More than Words: Twitter Chatter and Financial Market Sentiment

Travis Adams, Andrea Ajello, Diego Silva, Francisco Vazquez-Grande

2023-034

More than Words: Twitter Chatter and Financial Market Sentiment*

Travis Adams, Andrea Ajello, Diego Silva, and Francisco Vazquez-Grande[†]

Federal Reserve Board

May 2023

Abstract

We build a new measure of credit and financial market sentiment using Natural Language Processing on Twitter data. We find that the Twitter Financial Sentiment Index (TFSI) correlates highly with corporate bond spreads and other price- and survey-based measures of financial conditions. We document that overnight Twitter financial sentiment helps predict next day stock market returns. Most notably, we show that the index contains information that helps forecast changes in the U.S. monetary policy stance: a deterioration in Twitter financial sentiment the day ahead of an FOMC statement release predicts the size of restrictive monetary policy shocks. Finally, we document that sentiment worsens in response to an unexpected tightening of monetary policy.

1   Introduction

Does social media activity carry any meaningful signal on credit and financial markets' sentiment? We build a new real-time sentiment index derived from social media communications related to credit and financial markets. We rely on sentiment analysis of Twitter data and show that financial sentiment gauged from social media contains predictive information for stock returns and proves sensitive to monetary policy surprises, predicting tightening moves ahead of FOMC statement releases, as measured by several event-study monetary policy shocks developed in the literature.

We query a large sample of tweets that contain words and word clusters from financial- and credit-market dictionaries (Calomiris and Mamaysky, 2019), from the universe of social media posts available on Twitter since 2007. For each tweet in our sample, we measure sentiment using FinBERT a language model developed by Araci (2019) from BERT (Devlin et al., 2018) and specifically designed to measure sentiment of financial text. Our index draws from the universe of Twitter users who post financial content and is available in real time, as new tweets appear on the platform and their sentiment is assessed. Averaging sentiment values of posted tweets, we build a historical index of financial market sentiment and name it the Twitter Financial Sentiment Index (TFSI). We document that time variation in the TFSI can be attributed to changes in the extensive margin of users engaging in posting positive or negative sentiment tweets, rather than to the intensive margin—i.e., users posting tweets with higher or lower sentiment.

We show that the monthly TFSI correlates highly with market-based measures of financial sentiment, such as corporate bond spreads, the Excess Bond Premium (EBP) (Gilchrist and Zakrajšek, 2012), and survey-based measures of consumer confidence, such us the Michigan confidence index. We also find that our index correlates positively with market-based measures of borrowing costs, such as corporate credit spreads.

With the index at hand, we make two main contributions. First, we show that overnight Twitter sentiment can help predict daily stock market returns–i.e., the average tweeted sentiment between 4pm on day $t-1$ to 9am on day t helps forecasts stock market returns on day t after controlling for standard asset pricing factors. This fact speaks to the ability of tweeted sentiment to reflect information that will later be included in stock prices once U.S. markets open.

Second, the TFSI predicts the size of restrictive monetary policy surprises. We show that Fed-related tweets play a dominant role on FOMC days and, notably, that Twitter sentiment

after the first day of the FOMC meeting can predict the size of restrictive monetary policy shocks in connection with the release of the FOMC statement the following day. This last results holds across three measures of monetary policy shocks, identified by means of the event studies in Miranda-Agrippino and Ricco (2021), Jarociński and Karadi (2020), and Bauer and Swanson (2022). In other words, Twitter financial sentiment ahead of monetary policy decisions incorporates useful information that can help predict the market reaction around the FOMC statement release. We also find that the TFSI worsens in response to an unexpected tightening in the policy stance.

We contribute to the literature that attempts to measure financial market sentiment (see for example López-Salido et al. (2017); Danielsson et al. (2020), Greenwood and Hanson (2013), Shiller (2015b), Fama and French (1988), Baker and Wurgler (2000) and Lettau and Ludvigson (2001)), employing natural language processing to harness information from Twitter posts as a novel data source.

Time variation in average sentiment across tweets can broadly capture changes in expectations, risk appetite, beliefs, or emotions representative of a wide array of Twitter users. Traditional gauges of financial market sentiment are based on asset prices, portfolio allocation flows (Baker and Wurgler, 2006; Gilchrist and Zakrajšek, 2012), investors' surveys (Qiu and Welch, 2004), and news archives (Tetlock, 2007; Garcia, 2013). While measures based on portfolio allocations, prices, and news coverage can be monitored at high frequency, survey measures imply that sentiment is polled infrequently. Such sentiment measures are derived from actions, market outcomes, opinions and commentary from selected groups of actors rather than from the wider public.

Time variation in credit and financial market sentiment has proven to be an important predictor of asset returns (Shiller, 2015a; Greenwood and Hanson, 2013), and driver of credit and business cycles (López-Salido et al., 2017), and we aim to explore this transmission by means of our index in the near future and in future iterations of this paper.

Moreover, central bank decisions and communication strategies, intended to fine-tune the stance of monetary policy and share information on the state of the central bank's economic outlook, affect market participants' expectations, risk sentiment, and beliefs, as policy transmits to the broader economy (Gertler and Karadi, 2015; Miranda-Agrippino and Rey, 2020; Bekaert et al., 2013).

Our paper relates to a particular strand of literature that studies the role of text-based mea-

3

sure of financial market sentiment. Financial sentiment measured from news archives has been shown to predict stock market performance (Tetlock, 2007; Garcia, 2013). Research based on social media data show that the Twitter activity of institutions, experts, and politicians contains useful information to study various aspects related to central banking. As central banks have become more active on Twitter (Korhonen and Newby, 2019; Conti-Brown and Feinstein, 2020), Azar and Lo (2016) find that tweets that refer to FOMC communication can help predict stock market returns. Our results show that twitter sentiment can help predict stock market returns more systematically and can anticipates changes in the stance of monetary policy. Masciandaro, Peia, and Romelli (Masciandaro et al.) use dissimilarity between Fed-related tweets and FOMC statements to identify monetary policy shocks, while Meinusch and Tillmann (2017), Stiefel and Vivès (2019), and Lüdering and Tillmann (2020) use tweets to estimate changes in public beliefs about monetary policy and their impact on asset prices, although they do not explore the ability of twitter sentiment to forecast monetary policy shocks. Ehrmann and Wabitsch (2022) focus on studying view divergence and polarization in response to central bank communication. They show that following ECB communication, tweets primarily relay information and become more factual and that public views become more moderate and homogeneous. High-impact decisions and communications, such as Mario Draghi's "Whatever it takes" statement, instead trigger a divergence in views.

Recent work applies sentiment analysis to a wider set of central bank communication tools. Notably, Correa et al. (2020) measure sentiment in central banks' financial stability reports, introducing a dictionary tailored to financial stability communications, confirming that general dictionaries, including finance dictionaries such as Loughran and McDonald (2016), might not be suitable to assess tonality in a financial stability context. Binder (2021), Bianchi et al. (2019), Camous and Matveev (2021), and Tillmann (2020) show that tweets by former U.S. president Trump about the Federal Reserve and its policy stance affected long-term inflation expectations and confidence of consumers, suggesting that the wider public priced in future reductions in interest rates in response to the president's social media activity. Finally, Angelico et al. (2022) show that Twitter can be an informative data source to elicit inflation expectations in real time.

4

2   Methodology

In this section we describe our strategy to sample financial tweets from the Twitter historical and real-time enterprise-level Application Programming Interface (API) (Twitter, Inc., Inc.). We then describe how we filter the data, pre-process tweets, and compute sentiment to produce readings of our index at different time frequencies.

2.1   Sampling

We query a subset of tweets related to financial market developments from the universe of all tweets available since 2007. Calomiris and Mamaysky analyze news articles from the Thompson Reuters archive and isolate a set of 60 word roots related to financial discourse that we use to discipline the sample selection of historical and real-time tweets, downloaded from the Twitter APIs.

Downloading all tweets that contain any combination of word roots in the Calomiris and Mamaysky set proves undesirable and infeasible: word derived from roots in the set can have multiple meanings–e.g., the word "bond" can be used to mean "connection" as well as "fixed income obligation". A large, unsystematic query has a higher likelihood of contaminating the sample with non-financial tweets—and surpasses by at least one order of magnitude our contracted Twitter API download quota.

To discipline the sample of tweets, we use Keyword Clustering, pairing the set of word roots into groups that are semantically similar. We measure similarity across keywords by means of their cosine distances from machine-learning-generated semantic similarity vectors (Yamada et al., 2020): a trained machine assesses the similarity of our keywords based on their occurrence within the body of text of Wikipedia. Figure 1 shows that the three clusters loosely map into financial contracts (Group 1), entities (Group 2), and actions or contractual features (Group 3).

Our query uses logical operators to filter tweets that contain at least one word from each of the three clusters. Technical features of the Twitter API require that single tweets be downloaded as separate entities—that is, the search engine treats threads and quote tweets as disconnected tweets, while retweets are linked to the original tweet by means of a boolean operator and share their creating time and date with the original tweet even when they were posted at a later time. We pre-process the text of all tweets by removing excess white spaces, tags, hyperlinks, and information that is not part of the text body of the tweet. We only keep tweets with unique

text, filtering out full and near replicas of tweets, to reduce the number of bot-generated entries in our dataset.[1]

Figure 1: Semantic Similarity Clusters

| Group 1		Group 2		Group 3
OR	AND	OR	AND	OR
• Bond		• Corporate		• Coupon
• Debt		• Company		• Interest
• Security		• Subsidiary		• Rate
• Credit		• Municipal		• Term
• Loan		• Sovereign		• IPO
• Mortgage		• Program		• Liquidity
• Portfolio		• Market		• Yield
• Pension		• Federal Reserve		• Downgrade
• Federal Funds		• Collateral		• Outstanding
• Leverage		• Credit Agency		• Repayment
• Financing		• Etc.		• Default
• Etc.				• Etc.
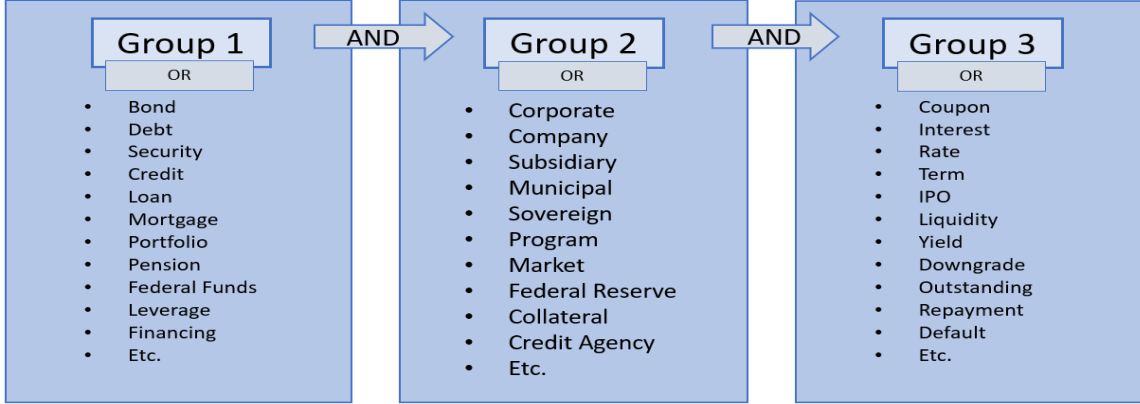
Note: This diagram displays an example of how financial words are clustered in three groups based on semantic similarity. See Appendix B for the full list of word and word roots, by cluster.

Our data query and preprocessing deliver a total of 4.4 million single tweets from 2007 to April 2023. Figure 2, plots the number of tweets downloaded per month since the beginning of the sample. Two structural features affect our data query. First, prior to 2011, as Twitter was burgeoning as a social media platform and its popularity was low, the number of tweet pulls averages around one hundred tweets per day, offering limited amount of text to measure sentiment at daily or weekly frequency. Second, in November 2017 Twitter increased the maximum character length of tweets from 140 to 280 characters, a change that makes it more likely to detect any three-word sequence in our query within a single post, resulting in a discrete jump in tweets pulled each month thereafter. Worth of note is the fact that discrete events, such as the start of the COVID-19 pandemic, the pivot in communication toward a tightening cycle in September 2021, and more recently the collapse of Silicon Valley Bank positively affect the number of tweets in the data pull from our query. The sample we use for our baseline analysis in sections 3 and 4 starts in September 2011—after the step increase in the number of monthly pulls visible in Figure 2—and includes 4.3 million tweets.

---

[1]In a similar spirit, we filter out tweets that advertise credit cards, crypto currency trades, and tweets related to topics that are only seemingly related to financial or credit market discourse, such as those that include words like "social security". Appendix B contains the full list of words from Calomiris and Mamaysky clustered in the three groups, and a detailed methodology to replicate our tweet selection and data cleaning.

Figure 2: Number of Tweets Selected per Month
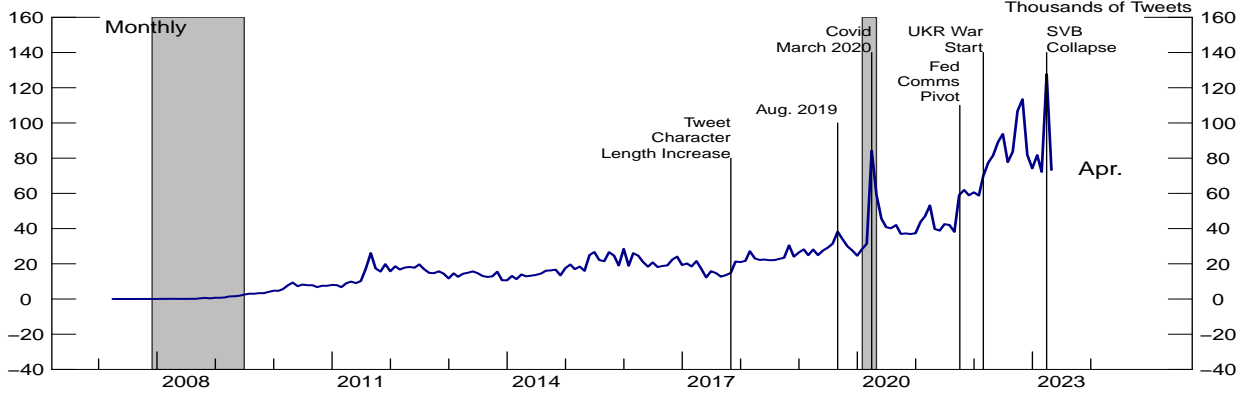
Note: This figure represents the number of tweets in the sample per month. The shaded bars indicate periods of business recession as defined by the National Bureau of Economic Research: December 2007–June 2009, February 2020–April 2020.
Source: Authors' calculation based on Twitter enterprise-level API data.

2.2 Measuring Sentiment

We use FinBERT (Araci, 2019) as our baseline tool to compute a sentiment value for each tweet of our sample. FinBERT is a language model based on Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018) to tackle natural language processing tasks in financial domain. An advantage of this tool relative to other text sentiment gauges is that it is specifically designed and trained to perform well measuring sentiment of financial text, making it the ideal candidate for our purpose.

We use FinBERT's compound score to assign a numeric value of sentiment to each tweet. FinBERT provides three probabilities to measure the odds that the analyzed text conveys positive neutral or negative sentiment, and also offers a compound sentiment score computed as the difference between the probability of the text having positive sentiment and the probability of the text having negative sentiment. Therefore FinBERT, provides us with a sentiment score between -1 and +1 for each tweet in our sample. For the purpose of measuring the evolution of average sentiment over time, given a time window, we first assign a sentiment score of zero to tweets that are labeled as neutral, and then sum the sentiment score of all remaining tweets

7

and divide by the total number of tweets posted over the desired time frame. Following this methodology, we can average sentiment across all tweets sampled in any desired time span and compute financial sentiment values at different time frequencies.[2]

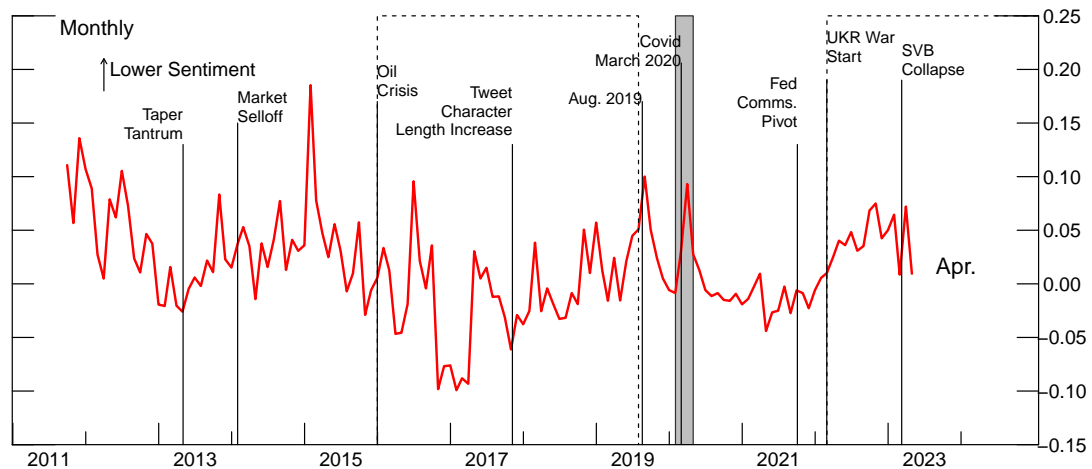# 3 The Twitter Financial Sentiment Index

The TFSI is calculated as the average sentiment across tweets in our sample for any given time period, as such it can be displayed in real time and at any time frequency. Figure 3, plots the TFSI at daily and monthly frequencies (top and bottom panel), since 2011. The daily index is particularly volatile at high frequencies especially before the Twitter CLI in 2017, after which the sentiment signal appears more informative day-to-day. For ease of comparability with other gauges of financial conditions, the index is oriented so that higher values indicate a deterioration in sentiment: the index rises in alignment with episodes of elevated stress in the U.S. financial system or tightening of financial conditions. These episodes include the Taper Tantrum in 2013, market selloff related to emerging market stress in 2014, and turbulence in late 2015 and early 2016 associated with fears related to the Chinese economy and a pronounced drop in oil price, the increase in likelihood that the US economy would enter a recession in August 2019, the COVID recession in 2020, and in the souring of financial sentiment associated with the onset of the Russian conflict in Ukraine and the beginning of tightening of monetary policy with the Fed communication pivot in September 2021.

---

[2]For robustness, we also measure sentiment using VADER, a lexicon and rule- based sentiment analysis tool that is specifically designed to measure sentiment expressed in social media (see Hutto and Gilbert (2014)). An advantage of this tool relative to other text sentiment gauges is that is better equipped to parse modifier words, and emojis to assess sentiment in social media text. Results using VADER sentiment are available upon request.

Figure 3: Twitter Financial Sentiment Index, Daily (Up), Monthly (Down)
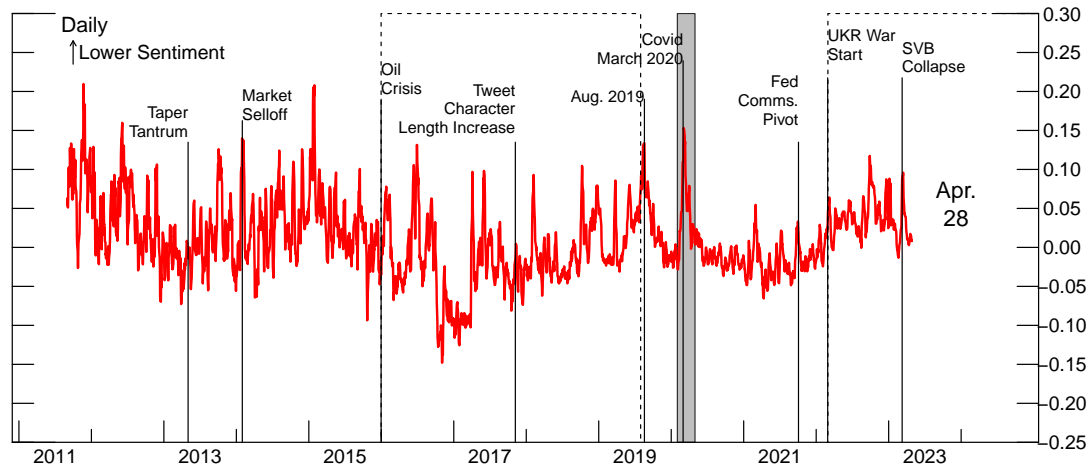
(a) Daily Twitter Financial Sentiment Index — Seven-Day Moving Average

(b) Monthly Twitter Financial Sentiment Index

Note: Increases in the TFSI point to a worsening of sentiment. Data sample starts from September 2011 and ends in April 2023. The dashed boxes indicate monetary policy tightening cycles: January 2016- August 2019, March 2022-present. The shaded bar indicate periods of business recession as defined by the National Bureau of Economic Research: February 2020–April 2020.
Source: Authors' calculation based on Twitter enterprise-level API data.

We also find that time-variation in the index can be mostly explained by the share of users that post tweets with positive or negative sentiment, rather than by the intensity of the tweeted sentiment. In principle, the value of the index would vary by changes in the intensive or the extensive margin, that is, it could be driven by changes in the sentiment value of the tweets or by the share of tweets with positive or negative sentiment values posted in a given time interval. Figure 4 compares our baseline index (in red) with the share of negative minus the share of positive tweets (in green), a measure of engagement on either side of the sentiment fronts. The similarity between the two lines demonstrates how most of the variation in the index is related to users' engagement on the extensive margin, rather than in the intensity of their sentiment expressed in their tweets.

Figure 4: Twitter Financial Sentiment Index vs. Share of Negative minus Share of Positive Tweets
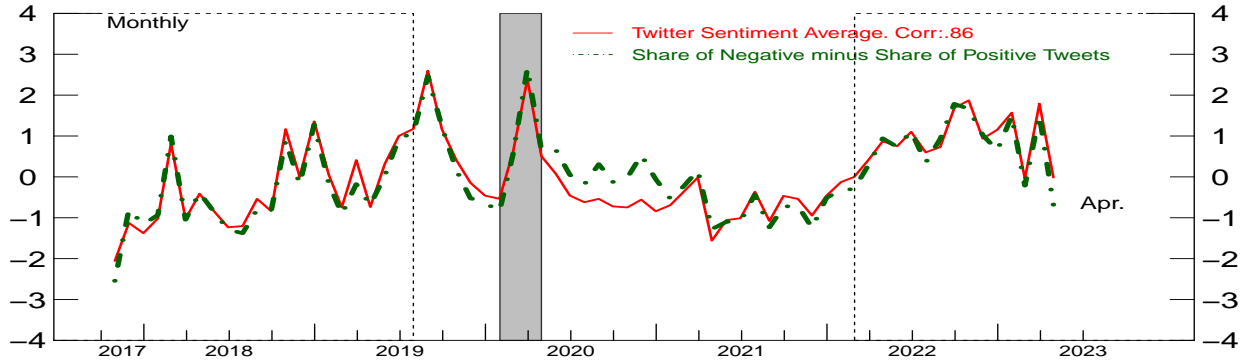
Note: The chart plots the Twitter Financial Sentiment Index (red solid line), against the difference in share of negative- and positive-sentiment tweets (green dashed line) at a monthly frequency and both standardized. Data sample starts from September 2011 and ends in April 2023. Increases in the TFSI point to a worsening of sentiment. The dashed boxes indicate monetary policy tightening cycles: January 2016- August 2019, March 2022-present. The shaded bar indicate periods of business recession as defined by the National Bureau of Economic Research: February 2020–April 2020.
Source: Authors' calculation based on Twitter enterprise-level API data.

4 Results

This section summarizes our main results. We show that the TFSI correlates with indexes and market gauges of financial conditions at monthly frequency. We also show that overnight twitter sentiment can help predict daily stock market returns. Finally, we show that Twitter financial sentiment can predict the size of restrictive monetary policy surprises and has a muted response to the realization of monetary policy shocks.

4.1 TFSI and Financial Conditions

Figure 5, compares the monthly TFSI with measures of financial conditions and economic and financial sentiment based on surveys and market prices since the Twitter CLI: the Baa corporate bond spread (top), and the Excess Bond Premium (EBP) of Gilchrist and Zakrajšek (2012) (middle) and the University of Michigan Consumer Sentiment index (bottom). The TFSI, while noisier, generally co-moves positively with these measures. These figure show that our sample selection and sentiment measure, that does not depend at all on market prices or surveys, presents a quantitatively and qualitatively similar picture to the most common metrics of economic and financial conditions.

Figure 5: TFSI vs Measures of Financial Conditions and Sentiment

TFSI and Baa Corporate Bond Spreads

TFSI and Excess Bond Premium
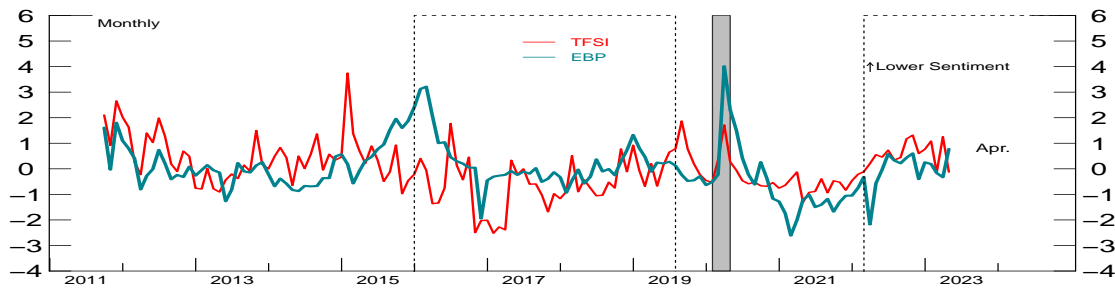
TFSI and U. Michigan Consumer Sentiment Index
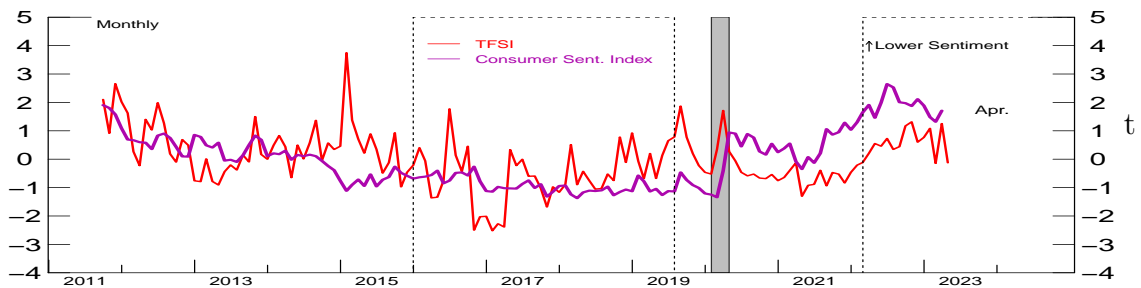
Note: Increases in the TFSI point to a worsening of sentiment. The dashed boxes indicate periods of monetary policy tightening cycles: January 2016-August 2019, March 2022-present. The shaded bars indicate periods of business recession as defined by the National Bureau of Economic Research: February 2020–April 2020.

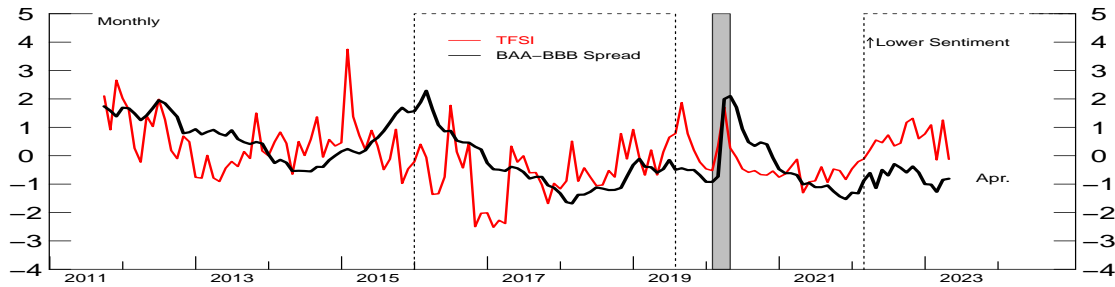Source: TFSI: Authors' calculation based on Twitter Enterprise-level API; Baa Spreads: Moody's via FRED; EBP: Federal Reserve Board, Favara et al. (2016); Consumer Sentiment Index: University of Michigan via FRED

12

4.2 TFSI and Stock Market Returns

We show that the TFSI can be used to forecast intraday returns of the S&P 500 index, even after controlling for other common predictors such as the VIX, the Fama-French stock market factors (Fama and French, 2015), financial sentiment present in official media sources (Shapiro, 2020), and lagged S&P 500 returns. One advantage of Twitter data is that it is available in real time, 24 hours a day. We take advantage of this feature to construct a measure of sentiment available when financial markets are closed. We compute the sentiment of all tweets in our sample that are posted overnight, that is, between 4pm at date $t-1$ and 9am at date t. We then run the following daily regressions:

$$SP500_{t_{(9am)}->t_{(4pm)}} = \alpha + \beta\ TFSI_{t-1_{(4pm)}->t_{(9am)}} + \gamma'X_{t-1} + \varepsilon_t$$

where $SP500_{t_{(9am)}->t_{(4pm)}}$ are the daily intraday Standard and Poor's 500 index market returns (from S&P Global, CapitalIQ), $TFSI_{t-1_{(4pm)}->t_{(9am)}}$ is our measure of overnight sentiment, and X_{t-1} is a vector of controls.

Table 1 displays the results. Each column of the table adds common predictors of daily returns as controls, that is, lagged S&P 500 index returns, the overnight return on the S&P 500—the returns between the close of market on day $t-1$ and opening of the market on day t—the VIX, and the three stock market factors of Fama and French (HML, High minus Low, SMB, Small minus Big, and MOM, Momentum). In all specifications the overnight sentiment index has a negative and significant coefficient. Each column also adds controls for financial sentiment in newspaper articles to account for sentiment in conventional media as measured by Shapiro (2020). We find that lower sentiment overnight predicts lower stock returns the following business day. In terms of magnitude a one-standard-deviation increase overnight in the TFSI, everything else equal, leads to a decrease of about 6 basis point in daily S&P 500 index returns.[3]

The TFSI also correlates contemporaneously with stock returns. Table 2 presents the results of regressing daily S&P 500 index returns on the contemporaneous observation of the TFSI (measured between 4pm at date $t-1$ and 4pm at date t) and the same control variables as in Table 1, excluding overnight returns. The contemporaneous TFSI also displays a negative and

---

[3]We build and backtest a trading strategy that conditions long or short trades on the S&P 500 index on a threshold value for overnight TFSI (e.g., buy at open and sell at close if overnight sentiment is positive and vice versa if sentiment is negative). We find that such strategy outperforms a simple benchmark that goes long daily on the S&P 500. Results are available upon request.

significant coefficient which implies that the worse the Twitter sentiment, the lower the daily S&P 500 index returns—a one-standard-deviation increase in the TFSI, everything else equal, corresponds to a decrease of about 10 basis point in daily aggregate returns. We do not find a statistically significant relation to aggregate market returns using one-day-lagged TFSI as a regressor (results not shown).[4]

---

[4]We test the assumption that the residuals of all models in tables 1 and 2 are i.i.d. (White, 1980), and we find that the assumption is rejected for all models, excluded model (2) that controls for news sentiment. To account for the role of heteroskedasticity on the uncertainty around the models' estimated coefficients, we report HAC-robust standard errors (Newey and West, 1987).

Table 1

	Dependent variable:				
	$SP500_{t_{(9am)}\rightarrow t_{(4pm)}}$				
	(1)	(2)	(3)	(4)	(5)
$TFSI_{t-1_{(4pm)}\rightarrow t_{(9am)}}$	-0.05**	-0.06***	-0.06***	-0.06***	-0.06***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$News_t$		-0.24**	0.01	0.01	0.01
		(0.10)	(0.15)	(0.15)	(0.15)
$FOMC_t$				0.06	0.07
				(0.08)	(0.08)
HML_{t-1}					-0.04
					(0.03)
SMB_{t-1}					-0.004
					(0.04)
MOM_{t-1}					-0.06***
					(0.02)
$SP500_{t-1}$			-0.04*	-0.04*	-0.05**
			(0.03)	(0.03)	(0.03)
$SP500_{t-1_{(4pm)}\rightarrow t_{(9am)}}$			0.52***	0.52***	0.52***
			(0.07)	(0.07)	(0.07)
VIX_{t-1}			0.01	0.01	0.01
			(0.006)	(0.006)	(0.006)
Constant	0.03**	0.03**	-0.11	-0.11	-0.11
	(0.02)	(0.02)	(0.10)	(0.10)	(0.10)
Observations	2,956	2,955	2,955	2,955	2,955
Adjusted R^2	0.002	0.004	0.08	0.08	0.08

Note: *p<0.1; **p<0.05; ***p<0.01

This table regresses returns of the S&P 500 index on a twitter based measure of "overnight" sentiment and an expanding set of controls used in the literature to forecast stock market returns:

$$SP500_{t_{(9am)}->t_{(4pm)}} = \alpha + \beta\, TFSI_{t-1_{(4pm)}->t_{(9am)}} + \gamma' X_t + \varepsilon_t$$

$News_t$ represents the sentiment in official news sources as calculated by Shapiro (2020). $SP500_{t-1}$ are the daily Standard and Poor's 500 index market returns. $TFSI_{t-1_{(4pm)}->t_{(9am)}}$ is our sentiment index from $4pm_{t-1}$ to $9am_t$. $FOMC_t$ is a binary variable indicating if the day in question was an FOMC meeting day. The variables HML_t, and SMB_t, represent the High-minus-low and Small-minus-big Fama-French factors (1993) respectively. The variable MOM_t represent the momentum factor as defined by Cahart (1997). $SP500_{t-1}$ denotes Standard and Poor's 500 index market returns from the previous day. $SP500_{t-1_{(4pm)}->t_{(9am)}}$ denotes Standard and Poor's 500 index market returns from the close of the market in the previous day to the opening of the market in the current day. VIX_t represent the implied volatility index from the Chicago Board Options Exchange. The sample goes from September 2011 until April 2023. The table reports HAC-robust standard errors for all coefficient estimates (in parentheses).

Table 2

	Dependent variable:				
	$SP500_t$				
	(1)	(2)	(3)	(4)	(5)
$TFSI_t$	-0.10***	-0.12***	-0.10***	-0.10***	-0.09***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
$News_t$		-0.46***	-1.66***	-1.66***	-1.58***
		(0.14)	(0.24)	(0.24)	(0.24)
$FOMC_t$				0.15	0.17
				(0.11)	(0.11)
HML_t					-0.08*
					(0.05)
SMB_t					0.21***
					(0.05)
MOM_t					-0.19***
					(0.03)
$SP500_{t-1}$			-0.17**	-0.17***	-0.18**
			(0.07)	(0.07)	(0.02)
VIX_t			-0.05***	-0.05***	-0.04***
			(0.010)	(0.010)	(0.010)
Constant	0.06***	0.05**	0.87***	0.86***	0.84***
	(0.02)	(0.02)	(0.16)	(0.16)	(0.17)
Observations	2,956	2,955	2,955	2,955	2,955
Adjusted R^2	0.01	0.04	0.04	0.09	0.09

Note: *p<0.1; **p<0.05; ***p<0.01

Note: This table regresses returns of the S&P 500 index on the daily TFSI (a twitter based measure of financial sentiment) and an expanding set of controls used in the literature to forecast stock market returns:

$$SP500_t = \alpha + \beta_1 TFSI_t + \beta_2 FOMC_t + \beta_3 X`_t + \beta_4 Returns_{t-1} + \beta_5 VIX_t + \varepsilon_t$$

$News_t$ represents the sentiment in official news sources as calculated by Shapiro (2020). $SP500_{t-1}$ are the daily Standard and Poor's 500 index market returns. $TFSI_t$ is the daily simple average of all unique finance-related and credit-related tweets on a scale from -1 to 1. $FOMC_t$ is a binary variable indicating if the day in question was an FOMC meeting day. The variables HML_t, and SMB_t, represent the High-minus-low and Small-minus-big Fama-French factors (1993) respectively. The variable MOM_t represent the momentum factor as defined by Cahart (1997). $SP500_{t-1}$ denotes Standard and Poor's 500 index market returns from the previous day. VIX_t represent the implied volatility index from the Chicago Board Options Exchange. The sample goes from September 2011 until April 2023. The table reports HAC-robust standard errors for all coefficient estimates (in parentheses).

16

4.3 TFSI and Monetary Policy

We find that the tweets in our sample relate strongly to federal reserve communications in and around FOMC days. Figure 6 shows two word clouds obtained from tweets in our sample. In such diagrams, the size of the words displayed is proportional to the word's frequency in the body of text. On the left we show the word cloud across all the tweets in our sample, and on the right the word cloud on FOMC days. Words associated with Federal Reserve communication are clearly displayed more prominently in the FOMC-days-only word cloud, which suggests that the twitter discourse in our sample on FOMC days is driven by monetary policy decisions.

Figure 6: Frequent Words: All Sample and on FOMC Days
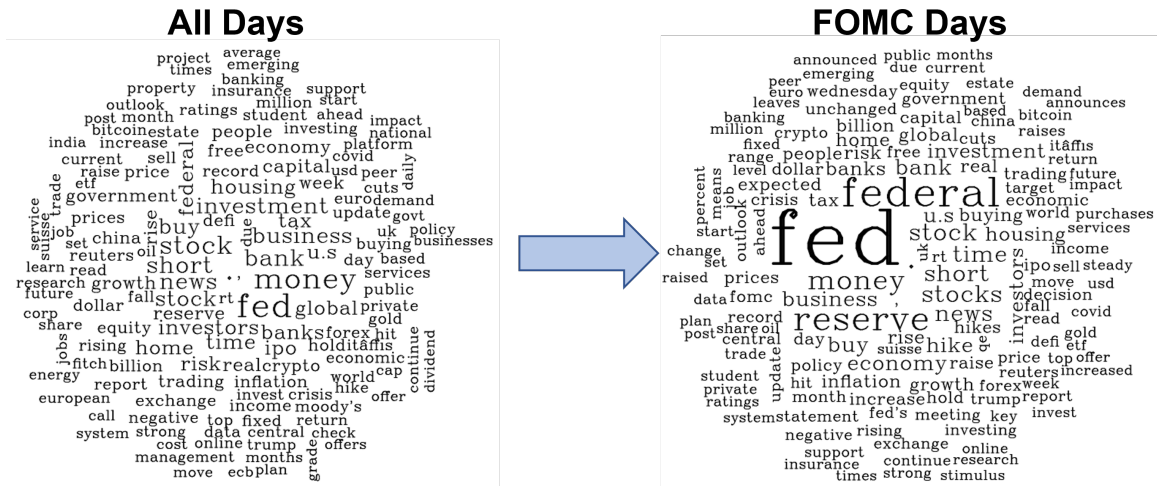
It is also worth noting that in proximity of an FOMC meeting the prevalence of tweets related to the Fed and to monetary policy increases. Figure 7 plots the average share of Fed-related tweets in our sample against the number of calendar days away from the second day of the FOMC meeting. On FOMC days the share of fed related tweets is about 25 percent on average. The share remains significantly above average, between one day before and 5 days after the FOMC meeting, reverting back close to its sample mean of 12 percent (the dashed line).

With these observations at hand, we study how Twitter sentiment behaves ahead and after monetary policy decisions. We find that the TFSI helps predict the size of restrictive monetary policy surprises, while it is uninformative on the size of easing shocks.

Twitter sentiment measured ahead of the release of the official monetary policy determina-

17

Figure 7: Average share of Fed-related tweets against calendar days away from FOMC date.
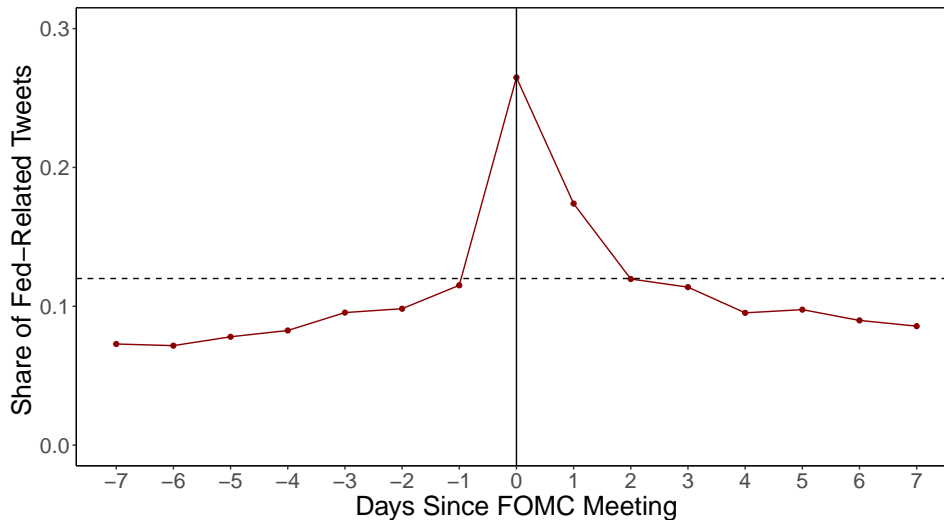
Note: The red line plots the daily share of Fed-related tweets defined in Appendix A, one week before and one week after an FOMC statement release. The dashed line represents the full sample average of Fed-related tweets standing at 12 percent.
Source: TFSI: Authors' calculation based on Twitter enterprise-level API data

tions of the Federal Open Market Committee (FOMC) can predict the size of restrictive monetary policy shocks as gauged by event-study monetary policy shocks. Our finding holds across three measures of monetary policy shocks that control for the central bank information effect—or changes in policymakers' assessment of the macroeconomic outlook conveyed by the policy statement: (Miranda-Agrippino and Ricco, 2021, henceforth MAR), (Jarociński and Karadi, 2020, henceforth JK), and (Bauer and Swanson, 2022, henceforth Bauer and Swanson).[5] Our results imply that tweeted financial sentiment ahead of monetary policy decisions contains information that can help predict the market reaction around the FOMC statement release.

Tables 3 regresses three different measures of monetary policy surprises on the TFSI index value measured over the time window between 4pm the day before the FOMC statement release

---

[5]MAR shocks are computed from 30-minute-window changes in the 2-year on-the-run Treasury Yield around policy announcements over a sample that starts in (sample: September 2011 to December 2022). JK shocks are computed from 30-minute-window changes in the three month ahead monthly Fed Funds futures (FF4) quotes around policy announcements, limiting the sample to those episodes in which the sign of the FF4 surprise and SP500 surprise have the opposite sign (sample: September 2011 to December 2019). Bauer and Swanson's shocks are computed from 30-minute-window changes in the FF4 quotes around policy announcements orthogonalized with respect to macroeconomic and financial data that pre-date the announcement (sample: September 2011 to December 2022).

and 2pm (excluded) on the day of the release. The first column of each table uses all monetary policy shocks in the sample, while the second and third columns split the sample in restrictive and easing shocks, respectively. The first columns of each table suggests that no systematic significant correlation holds between monetary policy shocks and values of the TFSI, for any of the three types of monetary policy shocks. After splitting the sample into tightening and easing shocks, however, the second columns reveal that the TFSI ahead of the policy announcement is a significant predictor of the size of restrictive monetary policy shocks, while this is not the case ahead of easing shocks. In other words, the TFSI increases (and sentiment sours) ahead of tighter monetary policy shocks.[6] Unexpected monetary policy moves—that should be unforecastable—are in fact debated in the Twitter conversation and affect its sentiment ahead of FOMC decisions. A negativity bias (Baumeister et al., 2001) seems be at play by which the anticipation of a negative outcome (a monetary policy tightening) is more likely to be reflected in Twitter sentiment relative to the anticipation of a positive outcome (a monetary policy easing).

We also represent these results graphically for two out of the three series of publicly available shocks. Figure 8 plots the size of the monetary policy shocks of JK (top), and Bauer and Swanson (bottom) on the x axis against the TFSI on the y axis. As expected, we find a statistically significant relation between our measure of sentiment and the different gauges of tightening shocks. Larger contractionary monetary policy shocks are associated with souring sentiment— an increase in our measured sentiment values. Easing monetary policy shocks, however, do not elicit an improvement in sentiment.

Finally, we study whether the size of monetary policy shocks affect sentiment after the release of the FOMC statement. Models in Columns 1, 3, and 5 of Table 4 regress the TFSI after the monetary policy announcement on the full set of monetary policy shocks, tightening shocks, and easing shocks respectively. Columns 2, 4 and 6 also add the TFSI before the statement release as an additional control to the univariate models. Columns 3 of Table 4 suggest that the TFSI— measured between 2PM and 4PM on the day of the policy announcement—responds significantly to unexpected tightening in the policy stance across all three shock measures, but this effect weakens once we control for twitter sentiment measured before the policy statement release (Column 4). Columns 2, 4, and 6 suggest that sentiment ahead of the policy announcement is a

---

[6]The statistical significance of these results is preserved after controlling for financial sentiment in media, as measured by Shapiro (2020), the returns of the SP500 and the level of the VIX index. Results are available upon request.

Table 3: TFSI and Monetary Policy Shocks — Prediction

MAR shocks

	Dependent variable:		
	$MAR\,Shocks_t$		
	All	Tight	Ease
$TFSI_{t-1_{(4pm)}\rightarrow t_{(1:59pm)}}$	0.03	0.51***	-0.06
	(0.10)	(0.12)	(0.16)
Constant	0.001	-0.00	0.00
	(0.10)	(0.12)	(0.16)
Observations	93	52	41
Adjusted R^2	-0.01	0.24	-0.02

JK shocks

	Dependent variable:		
	$JK\,Shocks_t$		
	All	Tight	Ease
$TFSI_{t-1_{(4pm)}\rightarrow t_{(1:59pm)}}$	-0.13	0.58***	-0.21
	(0.13)	(0.15)	(0.18)
Constant	0.00	-0.00	-0.00
	(0.13)	(0.15)	(0.18)
Observations	62	30	32
Adjusted R^2	0.001	0.32	0.01

BS shocks

	Dependent variable:		
	$Bauer-Swanson\,Shocks_t$		
	All	Tight	Ease
$TFSI_{t-1_{(4pm)}\rightarrow t_{(1:59pm)}}$	0.07	0.36**	0.005
	(0.12)	(0.16)	(0.20)
Constant	-0.003	-0.00	0.00
	(0.13)	(0.16)	(0.19)
Observations	64	36	28
Adjusted R^2	-0.01	0.10	-0.04

Note:	*p<0.1; **p<0.05; ***p<0.01

Note: The first column regresses monetary policy shocks on TFSI measured ahead of the monetary policy announcement, according to the specification: $MPS_t = \alpha + \beta\,TFSI_{[t-1](4pm-1:59pm)} + \varepsilon_t$. The second and third columns restrict the sample to consider only tightening and easing shocks respectively. JK and Bauer and Swanson shocks are available for the sample period Sept. 2011- Dec. 2019. MAR shocks are for the 2-year on-the-run Treasury yield and constructed for the sample period Sept. 2011 - Dec. 2022. Note that MAR shocks beyond 2018 are currently confidential and available only within the Federal Reserve System. The tables reports OLS standard errors for all coefficient estimates (in parentheses).

Figure 8: Sentiment on FOMC days vs Monetary Policy Shocks

Jarociński and Karadi (2020) shocks

Bauer and Swanson (2022) shocks
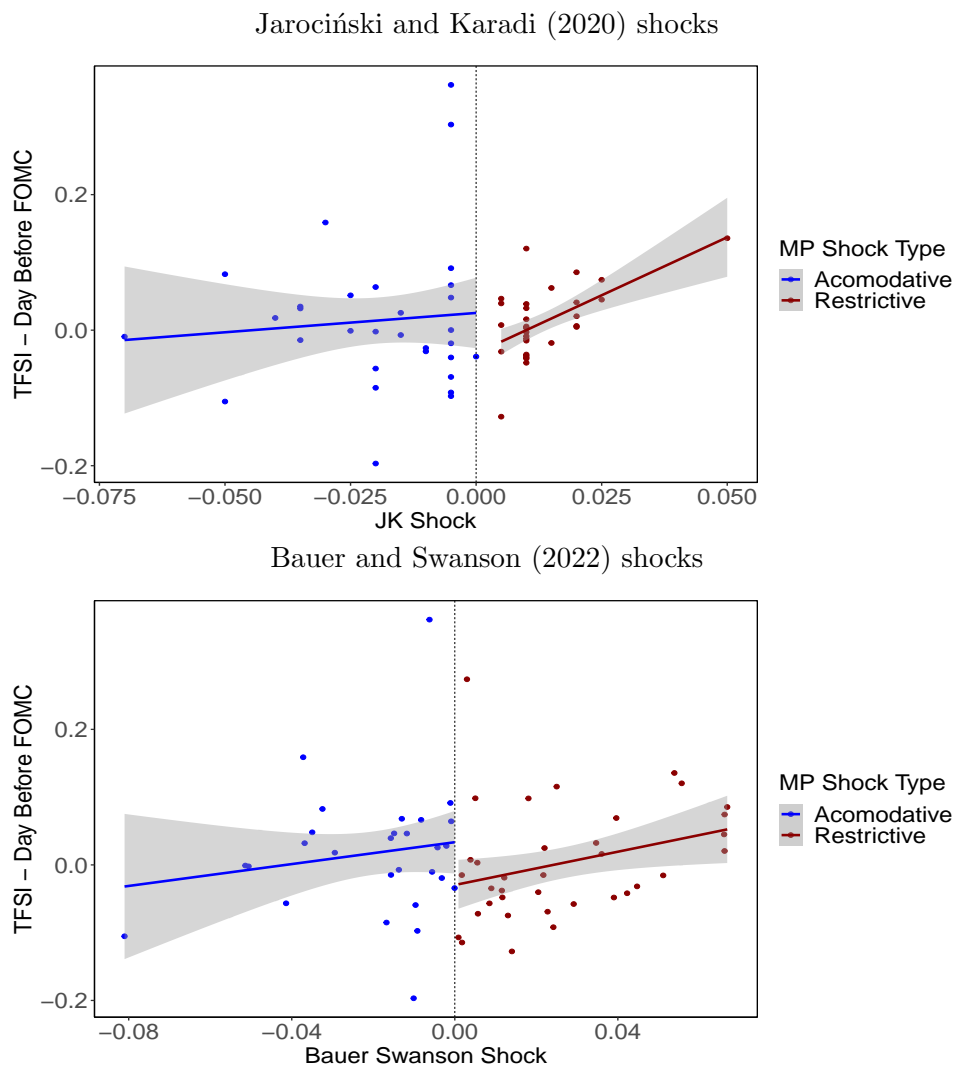
Note: The top panel displays the Jarocinski and Karadi (2020) monetary policy shocks against the TFSI computed between 4pm the day before and 2pm the day of the monetary policy decision (sample Sept. 2011 - Dec. 2019). The bottom panel displays the Bauer and Swanson (2022) monetary policy shocks against the TFSI computed between 4pm the day before and 2pm the day of the monetary policy decision (sample Sept. 2011 - Dec. 2019).

Source: TFSI: Authors' calculation based on Twitter enterprise-level API data; Jarocinski and Karadi (2020); Bauer and Swanson (2022)

significant predictor of sentiment after the policy announcement, independently of the sign of the monetary policy shock.[7] Our findings suggest that easing monetary policy shocks have no effect on the TFSI, while Twitter financial sentiment deteriorates both ahead and after a tightening monetary policy shock.

5   Conclusions

We build a real-time Financial Sentiment Index applying sentiment analysis to a query of tweets related to financial- and credit-market dictionaries. We find that changes in users' engagement–rather than in average tweeted sentiment–drives most variation in the index, that Twitter financial sentiment correlates highly with market-based measures of financial conditions and that overnight Twitter sentiment helps predict daily stock market returns. We document that Fed-related tweets play a dominant role on FOMC days and that sentiment deteriorates ahead of unexpected contractionary changes in the monetary policy stance. We also document that sentiment deteriorates further with the size of unexpected monetary policy tightening, while the relationship between sentiment and monetary policy accommodation is muted.

---

[7]The statistical significance of these results is preserved after controlling for financial sentiment in media, as measured by Shapiro (2020), the returns of the SP500 and the level of the VIX index.

Table 4: TFSI and Monetary Policy Shocks — Delayed Response

MAR shocks

	All	All	Tight	Tight	Ease	Ease
			Dependent variable:			
			$TFSI_{t_{(2pm)} \rightarrow t_{(4pm)}}$			
$MARShocks$	-0.01	-0.06	0.45***	0.35**	-0.01	0.06
	(0.11)	(0.10)	(0.13)	(0.14)	(0.16)	(0.14)
$TFSI_{t-1_{(4pm)} \rightarrow t_{(1:59pm)}}$		0.47***		0.24*		0.55***
		(0.09)		(0.14)		(0.14)
Constant	-0.01	-0.01	0.00	0.00	0.00	-0.00
	(0.11)	(0.10)	(0.12)	(0.12)	(0.16)	(0.13)
Observations	93	93	52	52	41	41
Adjusted R^2	-0.01	0.20	0.19	0.22	-0.03	0.26

JK shocks

	All	All	Tight	Tight	Ease	Ease
			Dependent variable:			
			$TFSI_{t_{(2pm)} \rightarrow t_{(4pm)}}$			
$JKShocks$	-0.06	0.04	0.41**	0.30*	-0.30*	-0.09
	(0.13)	(0.10)	(0.17)	(0.17)	(0.17)	(0.14)
$TFSI_{t-1_{(4pm)} \rightarrow t_{(1:59pm)}}$		0.61***		0.35*		0.67***
		(0.10)		(0.17)		(0.14)
Constant	0.00	0.00	-0.00	-0.00	-0.00	-0.00
	(0.13)	(0.10)	(0.17)	(0.16)	(0.17)	(0.13)
Observations	62	62	30	30	32	32
Adjusted R^2	-0.01	0.35	0.14	0.23	0.06	0.46

BS shocks

	All	All	Tight	Tight	Ease	Ease
			Dependent variable:			
			$TFSI_{t_{(2pm)} \rightarrow t_{(4pm)}}$			
$Bauer-SwansonShocks$	0.26**	0.23*	0.39**	0.35**	0.01	0.07
	(0.13)	(0.12)	(0.16)	(0.14)	(0.20)	(0.18)
$TFSI_{t-1_{(4pm)} \rightarrow t_{(1:59pm)}}$		0.49***		0.45***		0.42**
		(0.12)		(0.14)		(0.18)
Constant	0.11	0.06	0.00	-0.00	0.00	0.00
	(0.13)	(0.12)	(0.16)	(0.14)	(0.19)	(0.18)
Observations	64	64	36	36	28	28
Adjusted R^2	0.05	0.24	0.13	0.32	-0.04	0.11

Note: *p<0.1; **p<0.05; ***p<0.01

Note: The first two columns regress the TFSI after the release of the FOMC statement between 2pm and 4pm on monetary policy shocks, over all FOMC meeting since 2011 (columns 1 and 2) and separated by positive (columns 3 and 4) and negative shocks (columns 5 and 6). The even-numbered columns include the value of the TFSI before the release of the FOMC statement. The tables report OLS standard errors for all coefficient estimates (in parentheses).

References

Angelico, C., J. Marcucci, M. Miccoli, and F. Quarta (2022). Can we measure inflation expectations using twitter? *Journal of Econometrics 228*(2), 259–277.

Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *CoRR abs/1908.10063*.

Azar, P. and A. W. Lo (2016). The wisdom of twitter crowds: Predicting stock market reactions to fomc meetings via twitter feeds. *The Journal of Portfolio Management 42*, 123 – 134.

Baker, M. and J. Wurgler (2000). The equity share in new issues and aggregate stock returns. *The Journal of Finance 55*(5), 2219–2257.

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

Bauer, M. D. and E. T. Swanson (2022, April). A reassessment of monetary policy surprises and high-frequency identification. (29939).

Baumeister, R. F., E. Bratslavsky, C. Finkenauer, and K. D. Vohs (2001). Bad is stronger than good. *Review of General Psychology 5*(4), 323–370.

Bekaert, G., M. Hoerova, and M. Lo Duca (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics 60*(7), 771–788.

Bianchi, F., T. Kind, and H. Kung (2019, September). Threats to central bank independence: High-frequency identification with twitter. (26308).

Binder, C. (2021). Presidential antagonism and central bank credibility. *Economics & Politics 33*(2), 244–263.

Calomiris, C. W. and H. Mamaysky (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics 133*(2), 299–336.

Camous, A. and D. Matveev (2021, 02). Furor over the Fed: A President's Tweets and Central Bank Independence. *CESifo Economic Studies 67*(1), 106–127.

Conti-Brown, P. and B. D. Feinstein (2020). Twitter and the federal reserve. *Brookings Center on Regulation Markets Working Paper*.

Correa, R., K. Garud, J. M. Londono, and N. Mislang (2020, 04). Sentiment in central banks' financial stability reports. *Review of Finance 25*(1), 85–120.

Danielsson, J., M. Valenzuela, and I. Zer (2020). The impact of risk cycles on business cycles: A historical view. *Available at SSRN 3706143*.

Devlin, J., M. Chang, K. Lee, and K. Toutanova (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR abs/1810.04805*.

Ehrmann, M. and A. Wabitsch (2022). Central bank communication with non-experts – a road to nowhere? *Journal of Monetary Economics 127*, 69–85.

Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics 22*(1), 3–25.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Favara, G., S. Gilchrist, K. F. Lewis, and E. Zakrajšek (2016). Updating the recession risk and the excess bond premium. *FEDS Notes*.

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

Gertler, M. and P. Karadi (2015, January). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics 7*(1), 44–76.

Gilchrist, S. and E. Zakrajšek (2012, June). Credit spreads and business cycle fluctuations. *American Economic Review 102*(4), 1692–1720.

Greenwood, R. and S. G. Hanson (2013, 04). Issuer Quality and Corporate Bond Returns. *The Review of Financial Studies 26*(6), 1483–1525.

Hutto, C. and E. Gilbert (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media 8*(1), 216–225.

Jarociński, M. and P. Karadi (2020, April). Deconstructing monetary policy surprises—the role of information shocks. *American Economic Journal: Macroeconomics 12*(2), 1–43.

Korhonen, I. and E. Newby (2019). Mastering central bank communication challenges via twitter. (7/2019).

Lettau, M. and S. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance 56*(3), 815–849.

Loughran, T. and B. McDonald (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research 54*(4).

López-Salido, D., J. C. Stein, and E. Zakrajšek (2017, 05). Credit-market sentiment and the business cycle. *The Quarterly Journal of Economics 132*(3), 1373–1426.

Lüdering, J. and P. Tillmann (2020). Monetary policy on twitter and asset prices: Evidence from computational text analysis. *The North American Journal of Economics and Finance 51*(C), S1062940818302055.

Malo, P., A. Sinha, P. Takala, P. Korhonen, and J. Wallenius (2014, 04). Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the American Society for Information Science and Technology*.

Masciandaro, D., O. Peia, and D. Romelli. Central bank communication and social media: From silence to twitter. *Journal of Economic Surveys n/a*(n/a).

Meinusch, A. and P. Tillmann (2017, December). Quantitative Easing and Tapering Uncertainty: Evidence from Twitter. *International Journal of Central Banking 13*(4), 227–258.

Miranda-Agrippino, S. and H. Rey (2020, 05). U.S. Monetary Policy and the Global Financial Cycle. *The Review of Economic Studies 87*(6), 2754–2776.

Miranda-Agrippino, S. and G. Ricco (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics 13*(3), 74–107.

Moody's. Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity . *retrieved from FRED, Federal Reserve Bank of St. Louis [mnemonic: BAA10Y]*.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Qiu, L. and I. Welch (2004, September). Investor sentiment measures. Working Paper 10794, National Bureau of Economic Research.

Shapiro, Adam Hale ;Sudhof, M. W. D. (2020). Measuring news sentiment. *Federal Reserve Bank of San Francisco Working Paper*.

Shiller, R. J. (2015a). Irrational exuberance. *Princeton University Press*.

Shiller, R. J. (2015b). *Irrational Exuberance: Revised and Expanded Third Edition* (REV - Revised, 3 ed.). Princeton University Press.

S&P Global. Capital IQ Pro Platform.

Stiefel, M. and R. Vivès (2019, February). "Whatever it Takes" to Change Belief: Evidence from Twitter. AMSE Working Papers 1907, Aix-Marseille School of Economics, France.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance 62*(3), 1139–1168.

Tillmann, P. (2020). Trump, twitter, and treasuries. *Contemporary Economic Policy 38*(3), 403–408.

Twitter, Inc. Twitter enterprise-level API. *https://twitter.com*.

University of Michigan. University of Michigan Consumer Sentiment Index. *retrieved from FRED, Federal Reserve Bank of St. Louis [mnemonic: UMCSENT]*.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.

Yamada, I., A. Asai, H. Shindo, H. Takeda, and Y. Matsumoto (2020, 01). Luke: Deep contextualized entity representations with entity-aware self-attention. pp. 6442–6454.

A Technical Appendix

A.1 Background

Twitter data has multiple advantages over traditional data. First, it has the possibility to allow researchers to understand perceptions of financial conditions in the US by a broader set of participants compared to surveys of professional market participants and financial market prices. Twitter data enables researchers to analyze a mixed sample of expert and non-expert commentary. There are 35 million Twitter Daily Active Users generating more than 500 million tweets per day. Second, the Twitter platform can also act as an information transmission channel allowing institutions like the Federal Reserve to reach out to otherwise inaccessible sectors of the population. Third, given the character limit of each tweet, users are compelled to communicate in succinct messages to convey their ideas. Lastly, this real-time submission of comments allows for high frequency analysis. All these characteristics can make for a better understanding of the opinion of "Main Street" about financial conditions in the US.

However, certain challenges arise when working with Twitter data. The high volume of tweets available on different topics raises the challenge to obtain a pertinent set of tweets to the research question. Researchers use Google-like keyword searches to filter tweets, and, without careful consideration, can lead to overly broad or overly narrow data samples. In addition to this, most Twitter user accounts are not verified, allowing for robot-generated tweets inside the data. Furthermore, due to its relatively recent founding, in 2006, Twitter data analysis is only possible for most recent years. The number of tweets remains low until the user-base reached a critical mass. Given the challenges mentioned above, we designed a carefully curated query to download a feasible set of finance-related tweets using multiple Natural Language Processing tools, such as heuristic operators ('AND','OR') and semantic similarity between words.

A.2 How to Construct the Query.

We extracted finance-related tweets by using a multi=step financially intrinsic query. First, we obtain a list of financial words from Calomiris and Mamaysky (2019) and Danielsson et al. (2020). Given the sheer volume of tweets we would download by searching for all the words in this list, we use Search Engine Optimization (SEO) Keyword Clustering to group keywords that are semantically similar into three groups. Similar processes are used by SEO engineers

to ensure that search engines, like Google, show their websites amongst the most relevant results to a specific query or group of queries. This technique allows us to find financial tweets that are semantically relevant to our desired subject-matter. We assume that words which are semantically similar are interchangeable and can be separated by the Boolean operator, 'OR'. Second, we use a pre-trained machine learning model called Wiki2Vec to group our keywords by semantic similarity. This model, trained on the corpus of Wikipedia articles, converts words into computer-understandable vectors that allows to calculate word distance. We take these calculated similarities to determine optimal keyword groups, or clusters. We then generate three matrices based on 100, 300, and 500 vectors for each word to find similarity scores of each word compared to all other words in our dictionary. We highlight keyword pairs that have a similarity score higher than the matrix average to categorize word clusters. Based on our analysis and API limitations for monthly download of tweets, we chose to use three clusters.

We modify these clusters to allow for common phrases to be built from them. The first group, containing words like "bond", "security", "asset", or "currency", became the "noun/object" group. The second group is composed of "actors" or "subject" nouns, i.e., "company", "corporate", "market", or "Federal Reserve". The third group of words is comprised of "modifier" words or compounds, like "coupon", "term", "upgrade", and "default". The full set of words inside these groups can be found in Appendix A.

We eliminate from our query tweets that refer to advertising by excluding tweets that contain the specific phrases "social security" or "credit card". Furthermore, we do not download exact retweets to maximize our download quota. Regardless of this loss of retweets as individual observations, we obtain the retweet count and other "engagement" counts of each original tweet.

We extract the main informational set of each tweet, the text and date of publication, as well as their user metadata: username, verification status, and location, if available. We also have access to extended metadata related to each tweet including the aforementioned retweet count and other engagement counts, other tweets to which it might be in reply, and an assessment whether the tweet contains "possibly sensitive" or "NSFW" content. So far, we have processed 7.1 million tweets since the advent of Twitter, with an average increase of nearly 450,000 tweets per year. Due to the narrower scope of our search than general Twitter usage, there were few applicable results in the early days of Twitter. Fewer than 5,000 tweets were posted prior to 2009, with a little over 45,000 posted in 2009, and almost 150,000 posted in 2010. We choose to start our Twitter Financial Sentiment Index in 2011 given that more than 250,000 tweets were

posted. The number of tweets that falls within our query has gone up on average ever since. The month with the highest number of tweets downloaded is October 2022, with about 188,000 tweets, while the month with the lowest number of tweets is February 2011, with around 10,000 tweets.

A.3   Processing the Raw Data

Generating the Twitter Financial Sentiment Index requires a few steps after obtaining the data. First, we preprocess the text of each tweet by removing "@" tags, excess white space, hyperlinks, and replace common ascii plain text with the appropriate special character, i.e. "&amp" becomes "&". We also eliminate from our sample of tweets any tweet that refers to cryptocurrency and decentralized financial assets by removing tweets that contain the text strings "crypto","token","NFT","inu","shiba", and "defi". In addition to this initial text cleaning, we separate tweets with exact duplicate text into another data set to later be included in engagement counts on each unique tweet. Second, we shift the time zone of tweets to reflect Eastern Standard Time (EST) instead of the default Coordinated Universal Time (UTC) for easier comparison to United States events. Third, we calculate the sentiment values of each tweet in our sample using the Bidirectional Encoder Representations from Transformers (BERT) sentiment. Fourth, we set to zero the sentiment of tweets that contain values between -.1 and .1. Lastly, we obtain an index by taking the simple mean of all the remaining tweets from 2011 to 2022 by different frequencies– daily, weekly, and monthly.

A notable characteristic of our sample is the presence of tweets with the same text content and sentiment score. We address these tweets as "manual" retweets. Though our query explicitly excludes retweets from appearing in our search, this only prevents retweets that were made by clicking the retweet button. Oftentimes a manual retweet is done by bots, or by users sharing the exact same content from a non-twitter website, like news articles, editorials, and blog posts. To control for this, we remove tweets if the text is duplicated in a previous tweet, keeping the first incidence. There are 2M manual retweets in our sample. After removing, our sample size is 5.1M tweets. We then proceed to calculate the sentiment of each tweet.

A.4 FinBERT

We use Google's Bidirectional Encoder Representations from Transformers (BERT) as our starting point to analyze the sentiment of tweets. BERT is a state-of-the-art pre-trained machine learning model capable of understanding sentences alongside the context in which they are being applied. BERT is pre-trained on the Toronto BookCorpus (containing 800M words) and Wikipedia articles (containing 2.5B words). BERT converts words into vectors, and reads the text bidirectionally to classify sentences given the context in which words are being used. This unique ability to understand contextual representation, and doing so in both directions of the text allows BERT to significantly outperforms other machine-learning-based and dictionary-based models in tasks like text prediction and sentiment calculation. Furthermore, it can be pre-trained further and then fine-tuned to better understand a desired context, like financial jargon.

We use the model FinBERT as our baseline for sentiment scoring. FinBERT, from Araci (2019), is a refined version of BERT that is designed to understand text in the context of Financial sentiment. FinBERT is pre-trained using a large corpus of financial texts and fine-tuned with a dictionary of financial words and phrases from Malo et al. (2014). One caveat of FinBERT is that it was pre-trained using longer texts, so it splits sentences individually and then calculates sentiment on each one of them. Given the context of tweets can be better understood as a whole, rather than separated by sentences, we replace full sentence stops, ".", with a semi-colon, ";", to "desentencize" our text before calculating sentiment values on each of our tweets.

FinBERT produces five sentiment values. Three values represent the probabilities that the text is either positive, negative, or neutral. FinBERT also calculates a compound score as the positive probability minus the negative probability. Lastly, FinBERT provides trinary sentiment prediction which is based on the highest of the three probabilities. We drop tweets that are classified as neutral in this prediction (neutral probability is highest). Then, we obtain a sentiment score for each tweet in our sample.

We calculate our Twitter Financial Sentiment Index as the negative of the average the sentiment of all tweets by a given frequency, either daily, weekly, or monthly. The values of this measure range from -1 (extremely positive) to 1 (extremely negative). The monthly index has a mean of -0.09, and a standard deviation of 0.1.

A.5 Valence Aware Dictionary for Sentiment Reasoning

We also analyze a sentiment with a second model as a robustness check to our baseline BERT measure. We derive sentiment values from the Valence Aware Dictionary for sEntiment Reasoning (VADER) model. This dictionary-based sentiment scorer is specifically designed for analyzing sentiment on social media platforms, which tend to be shorter, more cryptic messages than the average texts analyzed. This dictionary is capable of understanding phrases, emoticons, and acronyms like 'XD" and "LOL". Furthermore, VADER can comprehend semantic modifiers from heuristic rules that other dictionaries lack. This key feature of VADER allows the calculation of the magnitude of sentiment, not only its polarity. For example, using VADER, the phrase "really bad" produces a more negative sentiment value than the single word "bad", while the phrase "not bad" produces a positive sentiment value.

Given VADER is a dictionary-based sentiment scorer, we seek to prevent selection bias by neutralizing the sentiment score of all the words from our constructed query. From all the words used in the query, only the words 'asset', 'credit', 'cut', 'debt', 'interest', 'low', 'pay', 'security', 'share', and 'treasury', and their plurals contained a sentiment valence other than zero. We then calculate the positive, neutral, negative, and compound sentiment values of each tweet in our sample.

Once the sentiment score is calculated, we filter out the tweets with an absolute compound score lower than 0.1. From our sample of 5.1M tweets, we removed 1.7M neutral tweets under this threshold. We argue that this tweets that show neutral sentiment are purely informational and serves the purpose of transmitting information to twitter users, rather than expressing sentiment. It is important to note that the distribution of informational tweets overtime is not uniform. From 2011 to 2017, informational tweets formed roughly 50% of the share of all tweets while the share of zero-valued tweets decreased to 25% from 2018 onward. This drastic change in the share of information tweets also came with an increase in the share of positive and negative tweets, from roughly 25% of the share overtime for both, to 50% and 25% of the share overtime respectively. This change in the composition of the sentiment-charged tweets in the sample is due to Twitter's executive decision to increase the character length limit of each tweet that took effect in December 2017. The increase in character limit from 140 to 280 increased the number of average words per tweets from 15.3 to 32.2. This gave more space for demonstrating more sentiment in each tweet by increasing the probability of writing more sentiment charged words. This is also an effect of VADER. We see, in general, that the longer a text that is provided to

VADER, the more polarized the sentiment score becomes.

When comparing to the BERT-based TFSI, we see a noticeable change in the shares of positive, negative, or neutral tweets when the character limit was increased in late 2017. This is likely due to finBERT's superior ability to understand context and account for differences in the number of sentiment-laden words. These shares are relatively consistent over time with roughly 50% of tweets being classified as neutral and the remaining 50% a relatively even split between positive and negative. The linear correlation between the BERT-based and VADER-based indexes is 0.68.

A.6 Twitter-based Financial Sentiment Index

We construct our Financial Sentiment Index by obtaining the simple mean of the compound sentiment score for all remaining after pre-processing steps take place. We do this at weekly, and monthly frequencies. Given the higher level of variance at higher frequencies, we use the 7-day moving average of TFSI values to obtain the daily TFSI.

We found our Twitter Financial Sentiment monthly index is highly correlated to other measures of financial conditions, like the Financing Conditions Index for Non-Financial Corporations, produced at the Federal Reserve Board. It has a correlation of 0.79 with the FCINFC. Also, it responds congruently to negative and positive shocks. The average sentiment across time is 0.12 and the standard deviation is 0.07. The volatility of earlier years in the index is higher, arguably due to the lack of a meaningful number of observations per month. There are two significant lows since the character limit increase, one in September 2019, and the other in March 2020.

A.7 Fed-worded Tweets on FOMC days

We also seek to understand the transmission of information on the Federal Reserve via tweets around events directly related to it. To do this, we analyze the traffic of Tweets that are related to the Federal Reserve around the Federal Open Markets Committee events. First, we generated a subsample of tweets that contained the text string 'Fed','Reserve','monetary', 'Powell', and 'Yellen' respective of their time as Federal Reserve Chairs. The average volume of tweets containing these strings is 65 per day. We found that, on average, this volume increased to 231 tweets 24 hours after an FOMC event took place. This increase in volume of tweets is 150 tweets above the 3rd quartile of the volume of all Fed-worded tweets overtime. This indicates

that Twitter is used as a transmission channel for the information conveyed in FOMC events. We then proceed to analyze the relationship of the frequency of this transmission and monetary policy shocks.

B   Word Clusters

Table 5: Word Clusters

Group 1	Group 2	Group 3
Bond	Corporate	Coupon
Debt	Company	Interest
Security	Subsidiary	Rate
Credit	Market	IPO
Loan	Municipal	Term
Mortgage	Sovereign	Liquidity
Portfolio	Program	Yield
Pension	Market	Downgrade
Federal Funds	Federal Reserve	Outstanding
Leverage	Collateral	Repayment
Financing	Credit Agency	Default
Rent	Sovereign	Initial Public Offering
Portfolio	Credit Agency	Lending
Asset	Program	Return on
Pension	Federal	Upgrade
Facility	Federal Reserve	
CPFF	Money Markets	
PDCF	Collateral	
MMMFLF	Junk	
PMCCF	High Yield	
MLF	Investment Grade	
MSCP	HY	
Fund	IG	
Currency		
Debenture		
Leverage		
Cash		
Finance		
Financial Leverage		