The Wisdom of Twitter Crowds: Predicting Stock Market Reactions to FOMC Meetings via Twitter Feeds

Pablo D. Azar and Andrew W. Lo

This Draft: March 11, 2016

Pablo D. Azar is a PhD student in the Department of Economics and Laboratory for Financial Engineering, Sloan School of Management, MIT. Email: pazar@mit.edu

Andrew W. Lo is Charles E. and Susan T. Harris Professor and the Director of the Laboratory for Financial Engineering, Sloan School of Management, MIT. Email: alo-admin@mit.edu

With the rise of social media, investors have a new tool to measure sentiment in real time. However, the nature of these sources of data raises serious questions about its quality. Since anyone on social media can participate in a conversation about markets—whether they are informed or not—it is possible that this data may have very little information about future asset prices. In this paper, we show that this is not the case by analyzing a recurring event that has a high impact on asset prices: Federal Open Market Committee (FOMC) meetings. We exploit a new dataset of tweets referencing the Federal Reserve and show that the content of tweets can be used to predict future returns, even after controlling for common asset pricing factors. To gauge the economic magnitude of these predictions, the authors construct a simple hypothetical trading strategy based on this data. They find that a tweet-based asset-allocation strategy outperforms several benchmarks, including a strategy that buys and holds a market index as well as a comparable dynamic asset allocation strategy that does not use Twitter information.

Investor sentiment has frequently been considered an important factor in determining asset prices. Traditionally, sentiment is measured by observing analyst estimates, survey data, news stories, and technical indicators such as put/call ratios and relative strength indicators. Two drawbacks of these indicators are that they are based on a relatively sparse subset of the population of investors and, except for technical indicators, are not measured in real time. The rise of social media allows us to overcome these drawbacks and measure the sentiment of a large number of individuals in real time. These data sources give the quantitative investor a new tool with which to construct portfolios and manage risk.

However, because social media data is generated by individual users and not investment professionals, the following questions arise about the quality of this data:

- Do user messages contain relevant information for asset pricing?
- Can this information be inferred from more traditional sources, or is it truly *new* information?
- Can social media data help predict future asset returns and shifts in volatility?

To answer these questions, we focus on a single recurring event that reveals previously unknown information to the market: Federal Open Market Committee (FOMC) meetings. Eight times a year, the FOMC meets to determine monetary policy. The decisions made by the FOMC are highly watched by all market participants, and often have a significant impact on asset prices.[1]

To understand how investors on social media behave around FOMC meeting dates, we create a new dataset of tweets that cite the Federal Reserve. Using natural language processing techniques, we can assign a polarity score to each Twitter message, identifying the emotion in the text. We show that this polarity score can be used to predict the returns of the CRSP Value-Weighted Index, even when limiting ourselves to articles and tweets that are published at least 24 hours *before* the FOMC meeting.

We use these results to construct trading strategies that bet more or less aggressively in a market index depending on Twitter sentiment. We find that portfolios using Twitter data can significantly outperform a passive buy-and-hold strategy.

[1](Bernanke and Kuttner, 2005; Cieslak, Morse, and Vissing-Jorgensen, 2014; Lucca and Moench, 2015).

Literature Review

There are now many studies that use social network data to analyze financial assets. Bar-Haim, Dinur, Feldman, Fresko, and Goldstein (2011) show that some users are "experts", in the sense that they are more consistently right about stock price movement than other users. Identifying these experts is useful for predicting asset movement using tweet data. The question remains on whether these experts are themselves influencing prices (by influencing others' beliefs and trades), or whether their tweets simply reflect existing market information better than non-sophisticated users' tweets. Giannini, Irvine, and Shu (2014) show that local investors are more likely to have accurate information about companies located in their area than non-local investors. Plakandaras, Papadimitriou, Gogas, and Diamantaras (2014) use social media data to predict changes in exchange rates using machine learning methods. Our paper differs from the existing literature in that we look at social media's reaction to a significant and recurring macroeconomic event, instead of analyzing events for individual equities.

Several studies have documented the fact that FOMC decisions have a significant impact on stock prices.[2] The analyses in the existing literature typically focus on easily quantifiable information. For example, Bernanke and Kuttner (2005) show that an unexpected 25-basis-point cut in the federal funds rate target induces a one percent rise in broad stock indices. The "surprise" component of the interest rate policy is measured as the deviation of the announced interest rate from that predicted via a vector auto-regression (VAR) model that an investor could calibrate based on previous decisions. One issue with this type of analysis is that it is not immediate that the VAR model captures investor sentiment, and it is not applicable during long periods of zero interest rates, such as the recent period between December 2008 and December 2015.

One important paper to highlight is that of Lucca and Moench (2015). They show that a significant fraction of the risk premium is earned on the 24 hours before the FOMC decision is announced, and suggests that a strategy that increases its equity holdings on the days before the FOMC decision should outperform the market. Surprisingly, we show that during

[2](Bernanke and Kuttner, 2005; Gurkaynak, Sack, and Swanson, 2005; Cieslak, Morse, and Vissing-Jorgensen, 2014; Lucca and Moench, 2015).

the 2010-2014 period that we study, such a strategy does not outperform,[3] while a strategy that adjusts its positions on FOMC decision days based on tweets performs much better.

Our main contribution is thus to add text data to the number of signals that we can use to measure the effect of the FOMC on markets. Furthermore—and in contrast to existing work on the FOMC and text[4]—we focus on text data generated by investors themselves and not by the FOMC. An advantage of this approach is that the information generated by investors appears before the FOMC announces their decisions, and can be incorporated in trading strategies that seek to anticipate market reactions to the FOMC meeting.

Data

We collected tweets using the Topsy API.[5] We gathered English language tweets between 2007 and 2014 that mentioned the terms "FOMC" or "Federal Reserve". We also gathered English language tweets that mentioned "Bernanke" or "Yellen" depending on whether the tweets were posted during Bernanke's term or Yellen's term.

Tweet sentiment was computed by using a Python package called "Pattern" (De Smedt and Daelemans, 2012). The outcome of this process is that each tweet is associated with a polarity score between -1 and $+1$ (-1 for purely negative and $+1$ for purely positive). We describe the process in slightly more detail in the appendix, and refer the interested reader to De Smedt and Daelemans' guide to the software package (De Smedt and Daelemans, 2012) for a more thorough description. Exhibit 1 shows some example tweets with their polarity scores, including some misclassified tweets that show how the algorithm may sometimes not capture sentiment about the Federal Reserve itself, but about other actors (e.g: "@jordangunderson: is proud that Jason Chaffetz is 1 of 28 Congressmen cosponsoring Ron Paul's Federal Reserve Transparency Act (HR 1207)".)

Each tweet is also associated with the number of followers of the user who posted the tweet. Following Giannini, Irvine, and Shu (2014), we weigh the polarity of each tweet by the number of people following the user who posted the tweet to measure the tweet's

[3]This may be due to the fact that our sample consists mostly of years after which the results of (Lucca and Moench, 2015) became known.

[4](Lucca and Trebbi, 2009; Azar, Li, and Lo, 2015; Jegadeesh and Wu, 2015)

[5]The API we used was originally available at `http://otter.topsy.api`. Unfortunately Topsy is no longer making this API available.

reach. For each day t, we construct a measure of tweet sentiment on day t by averaging the weighted polarities of the tweets on that day. Exhibit 2 summarizes the daily average weighted polarity, as well as the number of tweets and the average unweighted polarity. We also include the z-statistic from the augmented Dickey Fuller test for average polarity and average weighted polarity, to show that these series are likely stationary.

Exhibit 3 shows a time series plot of the average weighted polarity for each day between 2009 and 2014. Exhibit 4 shows the distribution of tweets during the year 2014. The spikes occur at times when the FOMC was meeting. Exhibit 5 shows a smoothed time series for the number of tweets per day between 2007 and 2014. We can see from exhibit 5 that tweets concerning the Federal Reserve were very scarce before 2009. Due to the scarcity of data before 2009, we restrict our study to the 2009–2014 period. We can also see from exhibit 5 that the smoothed series for number of tweets is not stationary, which is likely due to the growth in number of users between 2009 and 2014.

We use WRDS to obtain daily values for market returns and Fama French factors.

Tweet	Polarity
@GStuedler this was caused by the worst regulation of all time, the Federal Reserve Act of 1913. Decoupled money from reality #topprog #tcot	−1.00
@SenJohnMcCain Maybe instead of partisan bickering, you should all come together and go after the real bad guys... The Federal Reserve	−1.00
Bernanke Says Biggest Worry is That Politicians Abandon Banks: (CEP News) - U.S. Federal Reserve Chairman Ben Be.. http://tinyurl.com/cosb9m	−0.75
Treasurys rise as Fed meets: Treasury prices rose Tuesday amid speculation that the Federal Reserve will begin b.. http://tinyurl.com/c42qjb	0.60
RT @jordangunderson: is proud that Jason Chaffetz is 1 of 28 Congressmen cosponsoring Ron Paul's Federal Reserve Transparency Act (HR 1207).	0.80
Very impressed w/Federal Reserve Chair Ben Bernanke!	1.00

Exhibit 1: Example tweets with polarity scores.

Variable	Mean	SD	Min	Max	Median	DF
Num. Tweets Per Day	254.2	328.2	0	3445	158	
Avg. Polarity Per Day	0.0875	0.0761	−0.307	0.5	0.09	−30.348
Avg. Weighted Polarity Per Day	1141.0	5167.322	−33752.71	122957.5	601.2	−38.650

Exhibit 2: Summary statistics. The 'DF' column contains Dickey-Fuller test statistic. $N = 1,507$.

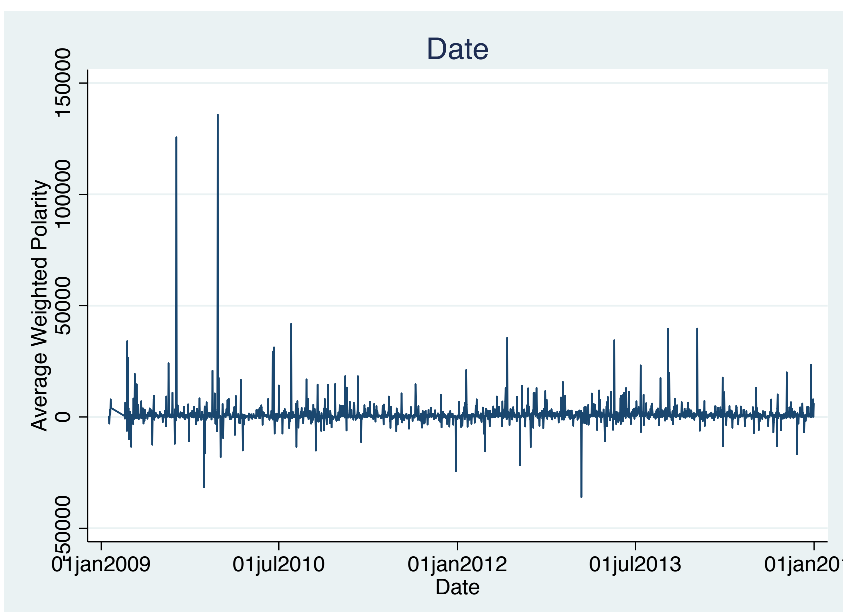

Exhibit 3: Average weighted polarity of tweets over time for the 2009–2014 period.

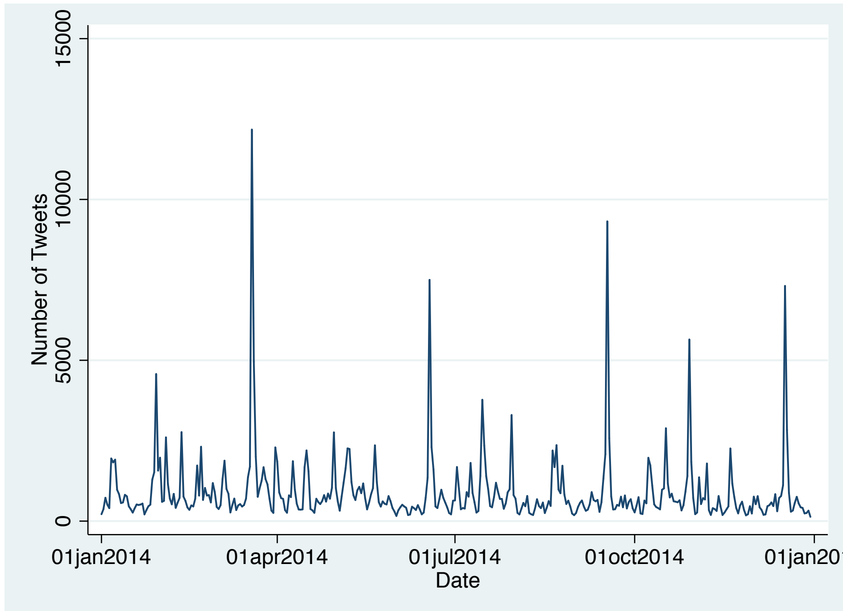

Exhibit 4: Number of tweets over the year 2014 mentioning the key words.

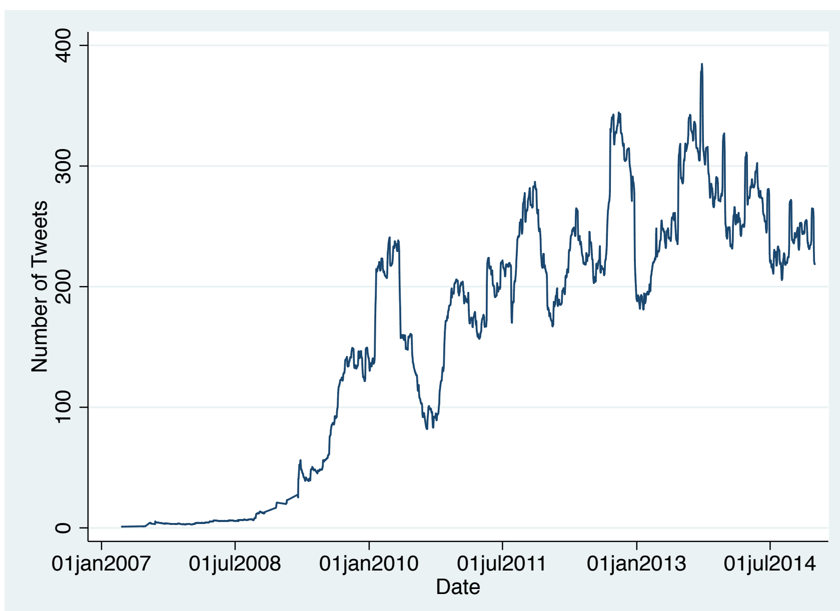

Exhibit 5: Smoothed number of tweets over time for the 2007–2014 period.

Regression Results

We show that tweets can be used to predict future returns and volatility, even after controlling for other factors. Exhibit 6 contains the results of regressions of the form:

$$\begin{aligned} R_t \quad = \quad & \alpha + \beta_1 \text{IndicatorFOMC}_t + \beta_2 \text{TweetPolarity}_{t-1} + \beta_3 \text{TweetPolarityFOMC}_{t-1} + \\ & + \gamma_1 \text{HML}_t + \gamma_2 \text{SMB}_t + \gamma_3 \text{UMD}_t + \gamma_4 R_{t-1} + \epsilon_t \end{aligned} \tag{1}$$

where the dependent variable R_t is the excess daily return (in percentage terms) on the CRSP value-weighted market index. The independent variables are

- IndicatorFOMC_t: a dummy variable equal to one if and only if there was a scheduled FOMC meeting that concluded at date t.

- $\text{TweetPolarity}_{t-1}$: the average weighted polarity of tweets in the time range 4:00pm on date $t-2$ to 4:00pm on date $t-1$. Notice that this variable is observable before market closes on date $t-1$, and can be used by a strategy that determines its date-t position at the market close of date $t-1$. This variable is normalized to have mean zero and variance 1.

- $\text{TweetPolarityFOMC}_{t-1} = \text{IndicatorFOMC}_t \times \text{TweetPolarity}_{t-1}$: this variable is the same as the tweet polarity variable, but is non-zero only on days when the FOMC meets.

- HML_t, SMB_t, UMD_t: the Value, Size and Momentum factors.

- R_{t-1}, the excess market return on date $t-1$.

Exhibit 6 shows four different regressions illustrating the effect of omitting or adding certain variables. From the exhibit we can see that excess returns are significantly higher—by about 0.3 percent—during our sample period when the FOMC meets. We can also see that tweet polarity on date $t-1$ can be used to predict returns on date t, and that this effect intensifies on days when the FOMC meets. Finally, we see that even if we account for contemporaneous movement in factors, tweet polarity on date $t-1$ is still significantly correlated with returns on date t, suggesting that—on days the FOMC meets—tweets have information that is not conveyed by market factors.

One important result to take away from exhibit 6 is that, once we account for the Fama French factors, the effect of Twitter sentiment on returns becomes negligible, *except* on days in which the FOMC meets. On these days, a one standard deviation increase in tweet sentiment will increase returns by 0.62 percent.

In the next section, we show that these regression results can be used to build trading strategies based on tweet information that outperform the market on multiple performance metrics.

VARIABLES	(1) Return	(2) Return	(3) Return	(4) Return
IndicatorFOMC	0.331*	0.338*	0.398**	0.343**
	(0.187)	(0.187)	(0.177)	(0.140)
TweetPolarity		0.0510**	0.0493*	0.0156
		(0.0249)	(0.0252)	(0.0195)
TweetPolarityFOMC			0.490	0.625**
			(0.529)	(0.296)
hml				0.843***
				(0.0757)
smb				0.857***
				(0.0640)
umd				-0.141**
				(0.0602)
L.Return	-0.0722*	-0.0715*	-0.0716*	-0.0334
	(0.0384)	(0.0384)	(0.0384)	(0.0296)
Constant	0.0630**	0.0628**	0.0628**	0.0516**
	(0.0313)	(0.0312)	(0.0313)	(0.0246)
Observations	1,506	1,506	1,506	1,506

Standard errors in parentheses

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

Exhibit 6: Regression of daily returns (in percent) on tweet polarity, FOMC meeting indicator, and Fama French factors. The interaction term is equal to TweetPolarity × IndicatorFOMC. Newey-West standard errors are computed using 4 lags.

Assessing Investment Impact

To develop a sense of the practical value of Twitter sentiment to investors, we simulate a simple trading strategy based on the Kelly Criterion. Suppose we invest in one risky asset and one risk-free asset and rebalance the position every day based on our forecast of tomorrow's risky-asset return. For simplicity, assume that the risk-free rate is $R_{f,t}$, and that the risky asset's returns on date t are drawn from a normal distribution $\mathcal{N}(\mu_t, \sigma_t^2)$ with mean μ_t and variance σ_t^2.

We wish to choose a fraction f_t of our wealth every day to invest in the risky asset (if our time t wealth is w_t, then our total investment in the risky asset will be $f_t \cdot w_t$). Under the assumption of a logarithmic utility function, the optimal investment strategy is to invest the fraction:

$$f_t^* \quad = \quad \frac{\mu_t - R_{f,t}}{\sigma_t^2} \ . \tag{2}$$

This policy also maximizes the expected geometric growth rate of the portoflio. If $f_t^* < 0$ or $f_t^* > 1$, we need to shortsell the risky asset or borrow (or short the riskless asset), respectively. We assume that there is some bound L on the leverage and shortselling available, so that $-L \leq f_t^* \leq L$. We also consider more conservative investment policies by multiplying f_t^* by a constant $\tau \in (0,1)$ to yield "fractional-Kelly" policies. This, in combination with the leverage and shortsales constraints, results in the following portfolio strategy:

$$\tilde{f}_t(\tau) \quad = \quad \begin{cases} \frac{\tau(\mu_t - R_{f,t})}{\sigma_t^2} & \text{if} \quad -L \leq \frac{\tau(\mu_t - R_{f,t})}{\sigma_t^2} \leq L \\ -L & \text{if} \quad \frac{\tau(\mu_t - R_{f,t})}{\sigma_t^2} \leq -L \\ L & \text{if} \quad \frac{\tau(\mu_t - R_{f,t})}{\sigma_t^2} \geq L \ . \end{cases} \tag{3}$$

In our empirical analysis below, τ is set to $\frac{1}{4}, \frac{1}{2}$ and 1.

To compute the portfolio weights, we need a model of returns that produces a prediction of the mean $\hat{\mu}_t$ and variance $\hat{\sigma}_t^2$ of the CRSP return for each date t. We use linear models of the form

$$R_t = \alpha + \beta X_t + \epsilon_t \tag{4}$$

where R_t is the excess return of the CRSP value-weighted market index, in excess of the risk-free rate $R_{f,t}$, and X_t is a vector of signals that are observed before time t. In this case, the model predictions are:

$$\hat{\mu}_t = \mathbb{E}[R_t|X_t] = \alpha + \beta X_t \tag{5}$$

$$\hat{\sigma}_t^2 = \mathrm{Var}[R_t|X_t] = \mathrm{Var}[\alpha + \beta X_t + \epsilon_t] \ . \tag{6}$$

To quantify the investment value of Twitter sentiment, we consider four models, corresponding to different choices of X_t:

Model 1: X_t is an empty vector, so the linear model is just $R_t = \alpha + \epsilon_t$

Model 2: $X_t = (\text{IndicatorFOMC}_t)$

Model 3: $X_t = (\text{IndicatorFOMC}_t, \text{TweetPolarity}_{t-1})$

Model 4: $X_t = (\text{IndicatorFOMC}_t, \text{TweetPolarity}_{t-1} \cdot \text{IndicatorFOMC}_t)$

The first model is the most basic, and does not capture the fact that the distribution of market returns can change on FOMC announcement dates. The second model adds an indicator variable to account for this change in the distribution of returns. The third model includes this indicator variable as well as our Twitter polarity measure on date $t-1$. The fourth model differs from the third only in that the Twitter polarity is interacted with the FOMC indicator, so that it is nonzero only on days when FOMC decisions are announced.

Each model will correspond to different estimates for $\hat{\alpha}$, $\hat{\beta}$, and therefore different portfolio weights $\{\tilde{f}_t(\tau)\}$ implemented on each date. The profitability of the models gives us a measure of the information content of tweets. If the third model is the most profitable, then tweets about the Fed always carry information, even if they are made on periods when the Fed does not make decisions. But if the fourth model is the most profitable, then tweets made on days far away from Fed decisions are more likely to be "noise", while tweets made right before the FOMC announcement are likely to be informative.

In a realistic simulation, we do not observe information from date $t+1$ when trading on day t. Thus, we cannot calibrate a model that will be used to predict R_t using tweets or returns from any date beyond t. In order to avoid this look-ahead bias, we initialize our portfolio simulations on January 2010. The models that are used to predict 2010 returns are trained only with data from 2009. The models that are used to predict 2011 returns are trained only with data from 2010 and 2011, and so forth.[6]

Exhibit 7 shows the manager's wealth over the 2010–2014 period for all four strategies using the quarter-Kelly criterion. The first row of the figure shows the results when the models are estimated on the entire 2010–2014 period, and thus are subject to look-ahead bias. The second row of the figure in exhibit 7 shows the out-of-sample results, where the model used to trade in year T is trained only with data from years before T. The columns correspond to different caps on leverage. We can see that, with high leverage, our models that use social media information outperform the models that use only market returns and FOMC indicators as predictors.

Exhibit 8 shows the performance metrics for the different portfolios using the out-of-sample estimation of the regressions and the quarter-Kelly criterion. The first column indicates which model is used to predict performance. The second column gives the bound L on the amount of leverage used, with L set to either 1, 2 or 4. The following columns show the annualized portfolio returns, the maximum drawdown, the annualized sharpe and information ratios, annualized alpha, and beta.[7] As can be seen from the exhibit, our strategies that uses information from tweets outperform when leverage is 2:1 or 4:1 and they have higher returns and higher information ratios than a strategy that just invests in a levered market index or a strategy that only uses the FOMC indicator as a signal. Furthermore, model 4 (where tweets are used only on days when the FOMC meets) outperforms model 3 (where tweets are used to predict returns every day). This suggests that the information value of social media is higher when there is real economic news.

It is important to highlight here that our fourth strategy (which uses tweet information only on days the FOMC is scheduled to meet) follows the CRSP value-weighted index passively every day except for *eight* trading days a year, and yet significantly outperforms the levered market benchmark with high information ratios. Model 2—which does not use

[6] One possible objection with our approach is that we do not have historical data for the number of followers of a Twitter user. Thus, we use the number of followers that the users had in August 2015. If users who are more accurate gained more followers between 2009 and 2014 due to their accuracy, this can bias our results. We leave this question for future work.

[7] The information ratio for strategy $i \in \{2, 3, 4\}$ is computed as $\frac{\mathbb{E}[R_i - R_1]}{\sqrt{\mathrm{Var}[R_i - R_1]}}\sqrt{252}$, where R_i is the return of strategy i and R_1 is the return on the Levered market strategy with leverage cap L.

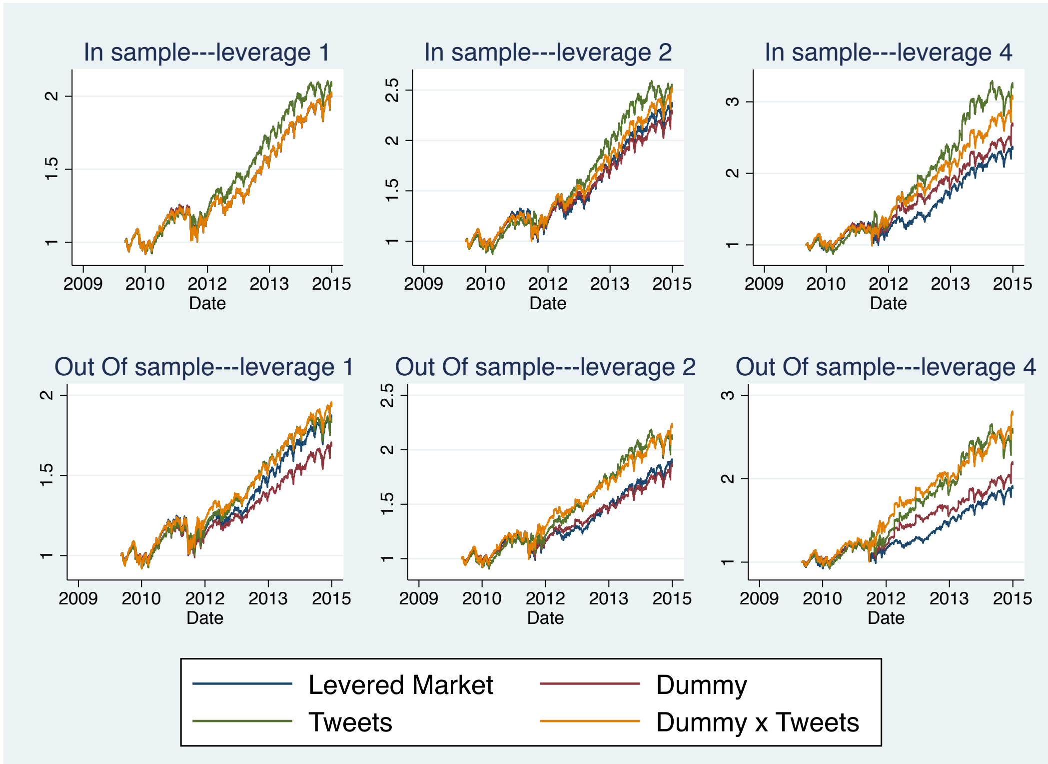

Exhibit 7: Wealth over the 2010–2014 period for four strategies using the quarter-Kelly criterion.

Twitter data—underperforms model 4, showing that information from social media can be very helpful for tactical asset allocation.

Exhibits 9 and 10 shows analogous results using the half-Kelly and full-Kelly criteria. As with the quarter-Kelly criterion, we can see that the portfolio metrics improve when using Twitter information in our prediction models.

Analysis of Portfolio Weights

To develop a better understanding of the practical implications of social-media data, consider the portfolio weights of our hypothetical trading strategies. Exhibit 11 shows the weight that the four strategies put on equities when they use a quarter-Kelly criterion and leverage 4. Note that models 2 and 4 change their position very infrequently (mostly on FOMC meeting days), and they behave very similarly: when the FOMC does not meet, these strategies use none or very little leverage. Model 2 increases its leverage to 4 every single day the FOMC meets. Model 4 is similarly aggressive, except on five days when Twitter sentiment was so

Strategy	L	Return	Drawdown	Sharpe	Info. Ratio	Beta	Alpha
Model 1	1.00	13.08	-7.77	0.88		0.94	-0.83
						(0.00)	(0.00)
Model 2	1.00	10.97	-4.89	0.88	-0.76	0.78	-0.67
						(0.00)	(0.00)
Model 3	1.00	12.87	-7.84	0.92	-0.11	0.85	0.18
						(0.01)	(0.00)
Model 4	1.00	14.06	-8.19	0.93	0.33	0.94	0.17
						(0.01)	(0.00)
Model 1	2.00	13.47	-7.77	0.87		0.99	-1.10
						(0.00)	(0.00)
Model 2	2.00	13.04	-3.72	0.93	-0.15	0.83	0.60
						(0.01)	(0.00)
Model 3	2.00	15.95	-9.26	0.95	0.35	0.97	1.52
						(0.01)	(0.00)
Model 4	2.00	17.11	-7.35	0.98	0.72	1.06	1.41
						(0.01)	(0.00)
Model 1	4.00	13.47	-7.77	0.87		0.99	-1.10
						(0.00)	(0.00)
Model 2	4.00	16.75	-2.71	0.91	0.28	0.93	3.16
						(0.02)	(0.00)
Model 3	4.00	20.54	-7.95	0.97	0.55	1.08	4.66
						(0.02)	(0.00)
Model 4	4.00	22.53	-5.42	1.07	0.82	1.14	5.46
						(0.02)	(0.00)

Exhibit 8: Out-of-sample performance metrics for our four strategies under different levels of leverage caps and the quarter-Kelly criterion. The information ratio is computed using the "levered market" strategy as a benchmark.

Strategy	L	Return	Drawdown	Sharpe	Info. Ratio	Beta	Alpha
Model 1	1.00	14.84	-8.19	0.94		1.00	0.00
						(0.00)	(0.00)
Model 2	1.00	14.84	-8.19	0.94		1.00	0.00
						(0.00)	(0.00)
Model 3	1.00	15.32	-7.27	0.98	0.13	0.97	0.89
						(0.00)	(0.00)
Model 4	1.00	15.91	-8.19	1.00	0.35	0.99	1.14
						(0.00)	(0.00)
Model 1	2.00	24.74	-16.56	0.88		1.89	-1.66
						(0.00)	(0.00)
Model 2	2.00	21.02	-10.42	0.88	-0.76	1.57	-1.33
						(0.01)	(0.00)
Model 3	2.00	24.70	-16.41	0.92	-0.11	1.71	0.37
						(0.01)	(0.00)
Model 4	2.00	26.85	-17.49	0.93	0.33	1.88	0.34
						(0.01)	(0.00)
Model 1	4.00	25.23	-16.56	0.87		1.99	-2.19
						(0.01)	(0.00)
Model 2	4.00	25.10	-7.61	0.93	-0.15	1.67	1.20
						(0.02)	(0.00)
Model 3	4.00	30.46	-19.59	0.95	0.35	1.94	3.06
						(0.02)	(0.00)
Model 4	4.00	32.67	-16.21	0.98	0.72	2.11	2.84
						(0.02)	(0.00)

Exhibit 9: Out-of-sample performance metrics for our four strategies under different levels of leverage caps and the half-Kelly criterion. The information ratio is computed using the "levered market" strategy as a benchmark.

Strategy	L	Return	Drawdown	Sharpe	Info. Ratio	Beta	Alpha
Model 1	1.00	14.84	-8.19	0.94		1.00	0.00
						(0.00)	(0.00)
Model 2	1.00	14.84	-8.19	0.94		1.00	0.00
						(0.00)	(0.00)
Model 3	1.00	14.99	-7.44	0.95	0.04	0.98	0.44
						(0.00)	(0.00)
Model 4	1.00	15.91	-8.19	1.00	0.35	0.99	1.14
						(0.00)	(0.00)
Model 1	2.00	28.29	-17.49	0.94	0	2.00	0.00
						(0.00)	(0.00)
Model 2	2.00	28.29	-17.49	0.94		2.00	0.00
						(0.00)	(0.00)
Model 3	2.00	29.52	-16.29	0.98	0.13	1.93	1.79
						(0.01)	(0.00)
Model 4	2.00	30.70	-17.49	1.00	0.35	1.97	2.30
						(0.01)	(0.00)
Model 1	4.00	40.70	-36.44	0.88		3.78	-3.29
						(0.01)	(0.00)
Model 2	4.00	36.53	-28.43	0.88	-0.76	3.13	-2.63
						(0.02)	(0.00)
Model 3	4.00	42.68	-35.36	0.92	-0.11	3.41	0.73
						(0.03)	(0.00)
Model 4	4.00	45.30	-37.58	0.93	0.33	3.75	0.67
						(0.02)	(0.00)

Exhibit 10: Out-of-sample performance metrics for our four strategies under different levels of leverage caps and the full-Kelly criterion. The information ratio is computed using the "levered market" strategy as a benchmark.

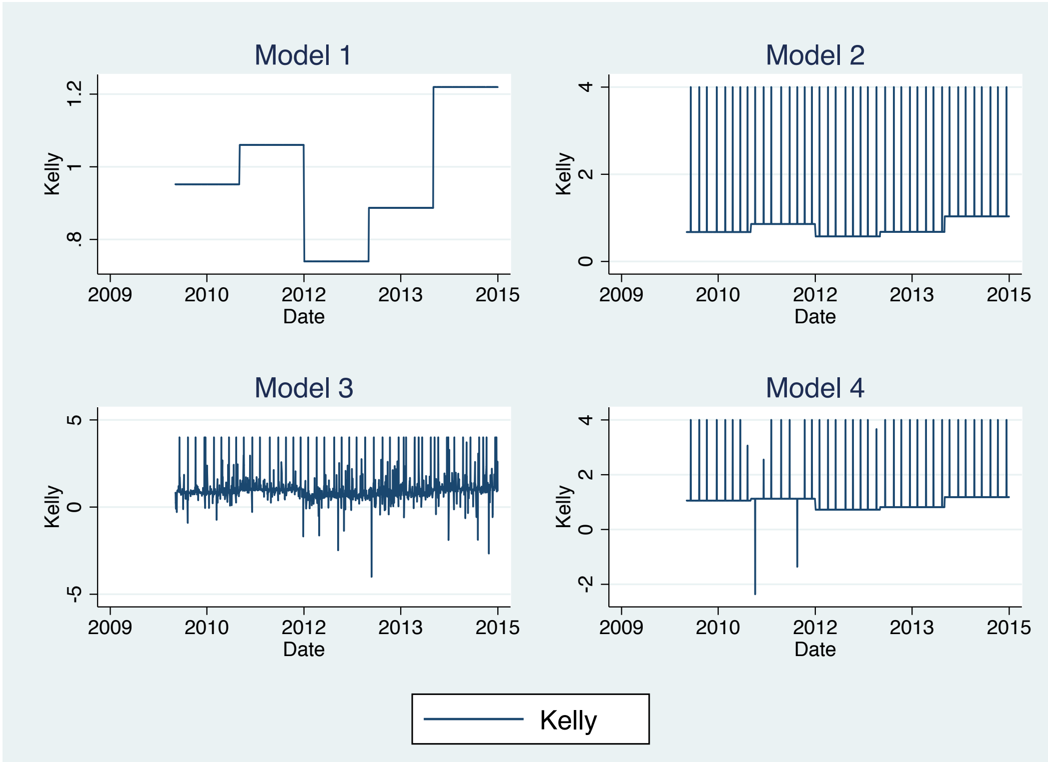

Exhibit 11: Kelly fractions over the 2010–2014 period for four strategies with the quarter-Kelly criterion and a leverage cap of 4.

negative that the strategy decides to reduce its leverage, or even go short. This is only a small number of trading days, but their effect on the portfolio behavior is important. We illustrate this by analyzing the two days where model 4 went short : January 26, 2011 and September 21, 2011. On January 26, the market return was 0.6%, giving model 4 a loss of 1.4% and model 2 a gain of 0.24%. On September 21, 2011, the market had a very negative reaction to the Fed's announcement of Operation Twist. The market return was -2.9%. Model 2 lost 11.68% of their assets on that day, while model 4 used the Twitter negativity before the announcement to take a short position and had a 3.9% gain.

As the models were retrained every year, model 2 learned that very large losses can come on Fed days, while model 4 explained this loss using Twitter sentiment. This allowed model 4 to bet more aggressively during the 2012-2013 periods, which were very good years for equities which showed strong positive market reactions to FOMC decisions.

We can observe this effect in the weights used by the full-Kelly strategies (Exhibit 12). Model 4 bets aggressively on all days, with leverage close to the maximum. It takes very

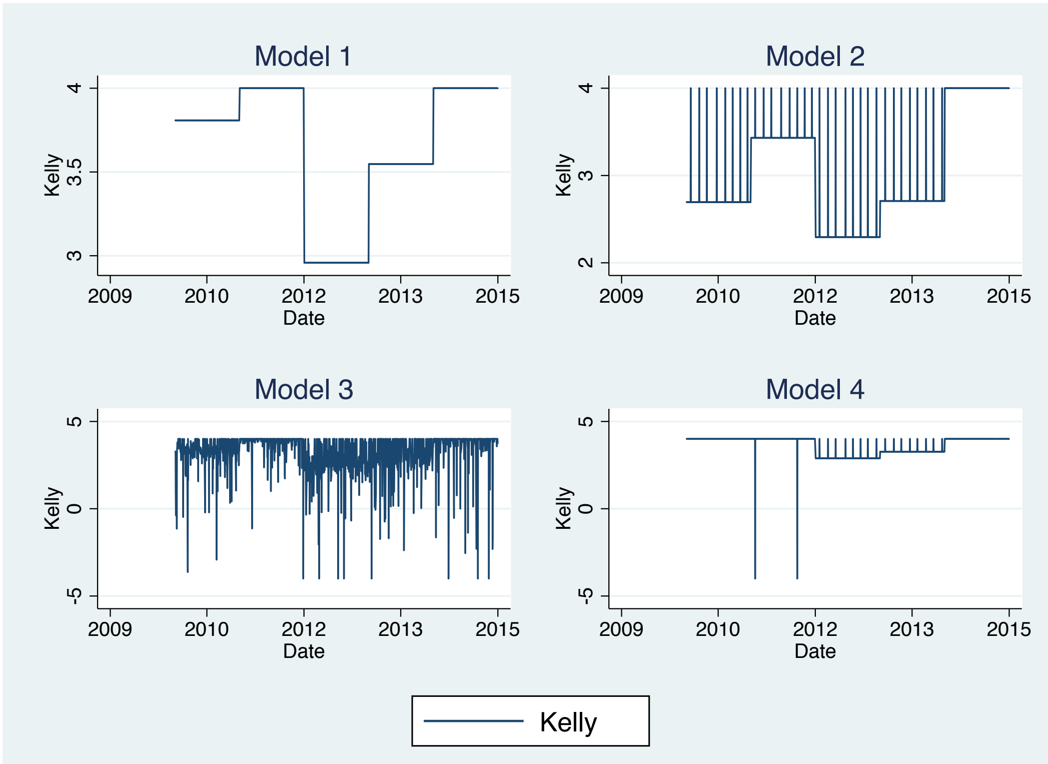

Exhibit 12: Kelly fractions over the 2010–2014 period for four strategies with the full-Kelly criterion and a leverage cap of 4.

strong short bets on January and September 2011, and uses the additional information it has to forecast a higher value (and therefore, use higher leverage) on days on which the FOMC does not meet.

Conclusion

A priori, it is not immediate that tweets contain information about asset prices, especially since individuals who tweet are not necessarily sophisticated or even invested in the market. In this paper, we show that these tweets do contain information, which can be used to predict returns even after controlling for common market factors. Furthermore, the information from tweets can be used to build portfolios that outperform on several dimensions a comparably levered benchmark market portfolio.

A caveat to our results is that the period for which tweets are available is one of rising markets and zero interest rates. It will be interesting to observe if this strategy continues to outperform the market during environments where the market index is receding and interest rates increase.

Acknowledgements. We thank William Li for helpful comments and discussion, Srijith Poduval for research assistance, and Jayna Cummings for editorial assistance. The views and opinions expressed in this article are those of the authors only and do not necessarily represent the views and opinions of any other organizations, any of their affiliates or employees, or any of the individuals acknowledged above. Research support from the MIT Laboratory for Financial Engineering is gratefully acknowledged.

Appendix

The Pattern algorithm for sentiment analysis assigns a polarity score between -1 and $+1$ to a given input text. This algorithm relies on the SentiWordnet database to assign scores. SentiWordnet is an annotated dictionary where each word is associated with multiple meanings, and each meaning is mapped to a triplet of numbers (p, n, o) measuring the positivity, negativity and objectivity of the word. The sum of the three numbers always adds up to one. For example, the word "good" may have one of the following meanings

- "Having desirable or positive qualities especially those suitable for a thing specified." The triplet for this meaning is $(p, n, o) = (0.75, 0, 0.25)$.
- "Commodity. Article of commerce". The triplet for this meaning is $(p, n, o) = (0, 0, 1)$.

The Pattern algorithm takes a tweet with n words and parses it into an array $(m_1, m_2, \ldots, m_n)$ of groups of words. If a word is an adjective or an adverb, it can amplify or decrease the polarity of a noun that they are modifying. A typical example is "not good", which has a negative polarity, instead of a positive polarity. The algorithm finishes by averaging the polarities of all groups of words, and assigning this score to the tweet.

References

Azar, P. D., W. Li, and A. W. Lo, 2015, "Decoding Fedspeak: A VAR Analysis of FOMC transcripts," .

Bar-Haim, R., E. Dinur, R. Feldman, M. Fresko, and G. Goldstein, 2011, "Identifying and following expert investors in stock microblogs," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1310–1319. Association for Computational Linguistics.

Bernanke, B. S., and K. N. Kuttner, 2005, "What explains the stock market's reaction to Federal Reserve policy?," *Journal of Finance*, 60, 1221?1257.

Cieslak, A., A. Morse, and A. Vissing-Jorgensen, 2014, "Stock returns over the FOMC cycle," *NBER Working Paper*.

De Smedt, T., and W. Daelemans, 2012, "Pattern for python," *The Journal of Machine Learning Research*, 13(1), 2063–2067.

Giannini, R. C., P. J. Irvine, and T. Shu, 2014, "Do local investors know more? A direct examination of individual investors? information set," in *A Direct Examination of Individual Investors? Information Set (August 8, 2014). Asian Finance Association (AsFA) 2013 Conference*.

Gurkaynak, R. S., B. Sack, and E. Swanson, 2005, "Do actions speak louder than words? The response of asset prices to monetary policy actions and statements," pp. 55–93.

Jegadeesh, N., and D. Wu, 2015, "Deciphering Fedspeak: The Information Content of FOMC Meetings," working paper, Emory University.

Lucca, D. O., and E. Moench, 2015, "The Pre-FOMC Announcement Drift," *The Journal of Finance*, 70(1), 329–371.

Lucca, D. O., and F. Trebbi, 2009, "Measuring central bank communication: An automated approach with application to FOMC statements," *NBER Working Paper 15367, National Bureau of Economic Research.*

Plakandaras, V., T. Papadimitriou, P. Gogas, and K. Diamantaras, 2014, "Market sentiment and exchange rate directional forecasting," working paper.