3):1367–1403.

Eraker, B., Johannes, M., and Polson, N. (2003). The impact of jumps in volatility and returns. *The Journal of Finance*, 58(3):1269–1300.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of financial economics*, 105(3):457–472.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of financial Economics*, 19(1):3–29.

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300.

Goldfeld, S. M. and Quandt, R. E. (1973). A markov model for switching regressions. *Journal of econometrics*, 1(1):3–15.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384.

Hamilton, J. D. (1994). *Time series analysis*, volume 2. Princeton university press Princeton.

Hribar, P. and McInnis, J. (2012). Investor sentiment and analysts' earnings forecast errors. *Management Science*, 58(2):293–307.

Huang, D. (2015). Retail traders wield social media for investing fame. Online. [Online; accessed 23-March-2018].

Kim, C.-J. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, 60(1-2):1–22.

Kollmeyer, B. (2016). The twitter accounts investors need to follow in 2017. Online. [Online; accessed 23-March-2018].

Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management science*, 48(8):1086–1101.

Kuhnen, C. M. and Niessen, A. (2012). Public opinion and executive compensation. *Management Science*, 58(7):1249–1272.

Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., and Chen, Y. (2014). The effect of news and public mood on stock movements. *Information Sciences*, 278:826–840.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

Loughran, T. and McDonald, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance*, 69(4):1643–1671.

Nielsen, F. Å. (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Openshaw, J. (2013). 3 ways investors can use twitter. Online. [Online; accessed 23-March-2018].

Roll, R. (1988). R2. *The Journal of Finance*, 43(3):541–566.

Ross, S. (2015). The recovery theorem. *The Journal of Finance*, 70(2):615–648.

Sanford, A. (2017). Recovery theorem with a multivariate markov chain.

Turner, C. M., Startz, R., and Nelson, C. R. (1989). A markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics*, 25(1):3–22.

Vasquez, J. (2018). In one tweet, kylie jenner wiped out $1.3 billion of snap's market value. [Online; accessed 12-March-2018].

World Wide Web Foundation (2018). Twitter usage statistics. [Online; accessed 16-March-2018].

A Appendix – Additional Results

This appendix presents another set of results for the OOS regressions using the Loughran and McDonald (2011) word dictionary. The regression specification for table 13 is the following:

$$r_t = \alpha + \sum_{i=1}^{k} \beta_i \left(\frac{Bad}{Good}\right)_{t-i} \cdot total_{t-1} + \epsilon_t \tag{12}$$

where r_t is the return at time t, α is the regression intercept, β_i is the regression coefficient on the independent variable at lag k, the ratio $\left(\frac{Bad}{Good}\right)_{t-i}$ is the previous period's count of good sentiment words to bad sentiment words, $total_{t-1}$ is the total number of tweets in the previous period, and finally ϵ_t is the regression error term. The ratio of good to bad is used here because it allows us to reflect the skewness of the sentiment about Apple in the market.

	Model 1	Model 2
(Intercept)	8.529^{**}	9.174^{**}
	(2.612)	(2.982)
$\left(\frac{Bad}{Good}\right)_{t-1} \cdot total_{t-1}$	-0.00463^{*}	-0.00461^{*}
	(0.00199)	(0.00224)
$\left(\frac{Bad}{Good}\right)_{t-2} \cdot total_{t-2}$		-0.00051
		(0.00222)
R^2	0.0390	0.0424
Num. obs.	133	132

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$

Table 13: Apple Out-of-Sample Return Regressions (scaled independent variable)

One important thing to note about table 13 is that the coefficients are for sentiment scores for every $1,000$ tweets. The coefficients in this regression specification are consistent when we add an additional lag into the model specification (model 2).