Can Twitter Help Predict Firm-Level Earnings and Stock Returns?

Eli Bartov, Lucile Faurel, Partha S. Mohanram

The Accounting Review, 2018.

孔彤阳 2021/01/16

Outline

- Introduction
- Research design
- Empirical analysis
- Supplementary test
- Conclusion

Background & Motivation

- Twitter is an emerging new source of information to the capital market, but it is unclear whether information from Twitter will be useful to investors.
- The academic literature has begun studying the role Twitter plays in the capital market only recently and whether firm specific information from Twitter is useful in predicting a firm’s earnings and stock returns has not been addressed.

Research Question

1. Does the aggregate opinion in individual tweets regarding a stock's prospects, OPI, predict its quarterly earnings? √
2. Does OPI predict the stock price reaction to the firm's earnings realizations? √
3. Does the information environment quality of a company explain the cross-sectional variation in the predictive ability of OPI (if it exists)? √

Contribution

1. While stock investing may be viewed as a zero-sum game, our results demonstrate that individuals use Twitter as an important channel to share information regarding stocks for their mutual benefit.
2. Our results show that Twitter can play a role in making the market more efficient and regulatory intervention does not seem warranted.

Literatures

- How companies exploit Twitter to communicate with investors. (Blankespoor et al., 2014; Jung et al., 2016; Lee et al., 2015)
- Whether information from Twitter predicts the overall stock market. (Bollen et al., 2011; Mao et al., 2012)
- How Twitter activity influences investor response to earnings. Curtis et al.(2016) find that high levels of activity are associated with greater sensitivity of earnings announcement returns to earnings surprises.

Literatures

- A broad stream of research has examined investors' use of Internet search engines, financial websites, forums, and other social media platforms and provided mixed evidence on whether this information helps predict future earnings and stock returns. (Da et al., 2011; Drake et al., 2012; Hirschey et al., 2000; Tumarkin, Whitelaw, 2001; Antweiler, Frank, 2004; Das, Chen, 2007; Chen et al., 2014)

Wisdom of Crowds

- The aggregation of information provided by many individuals often results in predictions that are better than those made by any single member of the group, or even experts.
- Hong and Page (2004) show that a diverse group of intelligent decision makers reaches reliably better decisions than a less diverse group of individuals with superior skills.
- Twitter has the most diverse set of users among social media platforms.

Sample selection & Data

- Data: Historical Twitter data
- Resources: GNIP
- Sample time interval: 2009Q1~2012Q4
- Sample selection:

1. Limit the sample to tweets with cashtags (e.g., $AAPL) .
2. Drop tweets containing multiple stock symbols.
3. Require that the firms are on Compustat.
4. Focus only on tweets written in the event window [-10;-2].

Final sample consists of 869,733 tweets, covering 33,186 firm-quarters from 3,604 distinct Russell 3000 firms.

Research design

- Can aggregate opinion predict earnings surprises?

$$\begin{aligned} ESURP = \ & \alpha + \beta_1 * OPI_{[-10;-2]} + \beta_2 * PRIOR\_ESURP \\ & +\beta_3 * EXRET_{[-10;-2]} + \beta_4 * RPOPI \\ & +\beta_5 * SIZE + \beta_6 * MB + \beta_7 * ANL \\ & +\beta_8 * INST + \beta_9 * Q4 + \beta_{10} * LOSS + \varepsilon \end{aligned} \tag{1}$$

- ***ESURP*** is the earnings surprise, measured using either standardized unexpected earnings (*SUE*) or analyst earnings forecast error (*FE*).
- *OPI* is the aggregate opinion about a firm extracted from individual tweets.

Research design

- *SUE:* measured using quarterly diluted earnings per share, excluding extraordinary items, and applying a seasonal random walk with drift model.
- *FE:* measured as I/B/E/S reported quarterly earnings per share less the latest I/B/E/S consensus analyst forecast, scaled by stock price as of the forecast date, multiplied by 100.
- *OPI:* we measure *OPI* using two textual-analysis methodologies: 1) *OPI_BAYES:* consider each tweet as a whole and classifies it as negative, positive, or neutral; 2) *OPI_VOCAB :* focus on the words included in each tweet and detects specific negative words in the tweet.

Research design

- *PRIOR_ESURP:* the lagged earnings surprise from the previous quarter, included to control for the well-documented positive autocorrelation in earnings surprises.
- *EXRET:* Carhart's (1997) four factor buy-and-hold abnormal stock returns for the firm over the window [-10;-2], multiplied by 100.
- *RPOPI:* a measure of the aggregate opinion in traditional news over the period -10 to -2, developed from the RavenPack database to control for information and opinion from traditional news media.

Research design

- *ANL:* number of analysts in the consensus I/B/E/S/ quarterly earnings forecast.
- *INST:* institutional investor holding.
- *LOSS:* an indicator variable for past quarterly loss.
- *MB*: market-to-book ratio.
- *Q4:* indicator variable for the fourth fiscal quarter.
- *SIZE:* firm size.

OPI

- *OPI_BAYES:* total number of tweets classified as positive less total number of tweets classified as negative during the nine-trading-day window [-10;-2], using an enhanced naïve Bayes classifier. Each positive or negative tweet is first weighted by the corresponding probability and by the number of followers of the user {1 + [Log(1 + Number of Followers)]}. The measure is scaled by one plus the sum of the probability levels.
- *OPI_VOCAB:* single factor from a factor analysis using three vocabulary-based measures. Number of words classified as negative in each tweet is first weighted by the number of followers of the user {1 + [Log(1 + Number of Followers)]}. Then, each measure is defined as minus one multiplied by the sum of the weighted number of negative words during the nine-trading-day window [-10;-2], scaled by one plus the number of words classified as either positive or negative.

OPI

- We distinguish between tweets that contain original information ($OPIo$) and those that relay or disseminate existing information ($OPId$). This classification allows us to compare and contrast the dual role of Twitter as a source of new information and a means of disseminating existing information.
- We distinguish between tweets that explicitly convey information about a firm's earnings, fundamentals, and/or stock price ($OPIf$) and tweets that contain other information($OPInf$). This refinement allows us to cast light on the nature of information that tweets convey.

OPI

#	Tweet	Followers	o / d	f / nf
1	$VCI getting demolished now, wait, you think the nightly news can't move markets, think again, fear mongering idiots	4,540	o	f
2	Simple Moving Average Crossover: stock hitting new low -$VCI - http://t.co/nylq1EXs	59	d	f
3	$VCI Valassis Announces Its Third Quarter 2011 Earnings Conference Call http://t.co/gR48OaG1	59	d	f
4	Valassis Communications, Earnings Estimate $VCI http://t.co/7u1oxQNC	16	d	f
5	Valassis Communications, Earnings Estimate $VCI http://t.co/UJOdmN1I	17	d	f

OPI

- *OPI_BAYES*

#	Result	Probability (%)
1	Negative	86.1
2	Negative	79.0
3	Neutral	51.4
4	Neutral	50.0
5	Neutral	50.0

$$\begin{aligned} OPI\_BAYES &= (-1 * 0.861 * \{1 + [Log(1 + 4{,}540)]\} + \\ &\quad -1 * 0.79 * \{1 + [Log(1 + 59)]\})/(1 + 0.861 + 0.79) \\ &= -4.578 \end{aligned}$$

- *OPI_VOCAB*

#	Number of Negative / Positive Words					
	Loughran and McDonald		Harvard IV-4		Hu and Liu	
	OPI_A		OPI_B		OPI_C	
	Neg .	Pos .	Neg .	Pos .	Neg .	Pos.
1	2	0	2	0	2	0
2	0	0	1	0	0	0
3	0	0	0	1	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0

$$OPI\_A \ = \ -1 \ * \ (2 * \{1 + [Log(1 + 4{,}540)]\}) \ / \ (1 + 2 + 0) \ = \ -6.281$$

$$OPI\_B = \ -1 \ * \ (2 * \{1 + [Log(1 + 4{,}540)]\} + \ 1 * \{1 + [Log(1 + 59)]\}) \ / \ (1 + 3 + 1) \ = \ -4.787$$

$$OPI\_C \ = \ -1 \ * \ (2 * \{1 + [Log(1 + 4{,}540)]\}) \ / \ (1 + 2 + 0) \ = \ -6.281$$

$$OPI\_VOCAB \ = \ single\ factor\ using\ OPI\_A, OPI\_B, OPI\_C \ = \ -4.082$$

Research design

- Can aggregate opinion predict announcement retuens?

$$EXRET_{[-1;+1]} = \alpha + \beta_1 * OPI_{[-10;-2]} + \beta_2 * EXRET_{[-10;-2]}$$
$$+\beta_3 * RPOPI + \beta_4 * ANL + \beta_5 * INST$$
$$+\beta_6 * Q4 + \beta_7 * LOSS + \varepsilon \quad (2)$$

- ***EXRET(%):*** Buy-and-hold abnormal returns measured using Carhart's (1997) four-factor model for the window specified, multiplied by 100. We measure the buy-and-hold abnormal returns, for firm i over three trading days, as follows:

$$EXRET_{[-1;+1]} = \prod_{t=1;3}(1 + R_{it}) - \prod_{t=1;3}(1 + ER_{it})$$

EXRET

- Where, R_{it} is the daily return for firm i on day t, and ER_{it} is the expected return on day t. We compute the daily abnormal returns using Carhart's (1997) four-factor model by first estimating the following model using a 40-trading-day holdout period:

$$R_{it} - RF_t = a_i + b_i(RMRF_t) + s_i(SMB_t) + h_i(HML_t)$$
$$+p_i(UMD_t) + e_{it}$$

- We then use the estimated slope coefficients from above equation, b_i, s_i, h_i, and p_i, to compute the expected return for firm i on day t as follows:

$$ER_{it} = RF_t + b_i(RMRF_t) + s_i(SMB_t) + h_i(HML_t) + p_i(UMD_t)$$

Research design

- Role of the information environment.

$$EXRET_{[-1;+1]} = \alpha_1 + \alpha_2 * POORINFO + \beta_1 * OPI_{[-10;-2]}$$
$$+\beta_2 * OPI_{[-10;-2]} \times POORINFO + \beta_3 * EXRET_{[-10;-2]}$$
$$+\beta_4 * RPOPI + \beta_5 * RPOPI \times POORINFO$$
$$+\beta_6 * ANL + \beta_7 * INST + \beta_8 * Q4 + \beta_9 * LOSS + \varepsilon \quad (3)$$

- We define an indicator variable labeled $POORINFO$, which equals 1 if analyst following, institutional investment, and traditional media coverage are all below sample medians in the same calendar quarter, and 0 otherwise.

Empirical analysis

Aggregate opinion in individual tweets and earnings surprises:

- The results provide consistent support that the aggregate opinion from individual tweets predicts earnings surprises.
- There is no statistical difference in the predictive ability between tweets that convey original information and tweets that disseminate existing information.
- Tweets specifically related to earnings, fundamentals, and stock trading are more important for predicting earnings than other tweets.

- Earnings Surprises and Twitter Opinion

Variable	Expected Sign	Coefficient (t-statistic)			
		ESURP = SUE		ESURP = FE	
		OPI_BAYES	OPI_VOCAB	OPI_BAYES	OPI_VOCAB
		Model I	Model II	Model III	Model IV
Intercept		-1.0572***	-1.3467***	-0.3084***	-0.3256***
		(-11.03)	(-13.23)	(-5.57)	(-5.66)
$OPI_{[-10;-2]}$	+	**0.0275*** **	**0.3138*** **	**0.0075** **	**0.0193** **
		(2.84)	**(13.98)**	**(2.53)**	**(2.40)**
PRIOR_ESURP		0.2831***	0.2801***	0.1871***	0.1871***
		(40.47)	(42.91)	(10.98)	(11.14)
$EXRET_{[-10;-2]}$		0.0219***	0.0213***	0.0084***	0.0085***
		(6.40)	(6.66)	(4.60)	(4.42)
RPOPI		0.3545***	0.3465***	0.0582***	0.0585***
		(5.27)	(5.27)	(2.68)	(2.72)
SIZE		0.1114***	0.1448***	0.0242***	0.0260***
		(7.59)	(9.39)	(3.17)	(3.32)
MB		-0.0112**	-0.0045	0.0018	0.0023
		(-2.10)	(-0.87)	(0.85)	(1.04)
ANL		-0.1111***	-0.0709**	0.0099	0.0123
		(-3.69)	(-2.41)	(0.53)	(0.65)
INST		0.1928**	0.0992	0.1975***	0.1918***
		(2.51)	(1.31)	(4.79)	(4.73)
Q4		0.2968***	0.3437***	-0.0924***	-0.0892***
		(6.23)	(6.99)	(-5.02)	(-4.85)
LOSS		0.8547***	0.8801***	-0.0577**	-0.0570**
		(18.21)	(18.96)	(-2.27)	(-2.18)
N		30,815	30,815	28,784	28,784
Adj. R^2 (%)		9.02	9.60	5.18	5.18

- Original and Dissemination Tweets

Variable	Expected Sign	Coefficient (*t-statistic*)			
		ESURP = SUE		*ESURP = FE*	
		OPI_BAYES	*OPI_VOCAB*	*OPI_BAYES*	*OPI_VOCAB*
		Model I	Model II	Model III	Model IV
Intercept		-1.0587***	-1.3913***	-0.3069***	-0.3321***
		(-11.09)	*(-14.64)*	*(-5.59)*	*(-5.58)*
OPIo$_{[-10;-2]}$	+	**0.0001**	**0.1814*****	**0.0050**	**0.0175***
		(0.01)	***(6.91)***	***(1.07)***	***(1.82)***
OPId$_{[-10;-2]}$	+	**0.0272*****	**0.1930*****	**0.0056****	**0.0100**
		(3.03)	***(7.45)***	***(1.98)***	***(1.18)***
PRIOR_ESURP		0.2831***	0.2805***	0.1871***	0.1872***
		(42.40)	*(42.82)*	*(10.84)*	*(10.76)*
EXRET$_{[-10;-2]}$		0.0220***	0.0218***	0.0084***	0.0085***
		(6.72)	*(6.75)*	*(4.51)*	*(4.77)*
RPOPI		0.3548***	0.3511***	0.0582***	0.0586***
		(5.18)	*(5.24)*	*(2.77)*	*(2.73)*
SIZE		0.1116***	0.1501***	0.0240***	0.0266***
		(7.71)	*(10.44)*	*(3.22)*	*(3.24)*
MB		-0.0112**	-0.0027	0.0018	0.0025
		(-2.03)	*(-0.50)*	*(0.86)*	*(1.21)*
ANL		-0.1117***	-0.0602**	0.0099	0.0141
		(-3.87)	*(-2.10)*	*(0.55)*	*(0.73)*
INST		0.1952**	0.0616	0.1972***	0.1877***
		(2.53)	*(0.80)*	*(4.85)*	*(4.77)*
Q4		0.2963***	0.3431***	-0.0921***	-0.0888***
		(6.21)	*(7.24)*	*(-4.69)*	*(-4.70)*
LOSS		0.8535***	0.8927***	-0.0580**	-0.0553**
		(17.81)	*(18.56)*	*(-2.18)*	*(-2.09)*
N		30,815	30,815	28,784	28,784
Adj. R^2 (%)		9.02	9.61	5.17	5.19
p-value of F-test $\beta_1 = \beta_2$		*0.11*	*0.79*	*0.92*	*0.61*

• Fundamental and Non-Fundamental Tweets

Variable	Expected Sign	Coefficient (*t-statistic*)			
		ESURP = SUE		*ESURP = FE*	
		OPI_BAYES	*OPI_VOCAB*	*OPI_BAYES*	*OPI_VOCAB*
		Model I	Model II	Model III	Model IV
Intercept		-1.0614***	-1.3898***	-0.3068***	-0.3308***
		(-11.12)	*(-14.48)*	*(-5.50)*	*(-5.47)*
OPIf $_{[-10;-2]}$	+	**0.0306*****	**0.2763*****	**0.0071****	**0.0159***
		(3.24)	***(11.18)***	***(2.37)***	***(1.96)***
OPInf $_{[-10;-2]}$	+	**-0.0175**	**0.0827*****	**0.0047**	**0.0091**
		(-1.27)	***(3.22)***	***(1.14)***	***(1.04)***
PRIOR_ESURP		0.2831***	0.2804***	0.1871***	0.1871***
		(42.74)	*(41.41)*	*(10.65)*	*(10.80)*
EXRET $_{[-10;-2]}$		0.0219***	0.0216***	0.0084***	0.0085***
		(6.65)	*(6.60)*	*(4.52)*	*(4.56)*
RPOPI		0.3589***	0.3510***	0.0577***	0.0590***
		(5.26)	*(5.27)*	*(2.78)*	*(2.74)*
SIZE		0.1125***	0.1508***	0.0239***	0.0267***
		(7.54)	*(10.49)*	*(3.13)*	*(3.32)*
MB		-0.0112**	-0.0035	0.0018	0.0024
		(-2.10)	*(-0.66)*	*(0.82)*	*(1.13)*
ANL		-0.1111***	-0.0646**	0.0099	0.0130
		(-3.93)	*(-2.24)*	*(0.54)*	*(0.68)*
INST		0.1951**	0.0715	0.1976***	0.1891***
		(2.55)	*(0.94)*	*(4.90)*	*(4.69)*
Q4		0.2996***	0.3421***	-0.0923***	-0.0892***
		(6.37)	*(6.96)*	*(-4.91)*	*(-4.92)*
LOSS		0.8535***	0.8899***	-0.0576**	-0.0560**
		(17.54)	*(18.99)*	*(-2.26)*	*(-2.25)*
N		30,815	30,815	28,784	28,784
Adj. R^2 (%)		9.03	9.61	5.18	5.18
p-value of F-test $\beta_1 = \beta_2$		*<0.01*	*<0.01*	*0.65*	*0.62*

Empirical analysis

Aggregate opinion in individual tweets and announcement returns:

- Twitter information predicts stock returns, and the value relevance of the aggregate opinion provided by Twitter for stock returns stems not only from predicting the immediate short-term earnings surprise, but also from other information relevant to stock valuation.
- Tweets that contain earnings, fundamental, and trade-related information provide information relevant to both earnings as well as stock valuation. $OPInf$, on the other hand, provides information irrelevant for short-term earnings yet still useful for valuation.

- Abnormal Stock Returns Around Earnings Announcements and Twitter Opinion

Variable	Expected Sign	Coefficient (t-statistic)			
		OPI_BAYES	*OPI_VOCAB*	*OPI_BAYES*	*OPI_VOCAB*
		Model I	Model II	Model III	Model IV
Intercept		-0.2146** *(-2.38)*	-0.2878*** *(-3.10)*	0.3298* *(1.95)*	0.2469 *(1.36)*
$\boldsymbol{OPI}_{[-10;-2]}$	+	**0.0599*** *(3.69)***	**0.2360*** *(2.83)***	**0.0532*** *(4.06)***	**0.1690* *(1.85)***
FE				1.3922*** *(10.54)*	1.3921*** *(10.51)*
$EXRET_{[-10;-2]}$		-0.0127 *(-1.64)*	-0.0129* *(-1.83)*	-0.0225** *(-2.34)*	-0.0225** *(-2.41)*
RPOPI		0.1633 *(1.18)*	0.1770 *(1.44)*	0.0554 *(0.39)*	0.0646 *(0.41)*
ANL		-0.1176 *(-1.56)*	-0.0637 *(-0.85)*	-0.2810*** *(-2.84)*	-0.2333** *(-2.42)*
INST		0.8296*** *(3.38)*	0.7566*** *(3.14)*	0.3958** *(2.08)*	0.3543* *(1.87)*
Q4		0.1895 *(1.15)*	0.2240 *(1.16)*	0.3778*** *(2.81)*	0.4032*** *(2.65)*
LOSS		-0.4554** *(-2.24)*	-0.4524** *(-2.47)*	-0.0372 *(-0.12)*	-0.0372 *(-0.12)*
N		33,114	33,114	29,388	29,388
Adj. R^2 (%)		0.18	0.23	5.29	5.31

- Original and Dissemination Tweets

Variable	Expected Sign	Coefficient (t-statistic) *OPI_BAYES* Model I	*OPI_VOCAB* Model II	*OPI_BAYES* Model III	*OPI_VOCAB* Model IV
Intercept		-0.2142**	-0.3065***	0.3288*	0.2195
		(-2.48)	*(-2.95)*	*(1.72)*	*(1.19)*
***OPIo*$_{[-10;-2]}$**	+	**0.0348****	**0.2227*****	**0.0169**	**0.1570*****
		(2.55)	***(6.75)***	***(0.82)***	***(4.40)***
***OPId*$_{[-10;-2]}$**	+	**0.0547*****	**0.1038**	**0.0559*****	**0.0759**
		(4.45)	***(1.37)***	***(3.92)***	***(0.92)***
FE				1.3922***	1.3915***
				(10.29)	*(9.85)*
EXRET$_{[-10;-2]}$		-0.0128*	-0.0125*	-0.0225**	-0.0221**
		(-1.85)	*(-1.72)*	*(-2.19)*	*(-2.55)*
RPOPI		0.1618	0.1821	0.0540	0.0680
		(1.24)	*(1.45)*	*(0.37)*	*(0.45)*
ANL		-0.1186*	-0.0396	-0.2815***	-0.2116**
		(-1.76)	*(-0.48)*	*(-2.80)*	*(-2.16)*
INST		0.8295***	0.7026***	0.3987**	0.3207*
		(3.41)	*(2.76)*	*(1.99)*	*(1.72)*
Q4		0.1893	0.2283	0.3764***	0.4059***
		(1.09)	*(1.19)*	*(2.77)*	*(2.59)*
LOSS		-0.4557**	-0.4339**	-0.0380	-0.0246
		(-2.29)	*(-2.37)*	*(-0.12)*	*(-0.08)*
N		33,114	33,114	29,388	29,388
Adj. R^2 (%)		0.18	0.26	5.29	5.33
p-value of F-test $\beta_1 = \beta_2$		*0.62*	*0.16*	*0.33*	*0.36*

• Fundamental and Non-Fundamental Tweets

Variable	Expected Sign	Coefficient (*t-statistic*)			
		OPI_BAYES	*OPI_VOCAB*	*OPI_BAYES*	*OPI_VOCAB*
		Model I	Model II	Model III	Model IV
Intercept		-0.2140** *(-2.36)*	-0.2989*** *(-3.18)*	0.3358* *(1.92)*	0.2242 *(1.15)*
$\boldsymbol{OPIf_{[-10;-2]}}$	+	**0.0467** (2.52)**	**0.1745** (2.05)**	**0.0414** (2.21)**	**0.1366* (1.70)**
$\boldsymbol{OPInf_{[-10;-2]}}$	+	**0.0619*** (4.18)**	**0.1497*** (2.68)**	**0.0562*** (3.89)**	**0.0929 (1.55)**
FE				1.3920*** *(10.35)*	1.3918*** *(10.37)*
$EXRET_{[-10;-2]}$		-0.0129* *(-1.86)*	-0.0129* *(-1.75)*	-0.0227** *(-2.40)*	-0.0224** *(-2.34)*
RPOPI		0.1536 *(1.20)*	0.1913 *(1.63)*	0.0462 *(0.32)*	0.0735 *(0.48)*
ANL		-0.1220* *(-1.73)*	-0.0454 *(-0.59)*	-0.2865*** *(-2.88)*	-0.2156** *(-2.14)*
INST		0.8294*** *(3.42)*	0.7122*** *(3.10)*	0.3935* *(1.92)*	0.3287* *(1.69)*
Q4		0.1897 *(1.11)*	0.2242 *(1.14)*	0.3778*** *(2.74)*	0.4035*** *(2.70)*
LOSS		-0.4517** *(-2.43)*	-0.4440** *(-2.43)*	-0.0341 *(-0.11)*	-0.0317 *(-0.10)*
N		33,114	33,114	29,388	29,388
Adj. R^2 (%)		0.18	0.25	5.29	5.32
p-value of F-test $\beta_1 = \beta_2$		*0.67*	*0.78*	*0.69*	*0.64*

Empirical analysis

Aggregate opinion in individual tweets and the information environment:

- Aggregate Twitter opinion plays a greater role in predicting announcement returns for firms in weak information environments.
- The incremental predictive ability for firms in weak information environments is driven by original tweets.

- Role of the Information Environment

Variable	Expected Sign	Coefficient (t-statistic)					
		OPI_B	*OPI_V*	*OPI_B*	*OPI_V*	*OPI_B*	*OPI_V*
		Model I	Model II	Model III	Model IV	Model V	Model VI
Intercept		-0.1667	-0.2064*	-0.1707	-0.2267	-0.1637	-0.2105
		(-1.26)	*(-1.37)*	*(-1.09)*	*(-1.49)*	*(-1.17)*	*(-1.42)*
POORINFO		-0.0780	-0.1454	-0.0749	-0.1426	-0.0846	-0.1574
		(-0.60)	*(-1.06)*	*(-0.54)*	*(-1.21)*	*(-0.63)*	*(-1.33)*
$\boldsymbol{OPI_{[-10;-2]}}$	+	**0.0648*** **	**0.1957** **				
		(4.07)	***(2.37)***				
$\boldsymbol{OPI_{[-10;-2]} \times POORINFO}$	+	**-0.0250**	**0.2144** **				
		(-0.66)	***(2.10)***				
$\boldsymbol{OPIo_{[-10;-2]}}$	+			**0.0319**	**0.1565*** **		
				(1.20)	***(4.48)***		
$\boldsymbol{OPIo_{[-10;-2]} \times POORINFO}$	+			**0.0157**	**0.3626** **		
				(0.15)	***(2.17)***		
$\boldsymbol{OPId_{[-10;-2]}}$	+			**0.0659*** **	**0.1160**		
				(6.62)	***(1.28)***		
$\boldsymbol{OPId_{[-10;-2]} \times POORINFO}$	+			**-0.0586**	**-0.0544**		
				(-0.90)	***(-0.69)***		
$\boldsymbol{OPIf_{[-10;-2]}}$	+					**0.0546*** **	**0.1530* **
						(2.58)	***(1.73)***
$\boldsymbol{OPIf_{[-10;-2]} \times POORINFO}$	+					**-0.0415**	**0.1095**
						(-1.16)	***(0.90)***
$\boldsymbol{OPInf_{[-10;-2]}}$	+					**0.0604*** **	**0.1112* **
						(2.92)	***(1.82)***
$\boldsymbol{OPInf_{[-10;-2]} \times POORINFO}$	+					**0.0111**	**0.2341* **
						(0.13)	***(1.87)***

Supplementary test

SeekingAlpha Coverage

- Chen et al. (2014) shows that user-generated research reports posted on the SeekingAlpha portal help predict stock returns.
- We rerun the regressions (2) after deleting all observations where the firm in question had a report on SeekingAlpha over the same time period [-10;-2].
- Results: the relation between our opinion variables and $EXRET$ stays robust. Information from SeekingAlpha can't confound our findings.

- Abnormal Stock Returns Around Earnings Announcements and Twitter Opinion, in Subsample without SeekingAlpha Coverage

Variable	Expected Sign	Coefficient (t-statistic)					
		OPI_BAYES	OPI_VOCAB	OPI_BAYES	OPI_VOCAB	OPI_BAYES	OPI_VOCAB
		Model I	Model II	Model III	Model IV	Model V	Model VI
Intercept		-0.1948* *(-1.85)*	-0.2568** *(-2.31)*	-0.1949* *(-1.73)*	-0.2729** *(-2.47)*	-0.1940* *(-1.87)*	-0.2666*** *(-2.72)*
OPI$_{[-10;-2]}$	+	**0.0617***** ***(3.74)***	**0.2073**** ***(2.46)***				
OPIo$_{[-10;-2]}$	+			**0.0254**** ***(2.05)***	**0.2080***** ***(5.90)***		
OPId$_{[-10;-2]}$	+			**0.0584***** ***(3.84)***	**0.0861** ***(1.11)***		
OPIf$_{[-10;-2]}$	+					**0.0522**** ***(2.46)***	**0.1538*** ***(1.95)***
OPInf$_{[-10;-2]}$	+					**0.0472**** ***(2.54)***	**0.1413**** ***(2.27)***

Supplementary test

Difference in opinions between Twitter and traditional media

- We examine whether this predictive ability is stronger in settings where the aggregate Twitter opinion differs greatly from the traditional media opinion.
- We partition our sample into three subsamples based on the absolute difference between OPI and $RPOPI$.
- Results: Twitter information is relevant for predicting earnings announcement returns particularly when the aggregate Twitter opinion differs from the opinion in traditional media sources.

Supplementary test

Size of the “Twitter Crowd”

- One of the stated assumptions underlying the Wisdom of Crowds is that the “crowd” has enough participants, such that the noise in individual opinion is diversified away, and the “truth” emerges.
- We first partition our sample into two subsamples based on the median number of distinct tweet users per firm-quarter, and then replicate the tests (2).
- Results: the Wisdom of Crowds needs a nontrivial number of distinct users providing their insights for the information to be useful to capital market participants.

• Abnormal Stock Returns Around Earnings Announcements and Twitter Opinion, by Number of Distinct Users Per Firm-Quarter

Variable	Expected Sign	Coefficient *(t-statistic)* Less than Five Distinct Users *OPI_BAYES* Model I	*OPI_VOCAB* Model II	Five or More Distinct Users *OPI_BAYES* Model III	*OPI_VOCAB* Model IV
Intercept		0.3156**	0.1878	-0.8413***	-0.8580***
		(2.55)	*(1.05)*	*(-3.25)*	*(-3.82)*
$OPI_{[-10;-2]}$	+	**0.0126**	**0.2758**	**0.0606*** **	**0.1272* **
		(0.27)	***(1.41)***	***(4.02)***	***(1.68)***
$EXRET_{[-10;-2]}$		0.0036	0.0030	-0.0265***	-0.0263***
		(0.28)	*(0.24)*	*(-3.94)*	*(-3.73)*
RPOPI		0.2396	0.2243	0.1822	0.1868
		(1.03)	*(0.91)*	*(1.08)*	*(1.02)*
ANL		-0.1974*	-0.1974*	0.0461	0.0781
		(-1.66)	*(-1.70)*	*(0.62)*	*(1.06)*
INST		0.6024	0.5984	0.9479***	0.8985***
		(1.60)	*(1.61)*	*(4.11)*	*(4.04)*
Q4		0.0559	0.0844	0.2497	0.2974
		(0.23)	*(0.32)*	*(1.45)*	*(1.46)*
LOSS		-0.3303**	-0.3360**	-0.5455**	-0.5417**
		(-2.14)	*(-2.17)*	*(-2.09)*	*(-2.11)*
N		15,678	15,678	17,436	17,436
Adj. R^2 (%)		0.04	0.08	0.37	0.37

Supplementary test

Twitter usage intensity

- Not all users are equally active on Twitter. The top one percent of Twitter users put out 542,890 tweets in our sample (62.4%).
- To assess the influence of the top users on our findings, we use two subsamples: one containing tweets only from the top one percent users and one with the remaining tweets.
- Results: when more sophisticated or credible users tweet, their opinion, whether positive or negative, is important for the capital market.

- Abnormal Stock Returns Around Earnings Announcements and Twitter Opinion, by Number of Tweets per Distinct User

Variable	Expected Sign	Coefficient (*t-statistic*)			
		Excluding Top 1% Users (*Users with Less than 159 Tweets in our Sample*)		Top 1% Users (*Users with More than 159 Tweets in our Sample*)	
		OPI_BAYES	*OPI_VOCAB*	*OPI_BAYES*	*OPI_VOCAB*
		Model I	Model II	Model III	Model IV
Intercept		-0.5116*** (*-4.75*)	-0.5553*** (*-4.25*)	-0.2711** (*-2.27*)	-0.3586*** (*-3.11*)
$\boldsymbol{OPI_{[-10;-2]}}$	+	**-0.0002** (***-0.02***)	**0.2109***** (***2.79***)	**0.0721***** (***4.54***)	**0.2623***** (***2.95***)
$EXRET_{[-10;-2]}$		-0.0153* (*-1.66*)	-0.0161 (*-1.64*)	-0.0146* (*-1.88*)	-0.0147* (*-1.92*)
RPOPI		0.2429** (*2.17*)	0.2415** (*2.07*)	0.1271 (*0.98*)	0.1466 (*1.08*)
ANL		-0.0846 (*-1.51*)	-0.0438 (*-0.70*)	-0.0883 (*-1.04*)	-0.0264 (*-0.31*)
INST		1.0286*** (*3.31*)	0.9580*** (*3.12*)	0.7870*** (*3.39*)	0.6997*** (*3.23*)
Q4		0.1130 (*0.52*)	0.1254 (*0.54*)	0.2793** (*2.07*)	0.3250** (*2.28*)
LOSS		-0.4695** (*-2.17*)	-0.4605** (*-2.25*)	-0.4913** (*-2.20*)	-0.4847** (*-2.23*)
N		20,671	20,671	29,936	29,936
Adj. R^2 (%)		0.20	0.26	0.22	0.28

Supplementary test

Extending the Twitter opinion window

- We measure Twitter opinion over a longer horizon, from day -30 to day -2.
- Results: the aggregate Twitter opinion measured over longer-term horizons is relevant to capital market participants.

Additional sensitivity tests

- We consider three sets of alternate deflators for *OPI_BAYES* and *OPI_VOCAB:* remove the deflators, firm size, the total number of tweets pertaining to the firm.
- The results are unaltered for all three sets of alternative specifications.

Conclusion

- The aggregate Twitter opinion helps predict quarterly earnings, and abnormal returns around earnings announcements.
- Twitter serves as a source of new information as well as a vehicle for the dissemination of existing information.
- Only $OPIf$ is important for predicting quarterly earnings. However, both $OPIf$ and $OPInf$ are associated with announcement returns.
- The aggregate Twitter opinion plays a greater role in predicting announcement returns for firms in weak information environments.