Appendix 1: Sample and variable definitions

Our sample consists of 3230 earnings announcements from S&P 500 firms and includes firm-quarters ending between March 31, 2012, and March 31, 2014. We collect accounting data from Compustat, analyst forecasts from I/B/E/S, return data from CRSP, and Tweets from the Twitter API (sentiment is calculated using the vaderSentiment package in Python). In the tables in Appendix 3, we only include observations with non-missing values for the variables required for each particular analysis.

Variable	Definition
#Analyst	The number of analysts following the firm taken from I/B/E/S.
analysts	The natural logarithm of #Analyst.
car	Cumulative abnormal returns (CAR), measured pre [-3,0], post [0,+3], or around [-3,+3] each earnings announcement. (Denoted with the suffixes _pre, _post, and _full, respectively.) Daily normal returns are defined as the difference between the normal return and the actual return. Normal returns are predicted using the following models: Model 0: $R = R_m$ Model 1: $R = \alpha + \beta(R_m - R_f)$ Model 2: $R = \alpha + \beta_1(R_m - R_f) + \beta_2 SMB + \beta_3 HML$ Model 3: $R = \alpha + \beta_1(R_m - R_f) + \beta_2 SMB + \beta_3 HML + \beta_4 UMD$ The coefficients are estimated using a 150 day estimation window [-165,-16], requiring a minimum of 120 non-missing return observations. Only Model 1 is used in the regression analysis.
eps | Earnings-per-share	Dividing the net income (niq) by shares outstanding (cshoq).
leverage | Leverage	Long-term debt plus debt in current liabilities divided by common equity ((dlttq + dlcq) / seqq).
lntweetsnum	The natural logarithm of the number of twitter post mentioning the company Ticker, measured pre [-3,0], post [0,+3], or around [-3,+3] each earnings announcement. (Denoted with the suffixes _pre, _post, and _full, respectively.) Firm-related tweets are defined as tweets containing cashtag plus the firm's ticker symbol (e.g., $AAPL for Apple Inc.; $PEP for PepsiCo Inc.).
mb | Market-to-book	Market value divided by book value of equity (mkvaltq / bkvlps).
Mean forecast	Consensus (mean) earnings-per-share forecast taken from I/B/E/S.
neg	The probability of a tweet having a negative sentiment, calculated using the python package vaderSentiment.
pos	The probability of a tweet having a positive sentiment, calculated using the python package vaderSentiment.
post	Indicator variable for quarters after SEC's 2013 guidance on social media disclosure.
Return on assets	Net income divided by total assets (niq / atq).
Return on equity	Net income divided by common book value equity (niq / bkvlps).
size | Total assets	The natural logarithm of total assets (atq) taken from Compustat.
ux | Unexpected earnings	The difference between the actual reported earnings-per-share and the mean analysts' earnings-per-share forecast taken from I/B/E/S.

Appendix 2: Figures

Figure 1: Calculating CAR

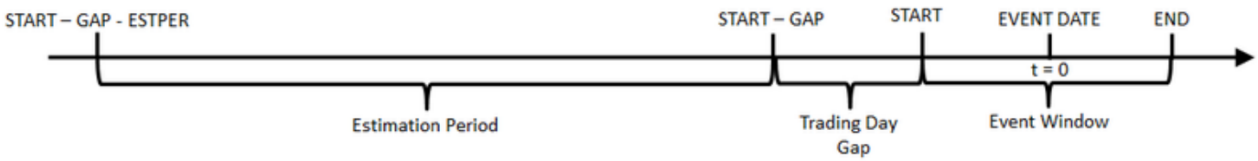
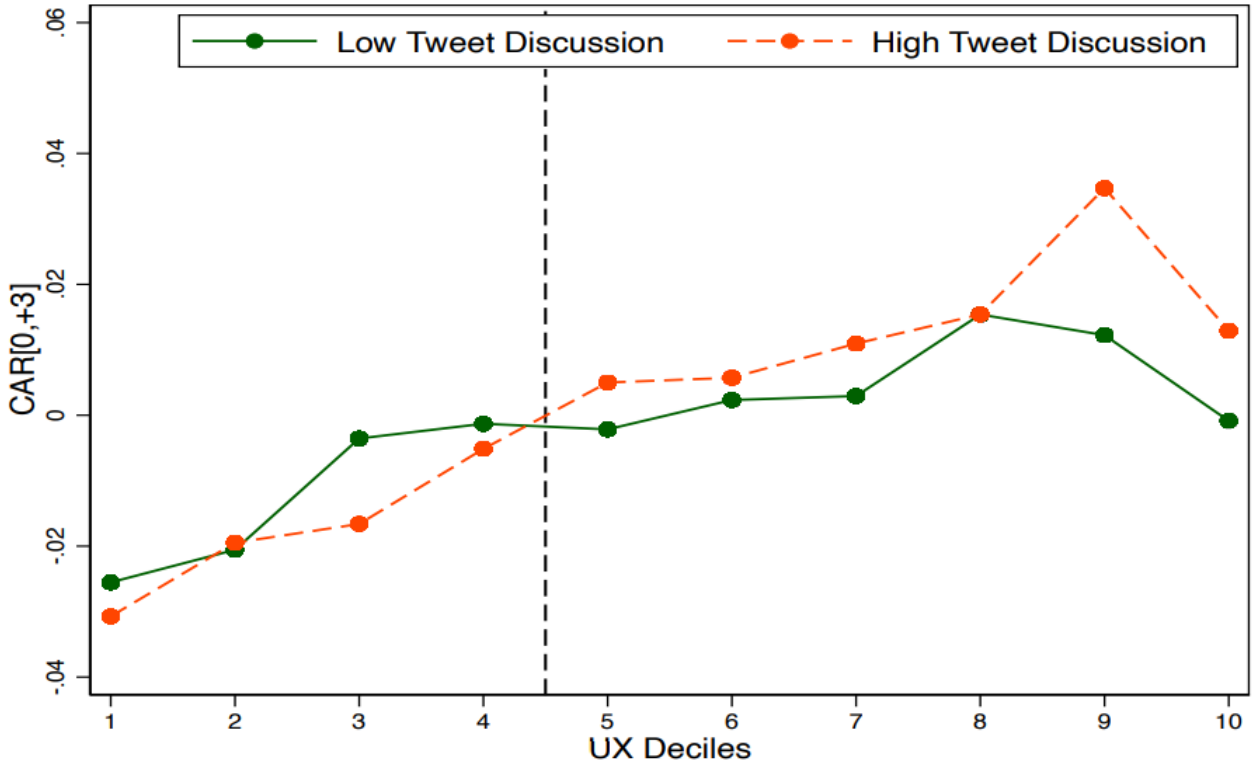


Figure 2: Twitter discussion and the earnings response coefficient (ERC)



Appendix 3: Tables

Table 1: Industry means

Industry	TweetsNum	TweetsTone	CAR[-3,+3]	UX
Utilities	85.1148	0.1051528	-0.0024514	0.0046209
Oil and Gas	152.832	0.1220753	0.0089792	0.0402538
Manufacturing	164.616	0.0907888	0.0038828	0.0346664
Consumer Durables	173.029	0.091398	0.0131532	0.0300735
Consumer NonDurables	241.442	0.0823722	0.0038728	0.0297189
Telephone and Television	278.963	0.076318	0.0072714	0.0203704
Other - Trans, Hotels, Finance	310.282	0.0773433	-0.0028433	0.0537081
Healthcare/Medical Equipment	346.855	0.0858497	-0.0073183	0.0273979
Business Equipment	578.728	0.0747156	0.0037538	0.0391979
Wholesale and Retail	599.604	0.0648143	-0.0025295	0.0113325

Table 2: Determinants of Twitter discussion

	(1) lntweetsnum_full
size	0.348^{***}
	(23.710)
eps	0.053^{***}
	(3.642)
leverage	0.000
	(1.195)
mb	6.146^{***}
	(2.604)
analysts	0.662^{***}
	(14.911)
_cons	-6.734^{***}
	(-2.857)
Quarter FE	*Yes*
Industry FE	*Yes*
R^2	0.345
N	3,060

Table 3: Alpha in pre-announcement tweets

	(1) ux	(2) car1_post
neg_pre	−0.052	0.053
	(−0.400)	(1.178)
pos_pre	0.009	−0.037
	(0.121)	(−1.465)
lntweetsnum_pre	0.009***	−0.000
	(3.890)	(−0.156)
_cons	0.004	0.002
	(0.470)	(0.718)
Quarter FE	*Yes*	*Yes*
Industry FE	*Yes*	*Yes*
R^2	0.016	0.010
N	3,229	3,219

Table 4: Tweet-market consistency

	(1) car1_pre	(2) car1_post	(3) car1_full
neg_pre	−0.068**		
	(−2.127)		
pos_pre	−0.016		
	(−0.929)		
lntweetsnum_pre	0.001		
	(0.956)		
neg_post		−0.742***	
		(−15.198)	
pos_post		0.181***	
		(5.979)	
lntweetsnum_post		0.004***	
		(4.232)	
neg_full			−0.470***
			(−7.988)
pos_full			0.088***
			(2.661)
lntweetsnum_full			0.003***
			(2.865)
_cons	0.001	−0.005	−0.006
	(0.702)	(−1.351)	(−1.288)
Quarter FE	*Yes*	*Yes*	*Yes*
Industry FE	*Yes*	*Yes*	*Yes*
R^2	0.013	0.087	0.033
N	3,219	3,212	3,219

Table 5: ERC pre/post SEC guidance on social media (2013)

	(1) Full Sample car1_post	(2) Active Firms car1_post	(3) Non-Active Firms car1_post
ux	0.046^{***}	0.053^{***}	0.040^{***}
	(6.289)	(4.628)	(4.329)
post × ux	0.039^{***}	0.047^{**}	0.033^{*}
	(2.915)	(2.247)	(1.961)
size	-0.001	-0.002^{*}	0.000
	(-1.357)	(-1.902)	(0.330)
eps	-0.000	-0.001	0.001
	(-0.117)	(-0.750)	(0.610)
mb	0.105	0.070	0.349
	(0.807)	(0.468)	(1.169)
analysts	0.001	-0.001	0.002
	(0.517)	(-0.146)	(0.681)
_cons	-0.099	-0.049	-0.361
	(-0.761)	(-0.327)	(-1.212)
Quarter FE	*Yes*	*Yes*	*Yes*
Industry FE	*Yes*	*Yes*	*Yes*
R^2	0.038	0.040	0.056
N	3,217	1,734	1,483

Table 6: Firm characteristics

Variable	Obs	Mean	Std. Dev.	Min	Max
Unexpected earnings	2948	0.035	0.153	-3.560	2.370
Total assets	2948	9.853	1.374	7.138	14.723
Market capitalization	2948	9.758	1.034	7.178	13.237
Earnings-per-share	2948	0.874	1.178	-12.381	16.895
Leverage	2948	1.141	3.154	0.000	91.455
Market-to-book	2948	1.000	0.008	0.915	1.212
Return on assets	2948	0.017	0.020	-0.201	0.262
Return on equity	2948	0.014	0.027	-0.601	0.438
#Analyst	2948	2.835	0.473	0.000	3.892
Mean forecast	2948	0.808	0.589	-0.770	5.990

This table shows summary statistics for quarterly firm observations measured for the quarter of the earnings announcement. For variable definitions, please see Appendix 1.

Table 7: Cumulative abnormal returns

Variable	Obs	Mean	Std. Dev.	Min	Max
pre-event (3:0)					
Car (Model 0)	2948	-0.001	5.354	-26.900	38.200
Car (Model 1)	2948	3.871	3.723	0.002	38.600
Car (Model 2)	2948	3.861	3.706	0.002	38.700
Car (Model 3)	2948	3.853	3.702	0.004	38.500
post-event (0:3)					
Car (Model 0)	2948	0.041	3.935	-24.100	35.500
Car (Model 1)	2948	2.699	2.897	0.000	37.900
Car (Model 2)	2948	2.701	2.902	0.002	38.900
Car (Model 3)	2948	2.688	2.881	0.001	40.000
combined window (3:3)					
Car (Model 0)	2948	-0.030	5.650	-28.200	37.400
Car (Model 1)	2948	4.105	3.943	0.004	39.200
Car (Model 2)	2948	4.111	3.935	0.003	40.000
Car (Model 3)	2948	4.089	3.928	0.003	40.600

This table shows cumulative abnormal returns for the pre-event window (three days prior to the earnings announcement) and the post event window (three days after the earnings announcement). Cumulative abnormal returns from three days prior to three days after the earnings announcement are shown in the combined measure. Model 0: $R = R_m$; Model 1: $R = \alpha + \beta(R_m - R_f)$; Model 2: $R = \alpha + \beta_1(R_m - R_f) + \beta_2 SMB + \beta_3 HML$; Model 3: $R = \alpha + \beta_1(R_m - R_f) + \beta_2 SMB + \beta_3 HML + \beta_4 UMD$. For variable definitions, please see Appendix 1.

Table 8: Twitter data

Variable	Obs	Mean	Std. Dev.	Min	Max
pre-event (3:0)					
# Tweets	2948	104.987	560.821	1.000	21646.000
Probability of negative	2948	2.276	2.371	0.000	24.042
Probability of neutral	2948	91.641	4.924	59.100	100.000
Probability of positive	2948	6.083	4.086	0.000	36.873
post-event (0:3)					
# Tweets	2948	207.737	639.382	1.000	14093.000
Probability of negative	2948	3.135	2.119	0.000	25.441
Probability of neutral	2948	90.317	4.119	53.100	100.000
Probability of positive	2948	6.548	3.362	0.000	37.439
combined window (3:3)					
# Tweets	2948	312.724	1150.594	4.000	35739.000
Probability of negative	2948	2.705	1.950	0.000	24.742
Probability of neutral	2948	90.979	4.001	62.067	100.000
Probability of positive	2948	6.316	3.261	0.000	35.310

This table shows summary statistics for data scrapped from twitter for the period Q1 2012 until Q1 2014 divided into pre-event window prior to the earnings announcement; post-event window after the earnings announcement and the combined window from three days prior to three days after the earnings announcement. For variable definitions, please see Appendix 1.

F&F Industry Classification	N	#Tweets	Prob. Positive	Prob. Neutral	Prob. Negative
Consumer NonDurables	202	110.86	5.68	91.95	2.38
Consumer Durables	61	54.74	5.85	92.40	1.74
Manufacturing	407	516.12	6.18	91.87	1.96
Oil, Gas, and Coal Extraction and Products	187	475.19	8.18	89.56	2.26
Business Equipment	438	179.04	5.58	92.26	2.16
Telephone and Television Transmission	74	893.78	5.61	92.03	2.36
Wholesale, Retail, and Some Services	332	206.63	5.75	90.84	3.41
Healthcare, Medical Equipment, and Drugs	231	127.70	6.50	90.95	2.55
Utilities	244	333.81	7.06	91.31	1.62
Other	772	811.79	5.68	92.16	2.16
Total	2948	105.02	6.08	91.64	2.27

This table reports means of twitter data for the pre-event window across Fama-French Industry Classification.

F&F Industry Classification	N	#Tweets	Prob. Positive	Prob. Neutral	Prob. Negative
Consumer NonDurables	202	138.36	6.74	90.06	3.20
Consumer Durables	61	131.59	6.68	90.67	2.64
Manufacturing	407	116.92	6.63	90.36	3.01
Oil, Gas, and Coal Extraction and Products	187	107.24	8.02	89.07	2.91
Business Equipment	438	418.37	6.02	90.88	3.10
Telephone and Television Transmission	74	212.85	6.40	90.50	3.10
Wholesale, Retail, and Some Services	332	364.05	6.84	88.69	4.47
Healthcare, Medical Equipment, and Drugs	231	180.77	6.44	90.44	3.12
Utilities	244	51.86	7.20	90.28	2.51
Other	772	174.45	6.11	90.96	2.93
Total	2948	207.8	6.55	90.31	3.14

This table reports means of twitter data for the post-event window across Fama-French Industry Classification.

Market Capitalization	N	#Tweets	Prob. Positive	Prob. Neutral	Prob. Negative
First quartile	737	485.98	5.48	92.57	1.95
Second quartile	737	116.28	5.87	92.04	2.09
Third quartile	737	540.12	6.06	91.74	2.20
Forth quartile	737	201.06	6.93	90.21	2.86
Total	2948	104.99	6.08	91.64	2.27

This table reports means of twitter data for the pre-event window for quartile split by market capitalization.

Market Capitalization	N	#Tweets	Prob. Positive	Prob. Neutral	Prob. Negative
First quartile	737	108.49	5.77	91.34	2.89
Second quartile	737	173.85	6.12	90.91	2.97
Third quartile	737	117.28	6.58	90.44	2.98
Forth quartile	737	431.33	7.72	88.58	3.70
Total	2948	207.74	6.54	90.32	3.14

This table reports means of twitter data for the post-event window for quartiles split by market capitalization.

Variable	Obs	Mean	Std. Dev.	Min	Max
pre-event (3:0)					
Unexpected earnings positive					
# Tweets	2169	102.761	432.110	1.000	11071.000
Probability of negative	2169	2.276	2.267	0.000	24.042
Probability of neutral	2169	91.665	4.718	63.962	100.000
Probability of positive	2169	6.059	3.888	0.000	36.038
Unexpected earnings negative					
# Tweets	779	111.186	819.149	1.000	21646.000
Probability of negative	779	2.273	2.641	0.000	22.917
Probability of neutral	779	91.576	5.458	59.100	100.000
Probability of positive	779	6.150	4.595	0.000	36.873
post-event (0:3)					
Unexpected earnings positive					
# Tweets	2169	210.965	627.731	1.000	10177.000
Probability of negative	2169	3.072	2.021	0.000	25.441
Probability of neutral	2169	90.418	3.941	53.100	100.000
Probability of positive	2169	6.510	3.191	0.000	37.439
Unexpected earnings negative					
# Tweets	779	198.747	671.091	1.000	14093.000
Probability of negative	779	3.311	2.365	0.000	22.305
Probability of neutral	779	90.035	4.570	60.951	100.000
Probability of positive	779	6.655	3.800	0.000	36.125

This table reports summary statistics of twitter data for the sample split by negative and positive unexpected earnings for the post event window period.