Communication and Comovement:

Evidence from Online Stock Forums*

Lei Jiang, Jinyu Liu, and Baozhong Yang

January 2016

Abstract

We develop a model of investor communication generating return comovement and test its predictions using a novel dataset on an active online stock forum in China. For each stock, we consider its "related stocks," which are frequently discussed in the sub-forum dedicated to the given stock. We find substantial comovement among the returns of a stock and related stocks. Comovement is greater when related stocks are more frequently discussed. Further, the effect of frequent communication on comovement is stronger for stocks associated with higher information asymmetry. Our results are robust in tests using a forum outage event as a natural experiment and tests controlling for media coverage, market, industry, and economic variables. Our findings highlight the impact of investor communication on stock return covariance.

Keywords: Comovement; Asset Returns; Communication

JEL Classification: G12, G14, D83

---

* We are grateful for helpful comments from Vikas Agarwal, Haiqiang Chen, Lauren Cohen, Zhi Da, Will Goetzman, Lixin Huang, Fuwei Jiang, Raymond Kan, Omesh Kini, Laura Xiaolei Liu, Ruichang Lu, Kevin Mullally, Lin Peng, Michael R. Powers, Zhen Shi, Ajay Subramanian, Rob Stambaugh, Yuehua Tang, Paula Tkac, Larry Wall, Baolian Wang, Hao Wang, Neng Wang, Rui Zhong, Chunyang Zhou, Guofu Zhou, and participants at the Central University of Finance and Economics, Federal Reserve Bank of Atlanta, Georgia State University, Peking University, the 2014 China Finance and Accounting Conference, the 2014 Chinese Finance Ph.D. Symposium at Xiamen, and the 2015 China International Conference on Finance. We appreciate the financial support from the 2015 CICF Yihong Xia Best Paper Award. Lei Jiang gratefully acknowledges financial support from the AXA research fund. This research is also supported by Tsinghua National Laboratory for Information Science and Technology. The authors thank Yaru Huang and Jie Liu for their excellent research assistance. Lei Jiang, School of Economics and Management, Tsinghua University, Beijing, 100084, China; Tel: 86-10-62797084; E-mail: jianglei@sem.tsinghua.edu.cn. Jinyu Liu, School of Economics and Management, Tsinghua University, Beijing, 100084, China; Tel: 86-137-2000-1753; E-mail: liujy.12@sem.tsinghua.edu.cn. Corresponding author: Baozhong Yang, Robinson College of Business, Georgia State University, Atlanta, GA 30303, USA; Tel: 1-404-434-7350; E-mail: bzyang@gsu.edu.

1. Introduction

One fundamental question in financial economics is how asset prices are determined. In the rational expectations paradigm, price variations reflect changes in fundamental values. However, the empirical literature documents that there can be comovement in stock prices that is difficult to explain by fundamental values.[1] Understanding the source and extent of comovement can shed light on the structure of asset prices and facilitate the design of portfolio management strategies.

In this paper, we study whether communication among investors can generate comovement of stock returns. In particular, we directly measure investor communication using a novel dataset for online stock forums in China. We document substantial comovement among stocks that are discussed by investors on online forums and study the factors that influence such comovement.

To motivate our empirical tests, we develop a simple Grossman and Stiglitz (1980)–type model in which investors communicate before trading. The model shows that asset returns can exhibit comovement beyond what is implied by fundamental values when investors communicate repeatedly. In the model, investors receive a sequence of signals when communicating with one another and update their beliefs before trading. Due to incomplete information or limited attention, investors do not fully incorporate the consequences of repeated communication in their beliefs. As a result, the model predicts that communication among investors can generate comovement in stock prices.

The model also predicts that comovement in asset returns is positively related to the frequency with which investors communicate before trading. Intuitively, more frequent communication leads to a greater dependence of investor beliefs on common signals, resulting in greater comovement. Further, the model

[1] See, for example, Lee, Shleifer, and Thaler (1991), Pindyck and Rotemberg (1993), and Froot and Dabora (1999).

predicts that the effect of communication on comovement is more pronounced when investors have less accurate beliefs, that is, for stocks associated with greater information asymmetry. The intuition is that, for stocks subject to greater information asymmetry, communication among investors exerts a larger influence on investor beliefs.

We test these predictions using a unique dataset from the East Money Stock Forum, one of the most active online stock forums in China. The Chinese stock market provides an ideal environment to study investor behavior. Established in the 1990s, the modern Chinese stock market has developed rapidly but still suffers from a number of issues, such as the irrationality and immaturity of individual investors (e.g., Xu (2001) and Wang, Shi, and Fan (2006)). While the importance of institutional investors has increased over time, individual investors still dominate trading. At the end of 2007, individual investors held 51.3% of the Chinese stock market by value, while institutional investors held 42.3%, and the government held 6.4%.[2] Trading by retail investors accounts for the majority of trades in China and has been used to explain the high market volatility in 2015.[3] In the Chinese stock market, individual investors frequently exchange information and ideas on online forums.[4] Whereas such communication can help to propagate and incorporate information into stock prices, it can also potentially lead to distortions in the market through mechanisms discussed above.

[2] For the data on equity holdings across investor categories, see the 2011 Annual Report of the China Securities Depository and Clearing Corporation Limited. The data are also available on the website http://daily.cnnb.com.cn/dnsb/html/2009-05/06/content_83379.htm.

[3] For example, see "FT Explainer: Why are China's stock markets so volatile?" July 2, 2015, *Financial Times*, available at http://www.ft.com/intl/cms/s/0/e5af8da0-1fc7-11e5-aa5a-398b2169cf79.html#axzz3xA55nFIu

[4] For example, an internet survey shows that 65.9% of individuals are willing to share information and ideas on online forums (the Sixth Survey of Chinese Internet Community Development (2010) by iResearch, available at http://zz.comsenz.com/2010publish/).

For any given stock, there is a sub-forum in the online forum devoted to discussion about it. We refer to the stock that the sub-forum focuses on as the *target* stock of the sub-forum. Investors are also free to discuss other stocks in a sub-forum. Based on our model, we expect the returns of the stocks discussed on the same sub-forum to have comovement. To test this hypothesis, for any target stock, we consider the most frequently discussed stocks (henceforth referred to as "most related stocks") on the target stock's sub-forum. We construct a *related portfolio* that consists of the top five related stocks for every target stock in each month. We then estimate time series regressions of target stock returns on the returns of their related portfolios to examine their correlation. We find that the correlation between the returns of a stock and those of its related portfolio is positive and highly significant, even after controlling for market and industry returns. This comovement is also economically significant; for example, a 1% increase in the related portfolio return is associated with a 0.21% increase in the target stock return.

To address the concern that the correlation may be spuriously generated by a temporal trend or comovement among industries, we conduct a falsification test. We first create for each target stock a *placebo* portfolio that consists of several placebo stocks randomly selected in the industries of related stocks. We then estimate the same regressions, replacing the returns of related portfolios with those of the placebo portfolios. We find the coefficients on the target stock's return in these regressions to be insignificant, suggesting that the comovement we document is unlikely to be caused by temporal or industry factors.

We next examine the prediction on the relation between the frequency of communication and stock comovement. We create a proxy variable for communication frequency by computing the number of investor posts about the top related stocks in the sub-forum for a target stock. We then include this frequency and its interaction with the related stock portfolio return as independent variables in the regressions of the target stock returns. We find that more frequent communication leads to higher comovement between the return of the target stock and its related stocks.

We then investigate the prediction that the effect of communication on return comovement is larger for stocks associated with greater information asymmetry. We use three proxy variables for the information asymmetry of stocks: stock illiquidity, market capitalization, and analyst coverage. We divide our sample of stocks into five quintile groups according to each of the information asymmetry variables and conduct our regressions separately for each group. Consistent with our model's prediction, we find that for more illiquid, smaller, and less covered stocks, the frequency of forum discussion has a greater effect on stock comovement.

To alleviate the concern about endogeneity in our results, we employ an exogenous variation in the extent of investor communication caused by an outage in the East Money forum in June 2010 (henceforth, the "outage month"). We show that communication in the online forum in the outage month is significantly lower than the months immediately before and after. We re-estimate our tests of comovement separately for the outage and adjacent months and find the comovement in the outage month to be the lowest, suggesting causality in our main results.

We conduct a large number of robustness tests. First, we carry out a time series robustness test by conducting our tests separately for two equal subperiods of our overall time period. Second, we use the number of *clicks* the posts receive (instead of the number of posts) to proxy for the frequency of investor communication and to define the portfolio of related stocks. Third, we control for Fama-French factors in our tests to address the possibility that comovement arises from style investing. Fourth, we include a host of industry, market, and macroeconomic variables in our tests. Fifth, we use the absolute values of returns in our tests to address the possibility of negative comovement. In all of these specifications, we find that our main results to be robust.

Further, we conduct a set of additional tests to rule out alternative explanations of our results. First, we control for the level of investor communication on target stocks in their own sub-forums to isolate the incremental effect of communication on related stocks in the same sub-forums. This can help to address the possibility that communication and comovement can both be correlated with general trends in investor attention or habitats. In particular, we i) include the portfolio returns of the most discussed target stocks in our tests, and ii) consider a relative communication intensity measure that is normalized by the total number of posts discussing the target stock. We find that our main results and the forum outage results continue to hold.

Second, we address the concern that our findings on comovement and communication may be driven by the public revelation of fundamental information about the firms. We collect news about firms from various media sources, including newspapers, television and online media. We control for media coverage about both target and related stocks in our tests and find our results to be qualitatively similar.

Our paper contributes to the literature that studies comovement in asset returns and its relation to investor behavior. To the best of our knowledge, this paper is the first to document the comovement of stock returns generated by communication in a social network. Pindyck and Rotemberg (1993) find excess comovement in stock prices. Froot and Dabora (1999) show that twin stocks, such as Royal Dutch and Shell, comove more with the local markets in which they are traded. Morck, Yeung, and Yu (2000) demonstrate that there is more stock price comovement in poor countries than in rich countries. Vijh (1994), Barberis, Shleifer, and Wurgler (2005), and Greenwood (2008) present evidence on an increase (decrease) in the correlation of a stock with other index stocks when it is added to (deleted from) the index portfolio. Kumar and Lee (2006) demonstrate that herding in individual investors' trades can lead to comovement. Pirinsky and Wang (2006) show that stocks with proximate headquarter locations comove more. Green and Hwang (2009) document that, after splits, stocks comove more with other lower-priced stocks. Leung, Agarwal, Konana, and Kumar (2013) find that stocks co-viewed by visitors on the Yahoo! Finance website exhibit comovement. We complement this literature by using a unique dataset on the communication of individual investors to study the effects of communication and its frequency on comovement.

Our paper is also related to the literature on information transmission in social networks and its effects on economic agents' behavior and asset prices (e.g., Hong, Kubik, and Stein (2006), Cohen, Frazzini, and Malloy (2008, 2010), Chawla, Da, Xu, and Ye (2015)). Analogous to these strands of literature, we show that communication among investors can have a substantial impact in the financial markets. Since the communication procedure reflects investor attention to online discussions, our paper is also related to the literature on investor attention and asset prices (e.g., Merton (1987), Peng (2005), Peng and Xiong (2006), Da, Engelberg, and Gao (2011) and Hirshleifer, Lim, and Teoh (2011)). Further, our paper is related to the literature on the effects of internet message board discussions on stock returns and volatility (e.g., Antweiler and Frank (2004) and Das and Chen (2007)). Whereas these papers consider the effects of internet messages on the return and volatility of individual stocks and the aggregate market, we focus on the comovement among different stocks that investors discuss on the same forum.

Finally, our model is related to a stream of theoretical literature that explains the comovement of stock prices from different angles. Calvo (1999) proposes a model in which the forced selling of emerging market securities may serve as a negative signal for uninformed investors, resulting in a market collapse. Kodres and Pritsker (2002) develop a multi-asset, rational expectations model on financial contagion arising from cross-market rebalancing by investors who experience idiosyncratic shocks. Peng and Xiong (2006) show that limited attention by investors can generate comovement in stock returns. Veldkamp (2006) uses the endogenous and costly production of information by investors to explain comovement in asset prices. Yang (2013) demonstrates that communication can produce a concentrated factor structure in asset returns, with assumptions about the structure of the social network. Our model emphasizes the role of repeated communication in generating price comovement and has the advantage of being testable using observable data.

2. The Model

In this section, we develop a simple Grossman and Stiglitz (1980)–type model to analyze the effects of communication on comovement in stock prices and motivate our empirical analyses. The basic structure of our model is similar to that of Veldkamp (2006). Consider an economy with two dates, $t = 0,1.$ There is a continuum of investors of unit mass with identical preferences. The preference function is dependent on the terminal wealth W at date 1 as follows:

$$U(W) = E[-e^{-\gamma W}]. \tag{1}$$

There is a risk-free asset and two risky assets in the economy. For simplicity, the risk-free rate is assumed to be zero. The values of the two assets at date 1 are given by stochastic quantities

$$\begin{aligned} v_1 &= x + y_1, \\ v_2 &= x + y_2, \end{aligned} \tag{2}$$

where x is a common component and $y_i, i = 1, 2,$ are idiosyncratic components. Note that without loss of generality, we assume that the coefficient on x to be 1 for both assets. The shocks x and y_i are independently and normally distributed. We assume that investors have identical prior beliefs as follows:

$$x \sim N(\mu_0, \sigma_0^2), y_i \sim N(\mu_{y_i}, \sigma_{y_i}^2),\ i = 1, 2. \tag{3}$$

Investors are endowed with initial wealth W_0 and trade after they form their posterior beliefs about the assets at date 0. The aggregate supply of asset i is S_i for i = 1, 2. The equilibrium is defined by the usual market clearing conditions and optimization of the investors' problem.

At date 0, all investors receive signals about the asset values before they trade the assets. For simplicity, we assume that they receive a sequence of signals, $z_j,\ j = 1, 2, \ldots, N,$ before they trade. Because we are concerned about potential comovement of stock prices, we focus on the case where the signals contain information about the common component x in the asset values.[5] Specifically,

$$z_j = x + \epsilon_j, \quad \epsilon_j \sim N(0, \sigma_\epsilon^2), \tag{4}$$

where ϵ_j are independent of the fundamental shocks x and y_i. In the model, for tractability, the source of these signals is treated as exogenous. In our context, these signals are posts on a message board or

[5] In our model, if communication is only about idiosyncratic shocks to stock values, then comovement is independent of communication. It is possible to build a model in which investors communicate only about idiosyncratic shocks and cause lower comovement among stock returns. Such a model requires the assumption of noisy aggregate supplies of stocks or the existence of other noisy traders in the market. We thank an anonymous referee for suggestion of this possibility. We address this possibility in our empirical tests in Section 4.6.

online stock forum. Some investors may have obtained information about stock values and posted their information to share with other investors. Such information sharing can be rational. For example, if an investor has completed building his positions, then revealing the information publicly will help stock prices to converge to the fundamental values faster and thus help the investor realize his profits earlier. Furthermore, as shown later in this section, if investors have limited attention or exhibit persuasion bias, even sharing publicly available information can lead to stronger comovement.

We begin by assuming that the signals z_j are independent signals, i.e., $\epsilon_j,\ j = 1, 2, ..., N,$ are independent of each other. We later consider the possibility that these signals are not independent. By Bayesian updating, we have the following proposition about the beliefs of the agents. For brevity, we include all proofs in the Internet Appendix.

Proposition 1. The investors have the following posterior beliefs about the common component:

$$x \sim N(\mu_N, \sigma_N^2), \tag{5}$$

where μ_N and σ_N^2 are given by:

$$\begin{aligned} \mu_N &= \frac{\sigma_0^{-2}}{\sigma_0^{-2} + N\sigma_\epsilon^{-2}}\mu_0 + \frac{N\sigma_\epsilon^{2}}{\sigma_0^{-2} + N\sigma^{-2}}\bar{z}, \quad \bar{z} = \frac{1}{N}\sum_{j=1}^{N} z_j, \\ \sigma_N^{-2} &= \sigma_0^{-2} + N\sigma_\epsilon^{-2}. \end{aligned} \tag{6}$$

Assume that an investor takes positions (α_1, α_2) in the risky assets at date 0; then, the date 1 wealth of the investors will be $W_0 + \sum_{i=1}^{2} \alpha_i (v_i - P_i),$ where P_i is the price of asset i at date zero. Therefore, investors choose their portfolios to solve the following optimization problem:

$$
\begin{aligned}
&\max_{(\alpha_1,\alpha_2)} E[-e^{-\gamma W} \mid I_N] \\
&s.t. \quad W = W_0 + \sum_{i=1}^{2} \alpha_i (v_i - P_i),
\end{aligned} \tag{7}
$$

where the expectation is taken with respect to investors' information set I_N after receiving all signals at date 0. The market clearing conditions together with (7) allow us to solve the asset prices.

Proposition 2. In equilibrium, the asset prices after communication are given by:

$$
\begin{aligned}
P_1 &= \mu_N + \mu_{y_1} - \gamma((\sigma_N^2 + \sigma_{y_1}^2)S_1 + \sigma_N^2 S_2), \\
P_2 &= \mu_N + \mu_{y_2} - \gamma(\sigma_N^2 S_1 + (\sigma_N^2 + \sigma_{y_2}^2)S_2).
\end{aligned} \tag{8}
$$

Using (6) and (8), we obtain the covariance of asset prices,[6]

$$
Cov(P_1, P_2) = \left(\frac{N\sigma_{\text{蝌}}^{-2}}{\sigma_0^{-2} + N\sigma_{\text{蝌}}^{-2}} \right)^2 Var(\overline{z}) = \left(\frac{N\sigma^{-2}}{\sigma_0^{-2} + N\sigma^{-2}} \right)^2 (\sigma_0^2 + \frac{1}{N}\sigma_{\text{ò}}^2). \tag{9}
$$

Note that the covariance of the intrinsic asset values is

$$
Cov(v_1, v_2) = Cov(x + y_1, x + y_2) = \sigma_0^2. \tag{10}
$$

The following proposition compares the covariances in fundamental values and asset prices.

Proposition 3. The covariances of fundamental values and asset prices satisfy

$$
Cov(v_1, v_2) > Cov(P_1, P_2). \tag{11}
$$

Therefore, when the signals received by investors are independent from each other and investors are fully rational, there is no comovement in asset prices beyond those in the fundamental values.

Next, we assume that the signals z_j are not independent from each other, yet investors still regard them as independent.[7] The motivation for this assumption is that it is unlikely that there are many

[6] Since the initial asset prices are constant, the covariance of prices here are equal to the covariance of changes in asset prices from the initial time. We follow the convention of studying changes in asset prices and their covariances in the framework of investors with Constant Absolute Risk Aversion (CARA) preferences and asset values with normal distributions, e.g., see Veldkamp (2006) and Banerjee (2011).

independent signals about firm values occurring in a short time period. Investors, however, have incomplete information about or pay limited attention to the sources of the signals (especially on online forums) and regard them as independent.[8] Our assumption is also similar to the persuasion bias of agents in DeMarzo, Vayanos, and Zwiebel (2003); that is, people fail to account for possible repetition of the information they receive.

For simplicity, we assume that all the signals z_j are identical and equal to $z = x + \epsilon$. This assumption does not change our results qualitatively. We now have the covariance of asset prices equal to

$$Cov(P_1, P_2) = \left(\frac{N\sigma_{\epsilon}^{-2}}{\sigma_0^{-2} + N\sigma_{\epsilon}^{-2}} \right)^2 Var(z) = \left(\frac{N\sigma^{-2}}{\sigma_0^{-2} + N\sigma^{-2}} \right)^2 (\sigma_0^2 + \sigma_{\epsilon}^2). \qquad (12)$$

The following proposition describes the properties of comovement in asset prices.

Proposition 4. i) The covariance of asset prices $Cov(P_1, P_2)$ is always greater than the covariance of fundamental values in the case where investors are fully rational.

ii) The following is always true:

$$\frac{\partial Cov(P_1, P_2)}{\partial N} > 0. \qquad (13)$$

iii) If $N\sigma_0^2 - 2\sigma_{\epsilon}^2 < 0,$ then

$$\frac{\partial^2 Cov(P_1, P_2)}{\partial N \partial \sigma_0} > 0. \qquad (14)$$

Part (i) of Proposition 4 shows that communication can give rise to stronger comovement when investors have limited cognition or exhibit persuasion bias. By part (ii), the model predicts that the extent

---

[7] Our results and intuition still hold in the case where investors treat the signals as correlated, so long as they underestimate the correlation among the signals. The results are available upon request from the authors.

[8] There is a large theoretical literature that studies incomplete information, limited investor attention, and asset prices. See, for example, Merton (1987), Peng (2005), Peng and Xiong (2006), and Hirshleifer, Lim, and Teoh (2011).

of comovement increases with the number of signals (N) that investors receive before trading. Intuitively, investors' beliefs and asset prices become more correlated when they receive a greater number of signals, but investors fail to consider the interdependence of these signals.

By part (iii) of Proposition 4, the model also predicts that the effect of communication on asset comovement is more pronounced for stocks subject to greater information asymmetry (higher σ_0).[9] The intuition is that for stocks with greater information asymmetry, communication among investors has a greater effect on their posterior beliefs and thus exerts a larger influence on stock return comovement.

3. Empirical Analysis

3.1 Data and Variables

We collect our data on investor communication by tracking all messages posted on an online forum: the East Money Stock Forum (*http://guba.eastmoney.com/*). We choose this forum because it is the oldest stock forum in China and one of the most active and influential forums.[10] When we search for the keywords "stock forum" on the most popular search engines in China (Baidu or Google (Hong Kong)), the East Money Stock Forum always appears among the top search results. Moreover, the forum is fully compatible with the East Money trading software, which is widely used by investors in China for placing orders to trade stocks. Investors can thus easily access the information posted on the stock forum when they use the software to trade. The East Money Stock Forum, therefore, provides a relatively

[9] The condition in part (iii) of Proposition 4 holds when the signals are not too precise relative to the prior beliefs of investors, which is likely to be the case for the online communications that we study in this paper.

[10] Recent studies (e.g., Hong, Jiang, Wang, and Zhao (2014) and Chang, Hong, Tiedens, Wang, and Zhao (2015)) use data from the East Money Stock Forum in their analyses.

representative and comprehensive dataset of communication among investors, which can be influential for stock trading and prices.

On the East Money Stock Forum (henceforth, the "forum"), there is a sub-forum for every stock. Investors can discuss and exchange information about a given stock in its sub-forum. We refer to the designated stock of a sub-forum as the *target stock*. On each such sub-forum, investors can also discuss other stocks, which we define as *related stocks* to the target stock of the sub-forum. Below are two example messages that discuss related stocks on the sub-forum for the target stock Wuhan Iron and Steel (stock exchange ID: 600005):

"The best sector in 2008 will be railroad industries; the indisputable leader in railroad stocks is Guangzhou-Shenzhen Railroad (601333)."

"Since FAW Automobile (000800) has tumbled, the prospect for Wuhan Iron and Steel won't be great."

As discussed in Section 2, communication on a sub-forum can potentially lead to comovement among the returns of a target stock and its related stocks.

Due to the limited availability of forum data prior to June 2008, we study the period from June 2008 to December 2012 in this paper. To ensure that there is sufficient discussion by investors on the forum, we focus on the sub-forums devoted to the component stocks of the Shanghai Stock Exchange (SSE) 180 Index, one of the most important benchmarks for the Chinese stock market. Similar to the S&P 500 index in the United States, the SSE 180 Index consists of stocks with large market capitalization. Besides being representative of the Chinese stock market, the SSE 180 stocks are associated with high trading volume, which helps attract investor attention. Therefore, there are large numbers of messages on the sub-forums dedicated to these stocks. We use stock returns and accounting data from the Resset Database (*http://www.resset.cn*) and the CSMAR Database (*http://www.gtarsc.com/*). During the period from 2008 to 2012, the composition of the SSE 180 Index experienced several adjustments, and a total of 296 stocks were included in the index. Our sample of stock return data includes 255,844 stock-trading-day observations for these stocks.

We download investor messages on the forum using a Perl program. Our program can retrieve message information, such as the identifiers of stocks mentioned in the message and the posting time of the message. Messages can be posted on both trading and non-trading days. Since the messages posted on non-trading days also convey information to investors, we include them in our sample. We retrieve a total of 13,528,136 messages for our sample of stocks from 2008 to 2012.

We use the daily return of stocks (*Ret*), the daily market return (*MKTRet*), and the daily industry sector return for a given stock (*INDRet*) in our empirical tests. To capture the returns of other stocks discussed on a sub-forum, we define a related stock return variable as follows. For each stock month, we consider all messages posted on a target stock's sub-forum during the month. We record the frequency of each related stock mentioned in these messages and rank the related stocks by such frequencies. We then construct the portfolio of the five most related stocks on a monthly basis. Note that although we require the target stock to be included in the SSE 180 Index, we do not impose the same restriction on its related stocks. We calculate the daily mean related stock return, or *MRR*, of the target stock as the daily average stock return of this portfolio, i.e.,

$$MRR_{m,t} = \frac{1}{5}\sum_{j=1}^{5} Ret_{jm,t},$$

where j indicates the rank of the related stock by discussion frequency, m indicates the target stock, t indicates the date, $MRR_{m,t}$ is the date t daily return of stock m, and $Ret_{jm,t}$ is the date t daily return of the related stock j.[11] Table A1 in the Appendix shows an example of the top five related stocks for one target stock, Wuhan Iron and Steel (ID: 600005), in the SSE 180 Index during a six-month period in our sample. In this example, a top related stock is mentioned on the sub-forum in 2 to 15 posts each month. Despite the relatively small number of posts, each post receives on average 803.7 clicks and likely much greater attention from investors because they can browse the list of posts without actually clicking on individual posts. We use the total number of times that the top five related stocks are mentioned on a sub-forum in a month, *Freq*, as a proxy for the intensity of communication among investors.[12]

Further, we consider a number of (Chinese) market and macroeconomic factors in our analysis: *Inflation*, the monthly growth rate of the Consumer Price Index; *GDP Growth*, the monthly growth rate of real gross domestic product, interpolated from quarterly data; *Term Spread*, the difference between the long-term (10-year) treasury bond yield and the short-term (3-month) treasury bond yield (Welch and Goyal, 2008); *IPO Activity*, the number of new firms that make an initial public offering in a month; *Turnover*, the turnover rate of the stock market; and *Economic Index*, the indicator for status of the economy calculated by the National Bureau of Statistics of China. Panel A of Table 1 reports summary statistics of the variables used in our empirical tests.

[11] If the target stock has less than five related stocks in a month, we use the actual number of related stocks mentioned in the sub-forum of this stock in the calculation of *MRR*.

[12] We also use the total number of clicks that the top five related stocks receive as an alternative measure for the extent of communication and find our results to be robust.

[Insert Table 1 Here]

Panel B of Table 1 reports the average cross-sectional correlations for our key variables: stock return (*Ret*), mean related stock return (*MRR*), market return (*MKTRet*), and the other variables. We find that the return of a target stock on a stock sub-forum is positively related to the mean return of its related stocks, which is consistent with the prediction of our model. All correlations are significantly different from zero at the 5% level.

3.2 Communication and Comovement of Stock Returns

In this section, we study the comovement of returns of target stocks and their related stocks discussed in the same sub-forums. As discussed in Section 2, our model predicts that investors' communication about a group of stocks can generate comovement among these stocks.

We first conduct time series regressions of each stock's returns on the returns of its related stock return (*MRR*) to study the comovement among them. In particular, we estimate the following model for each stock:

$$Ret_{m,t} = \alpha_m + \beta_m MRR_{m,t} + \varepsilon_{m,t}, \qquad \text{(Model 1)}$$

where $Ret_{m,t}$ is the daily return of the forum target stock *m* and $MRR_{m,t}$ is the mean related stock return for stock *m*. A positive β_m suggests positive comovement between the forum target stock and related stock returns.

The comovement among stocks studied in Model 1 could be generated by a market-wide stock movement that drives the returns of both the forum target stock and its related stocks. To alleviate this concern, we include market returns on the right-hand side of the regressions and estimate the following model:

$$Ret_{m,t} = \alpha'_m + \beta_{1m} MRR_{m,t} + \beta_{2m} MKTRet_{m,t} + \varepsilon'_{m,t}. \qquad \text{(Model 2)}$$

In Model 2, the coefficient β_{1m} indicates the comovement between the stock and related stock returns after controlling for market returns.

Table 2 reports the distributions of *t*-statistics and significant coefficients across all stocks for Models 1 and 2. Panel A of Table 2 shows that the average coefficient β_m of the related stock return *MRR* in Model 1 across all stocks is 0.717. The coefficients are positive and significant at 1% levels for all 296 stocks, with an average *t*-statistic of 23.85. This evidence suggests that there is strong comovement among target stocks and their related stocks within a sub-forum. Panel B of Table 2 shows that the average coefficient β_{1m} in Model 2 across all stocks is 0.213, which is positive and economically significant, with a mean *t*-value of 4.12. On average, a 1% increase in *MRR* leads to an economically significant 0.21% increase in daily target stock return. Further, this coefficient is positive and significant at 1% levels for 161 (or 68%) out of 296 regressions and is insignificant (or negative) in only 86 regressions (19%). Therefore, after controlling for market-level changes, we continue to find significant comovement among target stocks and related stocks within sub-forums.

[Insert Table 2 Here]

When examining the coefficients in Models 1 and 2 across all target stocks, it is possible to compute the overall *t*-statistics to assess the joint significance of the stock-by-stock regressions. However, the simple *t*-statistic (following the Fama-Macbeth method) for the average coefficient is calculated under the premise that the estimation errors are independent across regressions, which may be violated in the cross-sectional setting, leading to potential biases. To allow for cross-sectional correlation across residuals, we calculate overall *t*-statistics using the methodology developed by Chordia et al. (2000) (see also Avramov et al. (2012)). In particular, we calculate the variance of the mean coefficients as:

$$Var(\hat{\bar{\beta}}) = \frac{1}{M^2}[\sum_{m=1}^{M} Var(\hat{\beta}_m) + \sum_{m=1}^{M} \sum_{n=1, n \neq m}^{M} Cov(\beta_m, \hat{\beta}_n)],$$

where the variance and covariance matrices of the regression coefficients are estimated by

$$Var(\hat{\beta}_m) = \frac{(\hat{\varepsilon}_m' \varepsilon_m)}{(T - k)} (X_m' X_m)^{-1},$$

$$Cov(\hat{\beta}_m, \beta_n) = \frac{(\hat{\varepsilon}_m' \varepsilon_n)}{(T - k)} (X_m' X_m)^{-1} (X_m' X_n)(X_n' X_n)^{-1}.$$

In the above, X_m is the matrix formed by the sample values of the independent variables in the time-series regression for stock *m*, and $\hat{\varepsilon}_m$ is the sample deviation vector for stock *m*.

Panel C of Table 2 reports the mean coefficients and the overall *t*-statistics of the stock-by-stock regressions. The mean coefficient of *MRR* is positive and significant at the 1% level for both Models 1 and 2. These results confirm our finding that there exists strong comovement among the returns of a target stock and other stocks discussed on the same stock sub-forum.

3.3 Placebo Test

In the previous section, we document the existence of comovement among returns of stocks discussed on the online forum. However, it is still possible that temporal trends or other unobservable temporal factors, rather than information sharing among investors, drive the correlations between stock returns. We address this potential concern by conducting a placebo test.

For each target stock and month, we randomly select five stocks from the same industries of the top five related stocks in that month to form a *placebo portfolio* of stocks. Similar to the construction of the actual related stock portfolios, we adjust the composition of the placebo portfolios on a monthly basis. We define $RANDRet_{m,t}$ as the average date t return of stocks in the placebo portfolio of target stock m. We then conduct stock-by-stock time series regressions by replacing the related stock returns in Model 2 with the placebo portfolio returns:

$$Ret_{m,t} = \alpha_m + \beta_{1m} RANDRet_{m,t} + \beta_{2m} MKTRet_{m,t} + \varepsilon_{m,t}. \qquad \text{(Model 3)}$$

Table 3 reports the results of these stock-by-stock regressions. Panel A shows that the average t-value is only 0.12 across all regressions. For 78% of the target stocks, the coefficients of the placebo portfolio returns in Model 3 are insignificant. This stands in stark contrast to the results for Models 1 and 2 in Table 2, where the coefficients are significant at the 10% level or higher for nearly 90% of the stocks. Panel B shows the overall *t*-statistics for the mean coefficients, following Chordia et al. (2000). Consistent with the above results, the mean coefficient of *RANDRet* is insignificant, with an overall *t*-value of 1.49. In sum, the results of our placebo tests suggest that the comovement among target stocks and their related stocks is not likely driven by temporal trends or other temporal factors.

[Insert Table 3 Here]

3.4 Communication Intensity and Return Comovement

According to our model, as the number of rounds of communication between investors increases, investors update their beliefs about the stocks, leading to greater comovement among stock returns. Therefore, we expect comovement to be higher for stocks that are subject to more intense discussion. In this section, we use the frequency with which stocks are discussed on sub-forums as a proxy for communication intensity and test this prediction.

We include the frequency variable (*Freq*) and its interaction with the return of related stocks (*MRR*) in our time series regressions and estimate the following model for each target stock:

$$\begin{aligned} Ret_{m,t} &= \alpha_m + \beta_{1m} MRR_{m,t} + \beta_{2m} Freq_{m,t} \times MRR_{m,t} \\ &+ \beta_{3m} Freq_{m,t} + \beta_{4m} MKTRet_{m,t} + \varepsilon_{m,t}. \end{aligned} \qquad \text{(Model 4)}$$

The coefficient of the interaction term between *Freq* and *MRR* in Model 4 captures the marginal effects of more frequent discussion on the comovement between the target stock and related stock returns. In the regressions, we use the standardized *Freq* variable with mean equal to zero and standard deviation equal to one to facilitate interpretation of the coefficients of *MRR* and *Freq*×*MRR*.

We report the results of these stock-by-stock regressions in Table 4. Panel A shows that the average coefficients of *MRR* and the interaction term *Freq*×*MRR* are both positive. The coefficients of the interaction term are significant at the 10% or higher levels in 40% of the regressions. Panel B shows that the overall *t*-statistics of the mean coefficients in these regressions are 43.77 for *MRR* and 17.49 for the interaction term, both significant at the 1% level.

Taken together, the evidence in this section suggests that comovement is concentrated among stocks that are more frequently discussed by investors, consistent with our theoretical prediction.

[Insert Table 4 Here]

3.5 Information Asymmetry, Communication, and Return Comovement

In this section, we examine the relationship among information asymmetry, communication, and the comovement of stock returns. Our model generates the cross-sectional prediction that the noisier investors' prior beliefs are, the stronger the effect of communication is on comovement. To test this prediction, we examine whether stocks with higher information asymmetry have higher levels of return correlation with their related stocks.

We use three variables to proxy for information asymmetry: illiquidity, firm size, and analyst coverage. First, we employ the widely used Amihud illiquidity measure (Amihud, 2002), which is calculated as follows:

$$Amihud_{m,t} = \sqrt{\left|r_{m,t}\right| / (P_{m,t} \times Vol_{m,t})},$$

where $r_{i,t}$ is the daily return of stock *i*, and $P_{i,t}$ and $Vol_{i,t}$ are the daily price and trading volume of stock *i*. We use the natural logarithmic transformation of the Amihud measure to mitigate the effects of any outliers. Second, we use the logarithm of stock market capitalization as a proxy for firm size and information asymmetry. We average all daily measures to obtain quarterly measures. Third, we use the number of analysts who covered a stock in the previous year as an additional proxy since greater analyst coverage provides more information to the public.

We use the above three proxy variables of information asymmetry to construct subsamples. Specifically, we divide the 296 target stocks into five quintile groups according to the value of the information asymmetry variable in the lagged quarter. We readjust the composition of the five groups quarterly. We then estimate the regression of Model 4 separately for each quintile over time and compare the differences in stock return comovement among the different groups. Table 5 reports the results of these subsample analyses.

[Insert Table 5 Here]

Panel A of Table 5 shows that the coefficient of the interaction term *Freq*×*MRR* is increasing (from 0.033 in the bottom quintile to 0.095 in the top quintile) as the illiquidity of the stock increases. The difference between the coefficients of *Freq*×*MRR* in the top and bottom quintiles is 0.061 and is statistically significant at the 1% level, with a *t*-value of 2.21. Panel B shows that the coefficient of the interaction term decreases as stock market capitalization increases (from 0.125 in the bottom quintile to 0.028 in the top quintile; the difference between the top and bottom quintiles is statistically significant with a *t*-value of -4.09). Panel C shows that the difference of the coefficients of *Freq*×*MRR* for stocks with the highest analyst coverage and those with the lowest analyst coverage is negative, yet insignificant. Since stocks with higher illiquidity, smaller size, and less analyst coverage are subject to higher information asymmetry, these results suggest that the effects of communication on return comovement are more pronounced for stocks with higher information asymmetry.

3.6 Exogenous Variation in Communication: Forum Outage

To address the possibility that communication and comovement may be driven by unobservable variables, we employ exogenous variations in the degree of investor communication to establish a causal relationship between communication and comovement.

We consider a disruption in services on the East Money forum as a natural experiment. In June 2010, a number of online posts complain about the difficulty in accessing the website. These posts are dated between June 9 and June 30, and appeared on the East Money forum and other websites, such as Baidu Knows. This forum outage should affect communication among investors. Indeed, the average number of total posts in a target stock forum drops to 756.9 in June 2010 (henceforth, the "forum outage month"), compared to 903.1 in the previous month of May 2010 and 772.9 in the next month of July 2010.

We find no evidence of major nationwide internet outage during that period in the media, suggesting that this outage is specific to the forum website. To further rule out the concern that this outage is related to technical problems that directly affect investors' trading, we compute the average retail trading volume (defined as trades with values less than RMB 20,000)[13] normalized by outstanding shares for the sample of target stocks. We find the average normalized retail trading measure to be 1.54, 1.78, and 1.90, in May, June, and July of 2010, respectively. While there is a slight increasing trend in retail trading, the outage does not reduce retail trading in the outage month. Therefore, the outage was forum-specific and provides an ideal exogenous shock to communication between investors.

Table 6 reports the estimation of the baseline regressions in Model 2 for the forum outage month and the months before and after, respectively. To the extent that investors communicate less during the outage month, we expect comovement among the target stock and related stocks to be weaker in the outage month, compared to other months. In Panel A of Table 6, we find that the coefficient of the related portfolio return (*MRR*) in the forum outage month is 0.125, lower than the coefficient in the month before (0.22) and the month after (0.178), with the differences statistically significant at the 1% levels.

[13] A number of studies have used small trades to proxy for retail trades, e.g., Barber, Odean, and Zhu (2008) in the U.S. market; Ng and Wu (2007) and Liao, Liu, and Wang (2011) in the Chinese market.

To further rule out that the difference we find above is due to other unobservable temporal factors, we conduct placebo tests using June 2009 and June 2011 as the placebo treatment months, and May and July of 2009 and 2011 as the control months. Panel B of Table 6 reports the results of the placebo tests. The coefficient of *MRR* in June 2009 and 2011 is 0.281, compared to 0.272 in May 2009 and 2011, and 0.279 in July 2009 and 2011. The differences among these coefficients are statistically insignificant.[14] Taken together, the evidence in this section supports the premise that investor communication causes stronger stock comovement.

[Insert Table 6 Here]

4. Robustness Tests

4.1. Time Series Robustness Tests

In this section, we perform a robustness test by conducting our main regressions in two approximately equal subperiods of our sample, i.e., the periods from June 2008 to June 2010 and from July 2010 to December 2012. We estimate the regressions of Models 1 through 4 separately for the two subperiods and report the results in Table 7. For simplicity, we report only the mean coefficients of the stock-by-stock regressions and the overall *t*-statistics. In Models 1 and 2, the coefficients of the mean returns of related stocks are positive and statistically significant in both subperiods. In the placebo tests of Model 3, the mean return of a randomly chosen portfolio has either an insignificant positive coefficient or a negative coefficient in the two subperiods. In Model 4, the coefficients of the interaction term between

[14] We also consider the Spring Festival, the most important holiday in China, as a shock to communication in our tests. Indeed, we find that investors post fewer messages in the Spring Festival month. Further, target and related stocks comove less in the Spring Festival month than in the adjacent months. We present these results in the Internet Appendix (Table IA.1).

Freq and *MRR* for the two subperiods are both significant at the 1% level. Overall, the results of the above subperiod analyses are consistent with our findings in the previous sections.

[Insert Table 7 Here]

4.2. Alternative Measure of Communication and Other Alternative Specifications

Our dataset allows us to define an alternative measure of the degree of investor communication by the number of clicks the messages receive on the forum as of the end of 2012. We use the total number of clicks received on messages about related stocks to rank and obtain the top five related stocks, and we form the portfolio of related stocks each month. We then re-estimate Models 1 and 2 using this new definition of related portfolio returns. We also repeat the estimation of Model 4 by replacing the number of posts (*Freq*) with the total number of clicks on the posts (*Clicks*). Table 8 reports the results, which are consistent with our previous results using the number of messages to proxy investor communication.

[Insert Table 8 Here]

To address the possibility that our results are driven only by comovement of stocks in the same industry, we remove related stocks that are in the same industry as the target stock in our construction of the related portfolio, and then repeat our tests in Models 1, 2, and 4. The results are again qualitatively similar to our previous findings (for brevity, we report the results in Table IA.2 in the Internet Appendix.)

In the previous tests, we form the portfolio of the most related stocks in the same month in which we examine the correlations of stock returns. One alternative explanation of our findings is that communication among investors could instead arise from comovement among the target stock and its related stocks. To address this concern, we form the related stock portfolios using the top five related stocks of the target stock in the *previous* month and investigate the comovement of stock returns in the current month. We estimate the regressions in Models 1, 2, and 4 with the above modification and find our results to be robust (see Table IA.3 in the Internet Appendix).

4.3. Communication, Style Investing, and Comovement

The literature on comovement shows that comovement can arise when investors follow defined investment styles, such as large- vs. small-cap and growth vs. value investing (see, for example, Vijh (1994), Barberis, Shleifer, and Wurgler (2005), and Greenwood (2008)). We therefore conduct tests to distinguish communication-driven and style-driven comovement.

In particular, we perform the following two groups of tests. First, we augment Models 2 and 4 with the Fama-French small-minus-big and high-minus-low factors. To be consistent with factor models, we replace the dependent variable *Ret* and the independent variables *MRR* and *MKTRet* by excess returns, that is, differences between returns and risk-free rates. We calculate the Fama-French factors and risk-free rates in China following Fama and French (1993). Second, we modify Models 2 and 4 by replacing the dependent variable *Ret* with the Fama-French 3-factor alpha and re-estimate the models. We obtain the Fama-French 3-factor alphas as residuals of 3-factor regressions of daily returns over the entire sample period.[15] We report the results of these tests in Table 9.

We find in Panel A of Table 9 that the coefficient of *Excess MRR* is positive and significant at the 1% level in column 1. The coefficient of the interaction of *Freq* with *Excess MRR* is positive and significant

[15] Our results are robust to using alphas estimated in one-year rolling windows prior to each month.

at the 1% level in column 2. In Panel B, we similarly observe that the corresponding coefficients are positive and significant at the 1% level.[16] These results are in line with our main findings.

[Insert Table 9 Here]

4.4. Industry and Macroeconomic Conditions

In our tests of Models 2 through 4, we include market return in the independent variables to control for the effects of market-wide factors on the comovement of stock returns. To address the possibility that stock prices may move together in response to industry-wide information, other changes in the financial markets, and various macroeconomic conditions, we consider several additional controls in this section.

First, to control for industry-level changes, we add the control variable *INDRet*, the daily average return of stocks in the same industry as the target stock, to the list of independent variables in Model 2, i.e., we estimate the following model:

$$Ret_{m,t} = \alpha_m + \beta_{1m} MRR_{m,t} + \beta_{2m} MKTRet_{m,t} + \beta_{3m} INDRet_{m,t} + \varepsilon_{m,t}. \quad \text{(Model 5)}$$

Second, we control for other aggregate factors of the financial markets in the model. We include several aggregate market-level variables: *IPO Activity*, to capture whether the market is "hot" or "cool"; *Log(Turnover)*, to proxy for the trading activity in the market; and *Term Spread*, to represent effects from the bond markets. In particular, we estimate the following model:

$$\begin{aligned} Ret_{m,t} &= \alpha_m + \beta_{1m} MRR_{m,t} + \beta_{2m} MKTRet_{m,t} + \beta_{3m} IPOActivity_{m,t} \\ &+ \beta_{4m} Log(TURNOVER_{m,t}) + \beta_{5m} TermSpread_{m,t} + \varepsilon_{m,t}. \end{aligned} \quad \text{(Model 6)}$$

Third, to account for macroeconomic conditions, we include *Inflation*, *GDP Growth*, and the *Economic Index* in the independent variables of the regressions and estimate the following model:

[16] In unreported results, we further control for industry returns in these regressions and find qualitatively similar results.

$$Ret_{m,t} = \alpha_m + \beta_{1m} MRR_{m,t} + \beta_{2m} MKTRet_{m,t} + \beta_{3m} Inflation_{m,t} + \beta_{4m} GDPGrowth_{m,t} + \beta_{5m} EconomicIndex_{m,t} + \varepsilon_{m,t}. \quad \text{(Model 7)}$$

Finally, we include all the control variables and estimate the following model:

$$Ret_m = \alpha_m + \beta_{1m} MRR_{m,t} + \beta_{2m} MKTRet_{m,t} + \beta_{3m} INDRet_{m,t} + \beta_{4m} IPOActivity_{m,t} + \beta_{5m} Log(TURNOVER)_{m,t} + \beta_{6m} Inflation_{m,t} + \beta_{7m} GDPGrowth_{m,t} + \beta_{8m} TermSpread_{m,t} + \beta_{9m} EconomicIndex_{m,t} + \varepsilon_{m,t}. \quad \text{(Model 8)}$$

Table 10 (Panel A) reports the results of the regressions in Models 5 through 8. In all specifications, the coefficients of *MRR* are positive and highly significant, suggesting that the comovement we find among stocks discussed together on sub-forums is not due to industry, market, or macroeconomic factors.

[Insert Table 10 Here]

We next include the frequency of discussion, *Freq*, and its interaction with *MRR* in Models 5 through 8, to examine whether the results in Section 3.4 continue to hold with the additional industry, market, and macroeconomic variables. Panel B of Table 10 reports the results of regressions for these models. The coefficients of the interaction term $Freq \times MRR$ continue to be positive and highly significant in all specifications. This evidence corroborates our finding that more intensive communication is associated with greater comovement in stock returns.

4.5. Absolute Returns

In our model, we consider a common component in fundamental values that is positively correlated to the values of both assets and show that communication about the common component generates positive comovement. In reality, there can be shocks that affect one stock positively and another negatively, e.g., when the two companies are competitors. Communication about such shocks may lead to negative comovement in stock prices. In such a case, the comovement of the *absolute values* of returns should still increase when investors communicate. Therefore, we redo our analyses replacing all return variables by their absolute values and report the results in Table 11. All our results hold with this change in methodology. In fact, the coefficient on the absolute value of *MRR* is larger in magnitude and has higher statistical significance than in our tests using simple returns; this indicates that positive and negative comovement may coexist. The fact that we find substantial positive comovement on average in our previous tests suggests that positive comovement is the dominant phenomenon.

[Insert Table 11 Here]

4.6. Controlling for Habitat-Based Comovement

In our model and previous tests, we consider online posts discussing related stocks in a stock sub-forum that can cause the target and related stock returns to comove. There are also many posts in the sub-forum that focus only on the target stock itself. These posts are potentially more concerned with idiosyncratic shocks to the target stock and thus tend to increase its idiosyncratic volatility rather than its comovement with other stocks (see also foonote 5). Furthermore, it is possible that online posts are correlated with investor attention and habitats in stocks and investor habitats can lead to stock comovement (see Barberis, Shleifer, and Wurgler (2005) for a habitat-based model of comovement).

To address these concerns, we adopt the following two approaches. First, we consider a relative measure of communication intensity, *Relative_Freq*, that equals the monthly frequency of posts on related stocks in a target forum, normalized by the total number of posts on the target stock in the same sub-forum. Using this measure allows us to filter out the effect of general discussions (and discussions that focus on idiosyncratic shocks) about target stocks and isolate the incremental effect of discussions about related stocks. This also helps to control for general trends in investor attention to stocks in their habitats. Second, we rank all target stocks monthly by the total number of discussions of themselves in their own sub-forums and form a portfolio of target stocks in the top decile for each month. We then define a return variable, *Top_Discussed_PortRet*, as the equally weighted daily returns of this top-decile portfolio. If investors are more likely to discuss about stocks in their habitats, then including this return variable in our regressions should help to control for habitat-based stock comovement.

We include these two variables in our analyses and report the results in Table 12. Column 1 shows that higher communication intensity measured by *Relative_Freq* leads to significantly higher comovement. Column 2 demonstrates that our main results are robust to controlling the habitat-induced return, *Top_Discussed_PortRet*, and all other control variables that we have employed. Interestingly, the coefficient on *Top_Discussed_PortRet* is also positively significant, suggesting that a habitat-based comovement may coexist with communication-driven comovement. Column 3 confirms that our results continue to hold when we incorporate both variables, *Relative_Freq* and *Top_Discussed_PortRet*.

[Insert Table 12 Here]

We also include the *Top_Discussed_PortRet* variable and industry returns (*INDRet*) in the tests of the forum outage in Section 3.6.[17] Table 13 reports the results. We continue to find that comovement to be lower in the outage month, but not in the placebo tests. In summary, the evidence suggests that communication generates comovement above and beyond habitat-based comovement.

[Insert Table 13 Here]

4.7. Controlling for Media Coverage

In this subsection, we address the alternative explanation that our findings may be driven by fundamental information by controlling for news. We use media coverage to capture public information about fundamental values of the target and related companies. We aggregate all news from major newspapers, television, and online media, including press comments and analyst suggestions, over the period of 2008 to 2012 from the RESSET database. In total, there are 183,969 pieces of news for all stocks on the Shanghai Stock Exchange during our time period. For any target stock and its top five most related stocks during a given month (that are used to construct the *MRR* variable), we define the variables *News_Target* as the number of daily news about the target firm, and *News_Related* as the sum of numbers of daily news about the five related stocks. We use the natural logarithm of (one plus) the number of news in our regressions.

We first include the news variables, *News_Target* and *News_Related*, in Model 2 and its extensions that incorporate other control variables we have considered. Panel A of Table 14 reports the results. We

[17] Note that we cannot include the macroeconomic variables in these tests since those variables are available only at the monthly frequency and thus are constants in the regressions that involve daily observations within each month.

find that the coefficients of *MRR* continue to be positive and significant at the 1% levels in all specifications.

We next add the interaction of communication intensity, *Freq*, with *MRR* to the regressions in Panel A. Further, we add the interaction of news about the related firms, *News_Related*, with *MRR* in these regressions to test if the comovement is driven by news, rather than posts about related stocks on the target sub-forum. We report the results in Panel B of Table 14. In all specifications, the coefficients on *Freq* × *MRR* continues to be positive and statistically significant. In summary, our main results are robust to controlling for media coverage.

[Insert Table 14 Here]

5. Conclusion

In this paper, we use a novel dataset of online forum discussions in China to study stock comovement and the communication among investors. We develop a model in which investors receive informative signals through communication before trading. The model predicts that communication can generate comovement in stock returns.

We find that there exists substantial comovement among the returns of the target stock and its related stocks discussed in the same sub-forum. Comovement is greater when related stocks are more frequently discussed. Further, the effect of frequent discussion on comovement is stronger for stocks with higher information asymmetry, i.e., small, illiquid stocks, and stocks covered by fewer analysts. These findings are consistent with our model's predictions. We use the exogenous variation in communication in a forum outage event to establish causality in our results. Finally, we find our results to be robust in a host of different specifications, including tests controlling for investor habitats and media coverage, tests in different subperiods, tests that control for additional industry, investment style, market, and macroeconomic factors, and tests using alternative measures of discussion intensity.

Taken together, our evidence sheds light on the impact of investor communication on the covariance of stock returns. Our findings can potentially assist investors in managing their portfolios through better understanding of the comovement in stock returns.

Appendix

Table A1
Top Five Related Stocks for Wuhan Iron and Steel (600005)

This table lists the top five related stocks of the stock Wuhan Iron and Steel (600005), i.e., stocks that are discussed in the largest number of posts on the stock sub-forum for Wuhan Iron and Steel. The number of posts and ranking are calculated on a monthly basis. For brevity, we list the composition of the portfolio for the most recent six months in our sample, from June 2008 to December 2008. In the last two rows, we calculate the monthly mean (median) number of clicks per post for the top five related stocks and then average them over the period June to December 2008.

Year and Month	Top Five Related Stocks ID	Firm Name	Number of Posts
June 2008	600439	Henan Rebecca Hair Products	6
	600177	Youngor Group	5
	000423	Shan Dong Dong- E E-Jiao	5
	600240	Beijing Huaye Real Estate	4
	000806	Beihai Yinhe Industry Investment	3
July 2008	600255	Anhui Xinke New Materials	9
	000709	Hebei Iron and Steel	6
	002146	Rongsheng Real Estate Development	5
	002253	Wisesoft	4
	580024	Baoshan Iron and Steel CWB1	4
August 2008	000629	Pangang Group Vanadium Titanium and Resources	10
	600019	Baoshan Iron and Steel	7
	002224	Sanlux	6
	000005	Shenzhen Fountain Corporation	5
	000819	Yueyang Xingchang Petrochemical	4
September 2008	002005	Elec-Tech International	3
	600145	Guizhou Guochuang Energy Holding (Group)	3
	600580	Wolong Electric Group	3
	600299	Blue Star New Chemical Materials	3
	000731	Sichuan Meifeng Chemical Industry	2

(Continued)

Year and Month	Top Five Related Stocks ID	Firm Name	Number of Posts
October 2008	000488	Shandong Chenming Paper Holdings	6
	000605	Bohai Water Industry	5
	000635	Ningxia Younglight Chemicals	4
	600080	Ginwa Enterprise Group Inc.	3
	000522	Guangzhou Baiyunshan Pharmaceutical	3
November 2008	600782	Xinyu Iron and Steel	15
	600151	Shanghai Aerospace Automobile Electromechanical	5
	000546	Jilin Guanghua Holding Group	4
	002265	Yunnan Xiyi Industry	4
	002060	Guangdong No. 2 Hydropower Engineering	3
December 2008	000511	Ingenious Ene-Carbon New Materials	7
	600782	Xinyu Iron and Steel	6
	002271	Beijing Oriental Yuhong Waterproof Technology	4
	002267	Shaanxi Provincial Natural Gas	3
	600637	Bestv New Media	3
		Number of Clicks Per Post (Mean)	803.7
		Number of Clicks Per Post (Median)	506.4

Table A2. Definitions of Variables

This table provides definitions of the variables used in our empirical analysis.

Variables	Definitions
<u>Return Variables</u>	
Ret	Daily stock return
MRR	Average daily return of the top five related stocks of each target stock
MKTRet	Market daily average weighted return
INDRet	Daily average weighted return of all stocks in the same industry as our target stock. We use the industry sector definitions provided by the China Securities Regulatory Commission
RANDRet	Mean return of the five randomly chosen stocks for each target stock
<u>Other Variables</u>	
Freq	Total number of times that related stocks are mentioned on the forum for a target stock in a month. In regressions, we use the standardized version of this variable with zero mean and standard deviation equal to 1.
Inflation	Monthly growth rate of CPI (consumer price index)
GDP Growth	Monthly growth rate of GDP (gross domestic product), interpolated from quarterly data
Term Spread	Difference between long-term (10-year) and short-term yield (3-month) on national debt (Welch and Goyal, 2008)
IPO Activity	Number of new firms that make an initial public offering in a month
Log(Turnover)	Log value of the value-weighted monthly turnover rate of all stocks in the market
Economic Index	Indicator for status of the economy, calculated by the National Bureau of Statistics of China

References

Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets* 5, 31-56.

Antweiler, W., and M. Z. Frank, 2004. Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance* 59, 1259-1294.

Avramov, D., T. Chordia, and G. Jostova, 2012. The world price of credit risk. *Review of Asset Pricing Studies* 25, 112-152.

Banerjee, S., 2011. Learning from prices and the dispersion in beliefs. *Review of Financial Studies* 24, 3025-3068.

Barber, Brad M., Terrance Odean, and Ning Zhu, 2008, Do retail trades move markets?, *Review of Financial Studies* 22, 151-186.

Barberis, N., A. Shleifer, and J. Wurgler, 2005. Comovement. *Journal of Financial Economics* 75, 283-317.

Calvo, G. A., 1999. Contagion in emerging markets: When wall street is a carrier. Working paper, Columbia University.

Chang, Y., H. Hong, L. Tiedens, N. Wang, and B. Zhao, 2015. Does diversity lead to diverse opinions? Evidence from languages and stock market. Working paper, Princeton University, Shanghai Advanced Institute for Finance, and Stanford University.

Chawla, N., Z. Da, J. Xu, and M. Ye, 2015. Catching fire: the diffusion of retail attention on twitter. Working paper, University of Notre Dame and University of Illinois at Urbana-Champaign.

Chordia, T., R. Roll, and A. Subrahmanyam, 2000. Commonality in liquidity. *Journal of Financial Economics* 56, 3-28.

Cohen, L., C. Malloy, and A. Frazzini, 2008. The small world of investing: Board connections and mutual fund returns. *Journal of Political Economy* 116, 951-979.

Cohen, L., C. Malloy, and A. Frazzini, 2010. Sell side school ties. *Journal of Finance* 65, 1409-1437.

Da, Z., J. Engelberg, and P. Gao, 2011, In search of attention, *Journal of Finance* 66, 1461-1499.

Das, S. R., and M. Y. Chen, 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science* 53, 1375-1388.

DeMarzo, P., D. Vayanos, and J. Zwiebel, 2003. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics* 118, 909-968.

Froot, K. A., and E. M. Dabora, 1999. How are stock prices affected by the location of trade? *Journal of Financial Economics* 53, 189-216.

Green, T. C., and B. H. Hwang, 2009. Price-based return comovement. *Journal of Financial Economics* 93, 37-50.

Greenwood, R., 2008. Excess comovement: Evidence from cross-sectional variation in Nikkei 225 weights. *Review of Financial Studies* 21, 1153-1186.

Grossman, S. and J. Stiglitz, 1980. On the impossibility of informationally efficient markets. *American Economic Review* 70, 393-408.

Hirshleifer, D., S. S. Lim, and S. H. Teoh, 2011. Limited investor attention and stock market misreactions to accounting information. *Review of Asset Pricing Studies* 1, 35-73.

Hong, H., W. Jiang, N. Wang, and B. Zhao, 2014. Trading for status. *Review of Financial Studies* forthcoming.

Hong, H., J. D. Kubik, and J. C. Stein, 2005. Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *Journal of Finance* 60, 2801-2824.

Kodres, L. E. and M. Pritsker, 2002. A rational expectations model of financial contagion. *Journal of Finance* 57, 769-799.

Kumar, A. and C. M. C. Lee, 2006. Retail investor sentiment and return comovements. *Journal of Finance* 61, 2451-2486.

Lee, C. M. C., A. Shleifer, and R. H. Thaler, 1991. Investor sentiment and the closed-end fund puzzle. *Journal of Finance* 46, 75-109.

Leung, A., A. Agarwal, P. Konana, and A. Kumar, 2013. Online search and return comovement. Working paper, University of Miami.

Liao, Li, Bibo Liu, and Hao Wang, 2011, Information discovery in share lockups: evidence from the split-share structure reform in China, *Financial Management* 40, 1001-1027.

Merton, R., 1986. A simple model of capital market equilibrium with incomplete information. *Journal of Finance* 42, 483-510.

Morck R., B. Yeung, and W. Yu, 2000. The information content of stock markets: Why do emerging markets have synchronous stock price movements? *Journal of Financial Economics* 58, 215-260.

Ng, Lilian, and Fei Wu, 2007, The trading behavior of institutions and individuals in Chinese equity markets, *Journal of Banking & Finance* 31, 2695-2710.

Peng, L., 2005, Learning with information capacity constraints. *Journal of Financial and Quantitative Analysis* 40, 307-329.

Peng, L., and W. Xiong, 2006. Investor attention, overconfidence, and category learning. *Journal of Financial Economics* 80, 563-602.

Pindyck, R. S. and J. J. Rotemberg, 1993. The comovement of stock prices. *Quarterly Journal of Economics* 108, 1073-1104.

Pirinsky, C. A. and Q. Wang, 2006. Does corporate headquarters location matter for stock returns? *Journal of Finance* 61, 1991-2015.

Veldkamp, L. L., 2006. Information markets and the comovement of asset prices. *Review of Economic Studies* 73, 823-845.

Vijh, A. M., 1994. S&P 500 trading strategies and stock betas. *Review of Financial Studies* 7, 215-251.

Wang, X. L., K. Shi, and H. X. Fan, 2006. Psychological mechanisms of investors in Chinese stock markets. *Journal of Economic Psychology* 27, 762-780.

Welch, I., and A. Goyal, 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21, 1455-1508.

Xu, C. K., 2000. The microstructure of the Chinese stock market. *China Economic Review* 11, 79-97.

Yang, B., 2013. Communication, excess comovement and factor structures. *Journal of Business Finance and Accounting* 40, 1304-1325.

Table 1
Summary Statistics

This table reports summary statistics and correlations of the variables used in our empirical analysis. Our sample period is from June 2008 to December 2012. All variables are defined in Table A2 in the Appendix. Panel A reports the summary statistics and Panels B and C report the correlations.

Panel A. Summary statistics

Variable	Mean	Std. Dev.	Median	25th Pct.	75th Pct.	Observations
Return Variables						
Ret (%)	0.023	2.928	0.000	-1.507	1.504	255,844
MRR (%)	0.131	2.529	0.272	-1.1944	1.639	255,844
MKTRet (%)	0.070	2.090	0.289	-0.8991	1.303	255,844
INDRet (%)	0.071	2.238	0.237	-1.0383	1.325	255,844
Other Variables						
Freq	11.597	10.586	8	6	14	255,844
IPO Activity	1.936	1.800	2	0	3	255,844
Log(Turnover)	2.872	0.647	2.890	2.372	3.306	255,844
Inflation (%)	0.210	0.501	0.200	-0.18	0.5	255,844
GDP Growth (%)	4.810	26.100	13.640	11.48	18.703	255,844
Term Spread (%)	1.824	0.651	1.956	1.366	2.423	255,844
Economic Index	1.019	0.021	1.017	1.004	1.031	255,844

Panel B. Correlations between the main return variables

	Ret	*MRR*	*MKTRet*
Ret	1	0.6258	0.6823
MRR	0.6258	1	0.8330
MKTRet	0.6823	0.8330	1

Panel C. Correlations between stock return and other control variables

	Ret	*IPO*	*Log(Turnover)*	*Inflation*	*GDP Growth*	*Term Spread*	*Economic Index*	*INDRet*
Ret	1	-0.008	0.055	0.012	-0.039	0.059	0.011	0.713
IPO Activity	-0.008	1	-0.460	0.390	-0.039	-0.379	0.118	-0.015
Log(Turnover)	0.055	-0.460	1	-0.280	0.027	0.753	0.084	0.081
Inflation	0.012	0.390	-0.280	1	-0.092	-0.121	0.208	0.011
GDP Growth	-0.039	-0.039	0.027	-0.092	1	-0.107	0.071	-0.052
Term Spread	0.059	-0.379	0.753	-0.121	-0.107	1	0.438	0.091
Economic Index	0.011	0.118	0.084	0.208	0.071	0.438	1	0.026
INDRet	0.713	-0.015	0.081	0.011	-0.052	0.091	0.026	1

Table 2

Communication and Comovement: Regressions

This table reports results of the stock-by-stock time series regressions in Models 1 and 2. We estimate the regressions in Models 1 and 2 separately for the 296 target stocks in our sample. All variables are defined in Table A2 in the Appendix. Panel A reports the average coefficients, distribution of *t*-values for the coefficients, and the distribution of insignificant and significant coefficients in these regressions. Panel B reports the average coefficients and overall *t*-statistics calculated using the methodology in Chordia et al. (2000).

Panel A. Stock-by-stock regression results of Model 1

Model 1		
Independent Variable	*MRR*	Constant
Average coefficient	0.717	-0.001
Mean *t*-value	23.846	-1.028
Minimum *t*-value	6.174	-3.504
Maximum *t*-value	38.869	2.057
Number of stocks	296	296

Model 1: Coefficient of *MRR*			
	Range of *p*-values	Num. of Stocks	%
	* [0.05,0.1)	0	0.0
Significant	** [0.01,0.05)	0	0.0
	*** (0,0.01)	296	100.0
Insignificant	$p > 0.1$	0	0.0
Total		296	100.0

Panel B. Stock-by-stock regression results of Model 2

Model 2			
Independent Variable	*MRR*	*MKTRet*	Constant
Average coefficient	0.213	0.740	-0.001
Mean *t*-value	4.118	12.305	-0.912
Minimum *t*-value	-1.779	-0.159	-3.781
Maximum *t*-value	13.032	22.931	2.341
Number of stocks	296	296	296

Model 2: Coefficient of *MRR*			
	Range of *p*-values	Number of Stocks	%
	* [0.05,0.1)	17	3.4
Significant	** [0.01,0.05)	32	9.8
	*** (0,0.01)	161	67.9
Insignificant	$p > 0.1$	86	18.9
Total		296	100.0

Panel C. Overall *t*-statistics based on Chordia et al. (2000)

	Model 1	Model 2
MRR	0.717	0.213
	(106.13)	(45.21)
MKTRet		0.740
		(95.67)
Constant	-0.00076	-0.00059
	(-4.12)	(-5.14)
No. Stocks	296	296

Table 3

Placebo Test for Comovement

This table reports the results of the placebo regressions in Model 3. We estimate the regressions in Model 3 separately for the 296 target stocks in our sample. All variables are defined in Table A2 in the Appendix. Panel A reports the average coefficients, distribution of *t*-values for the coefficients, and distribution of insignificant and significant coefficients in these regressions. Panel B reports the average coefficients and overall *t*-statistics calculated using the methodology in Chordia et al. (2000).

Panel A. Stock-by-stock regression results of Model 3

	Model 3		
Independent Variable	*RANDRet*	*MKTRet*	Constant
Average coefficient	0.007	0.952	0.000
Mean *t*-value	0.121	12.709	-0.722
Minimum *t*-value	-4.458	4.097	-3.798
Maximum *t*-value	13.696	21.835	2.444
Number of stocks	296	296	296

	Model 3: Coefficient of *RANDRet*		
	Range of *p*-values	Number of Stocks	%
	* [0.05,0.1)	25	8.5
Significant	** [0.01,0.05)	16	5.4
	*** (0,0.01)	21	7.1
Insignificant	$p > 0.1$	232	78.4
Total		296	100.0

Panel B. Overall *t*-statistics based on Chordia et al. (2000) for Model 3

	Model 3
RANDRet	0.007
	(1.49)
MKTRet	0.949
	(122.43)
Constant	-0.00046
	(-3.80)
No. Stocks	296

Table 4

Communication Intensity and Comovement

This table reports the results of the stock-by-stock time series regressions in Model 4. We estimate the regressions in Model 4 separately for the 296 target stocks in our sample. All variables are defined in Table A2 in the Appendix. Panel A reports average coefficients, distribution of *t*-values for the coefficients, and distribution of insignificant and significant coefficients in these regressions. Panel B reports the average coefficients and the overall *t*-statistics calculated using the methodology in Chordia et al. (2000).

Panel A. Stock-by-stock regression results of Model 4

	Model 4				
Independent Variable	*MRR*	*Freq* ×*MRR*	*Freq*	*MKTRet*	Constant
Average coefficient	0.224	0.077	0.001	0.744	-0.00027
Mean *t*-value	3.923	1.379	0.381	12.331	-0.501
Minimum *t*-value	-4.830	-3.664	-2.932	0.337	-2.895
Maximum *t*-value	6.137	6.755	2.712	22.740	2.399
Number of stocks	296	296	296	296	296

		Coefficients of *MRR*		Coefficients of *Freq*×*MRR*	
	Range of *p*-values	Number of Stocks	%	Number of Stocks	%
	* [0.05,0.1)	15	5.1	30	10.1
Significant	** [0.01,0.05)	30	10.1	25	8.4
	*** (0,0.01)	197	66.6	65	22.0
Insignificant	$p > 0.1$	54	18.2	176	59.5
Total		296	100.0	296	100.0

Panel B. Overall *t*-statistics based on Chordia et al. (2000) for Model 4

	Model 4
MRR	0.224
	(43.77)
Freq×MRR	0.077
	(17.49)
Freq	0.001
	(10.23)
MKTRet	0.744
	(95.96)
Constant	-0.00027
	(-2.22)
No. Stocks	296

Table 5
Information Asymmetry, Communication, and Comovement

This table reports the results of regressions for subsamples of stocks with different levels of information asymmetry. We use three proxy variables for information asymmetry: the lagged quarterly Amihud illiquidity measure, market capitalization, and the number of analysts who covered the stock in the previous year, to divide our sample into five quintiles. The quintile classification is redefined quarterly. We then estimate the regressions in Model 4 separately for each quintile and report the average coefficients. All other variables are defined in Table A2 in the Appendix. Panels A, B, and C report results using the Amihud illiquidity measure, market capitalization, and analyst coverage, respectively.

Panel A: Subsample analysis: quintiles by Amihud illiquidity

Mean Coefficients	Bottom Quintile	2^{nd} Quintile	3^{rd} Quintile	4^{th} Quintile	Top Quintile
Freq ×*MRR*	0.033	0.053	0.073	0.129	0.095
	(1.75)	(3.54)	(4.07)	(6.55)	(4.28)
Freq	0.001	0.001	0.001	0.001	0.002
	(3.01)	(2.69)	(5.32)	(4.68)	(7.83)
MRR	0.242	0.238	0.219	0.237	0.234
	(18.95)	(21.35)	(18.05)	(17.78)	(13.75)
MKTRet	0.630	0.752	0.796	0.823	0.762
	(38.27)	(56.90)	(73.46)	(73.07)	(63.11)
Constant	-0.000755	-0.000641	-0.000291	-0.000198	0.000638
	(-3.79)	(-4.97)	(-3.32)	(-2.31)	(6.18)
Diff. of Coeff. of *Freq* ×*MRR* (Top – Bottom Quintile)					0.061
t-statistic					(2.21)

Panel B: Subsample analysis: quintiles by market capitalization

Mean Coefficients	Bottom Quintile	2nd Quintile	3rd Quintile	4th Quintile	Top Quintile
Freq × *MRR*	0.125	0.063	0.075	0.056	0.028
	(8.55)	(2.72)	(4.61)	(4.05)	(2.28)
Freq	0.002	0.001	0.001	0.001	-0.001
	(7.64)	(6.04)	(3.97)	(3.41)	(-0.67)
MRR	0.247	0.208	0.236	0.228	0.225
	(21.62)	(11.78)	(19.45)	(20.00)	(19.31)
MKTRet	0.814	0.800	0.785	0.718	0.607
	(71.18)	(69.22)	(66.44)	(54.86)	(38.00)
Constant	0.001	0.000	0.000	-0.001	-0.001
	(6.06)	(0.73)	(-3.97)	(-4.38)	(-3.39)
Diff. of Coeff. of *Freq* × *MRR* (Top – Bottom Quintile)					-0.097
t-statistic					(-4.09)

Panel C: Subsample analysis: quintiles by analyst coverage

Mean Coefficients	Bottom Quintile	2nd Quintile	3rd Quintile	4th Quintile	Top Quintile
Freq × *MRR*	0.059	0.099	0.104	0.090	0.054
	(3.82)	(5.71)	(5.58)	(6.58)	(3.18)
Freq	0.001	0.001	0.003	0.001	0.000
	(5.38)	(3.73)	(10.49)	(3.41)	(1.60)
MRR	0.213	0.195	0.231	0.240	0.230
	(16.44)	(16.37)	(17.65)	(21.72)	(15.64)
MKTRet	0.831	0.846	0.763	0.703	0.641
	(72.65)	(79.64)	(68.24)	(56.06)	(41.43)
Constant	-0.000517	-0.000411	0.000514	-0.000315	-0.000049
	(-4.18)	(-4.29)	(4.92)	(-2.54)	(-0.25)
Diff. of Coeff. of *Freq* × *MRR* (Top – Bottom Quintile)					-0.005
t-statistic					(-0.17)

Table 6
Forum Outage, Communication, and Comovement

This table compares the relationship between communication and comovement in the month of a forum outage (June 2010) with the month before (May 2010) and the month after (July 2010). Panel A reports the results of the stock-by-stock regressions in Model 2 in the specific months. Panel B reports the results using the placebo sample of May, June, and July of 2009 and 2011. Average coefficients and overall *t*-statistics are calculated using the methodology in Chordia et al. (2000).

Panel A: Forum outage

	(1)	(2)	(3)
Month	May 2010	June 2010 (Outage month)	July 2010
MRR	0.220	0.125	0.178
	(3.70)	(1.97)	(4.44)
MKTRet	0.753	0.754	0.815
	(9.73)	(8.87)	(14.04)
Constant	-0.00059	-0.00061	-0.00040
	(-0.43)	(-0.54)	(-0.62)
No. of Stocks	275	275	275
Diff. of Coeff. of *MRR* with (2)	0.094		0.053
t-statistic	(8.19)		(5.69)

Panel B: Placebo tests

	(1)	(2)	(3)
Month	May 2009 and May 2011	June 2009 and June 2011	July 2009 and July 2011
MRR	0.272	0.281	0.279
	(12.55)	(12.68)	(11.93)
MKTRet	0.614	0.643	0.698
	(15.57)	(13.20)	(20.42)
Constant	-0.00067	-0.00017	-0.00085
	(-1.36)	(-0.32)	(-2.08)
No. of Stocks	275	275	275
Diff. of Coeff. of *MRR* with (2)	-0.009		-0.002
t-statistic	(-0.25)		(-0.05)

Table 7
Time Series Robustness Tests: Subperiod Analyses

This table reports the results of regressions in Models 1 through 4 conducted for two subperiods in our sample. Panels A and B report the results for the subperiod June 2008 to June 2010 and the subperiod July 2010 to December 2012, respectively. All variables are defined in Table A2 in the Appendix. Cross-sectional averages of coefficients from time series regressions are reported with *t*-statistics in parentheses. The *t*-statistics are calculated following Chordia et al. (2000).

Panel A: Subperiod analysis: June 2008 to June 2010

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
MRR	0.760	0.236		0.245
	(84.84)	(32.68)		(28.26)
RANDRet			0.011	
			(1.60)	
Freq × *MRR*				0.065
				(7.52)
Freq				0.001
				(3.96)
MKTRet		0.732	0.959	0.737
		(67.21)	(88.89)	(66.85)
Constant	-0.00103	-0.00083	-0.00066	-0.00048
	(-3.86)	(-4.91)	(-3.67)	(-2.57)
No. of Stocks	289	289	289	289

Panel B: Subperiod analysis: July 2010 to December 2012

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
MRR	0.633	0.179		0.193
	(57.59)	(27.24)		(25.02)
RANDRet			-0.011	
			(-1.89)	
Freq × *MRR*				0.077
				(9.16)
Freq				0.001
				(8.54)
MKTRet		0.733	0.926	0.736
		(60.02)	(76.94)	(60.61)
Constant	-0.0005	-0.00036	-0.00026	0.000113
	(-1.98)	(-2.37)	(-1.64)	(0.68)
No. of Stocks	294	294	294	294

Table 8
Comovement and Communication:
Alternative Measure of Communication

This table reports the results of the stock-by-stock regressions in Models 1, 2, and 4 using the total number of clicks on posts about related stocks to proxy for investor communication. *Clicks* is the total number of hits that the posts concerning the top five related stocks receive in the target stock's sub-forum in the given month. The *t*-statistics are calculated following Chordia et al. (2000).

	(1) Model 1	(2) Model 2	(3) Model 4
MRR	0.712	0.261	0.261
	(114.85)	(63.77)	(60.06)
MKTRet		0.692	0.696
		(98.07)	(98.65)
$Clicks \times MRR$			0.049
			(11.99)
Clicks			0.00096
			(10.16)
Constant	-0.00082	-0.00064	-0.00046
	(-4.67)	(-5.92)	(-4.14)
No. of Stocks	296	296	296

Table 9

Style Investing, Communication, and Comovement

This table reports the results of regressions that modify Models 2 and 4 by controlling for Fama-French factors or investing styles. In Panel A, the excess return of the target stock, *Excess Ret*, is used as the dependent variable and is analogous to the factor models. In Panel B, the Fama-French 3-factor alpha is used as the dependent variable. Fama-French factors and risk-free rates are calculated following Fama and French (1993). All other variables are defined in Table A2 in the Appendix. Cross-sectional averages of coefficients from time series regressions are reported with *t*-statistics in parentheses. The *t*-statistics are calculated following Chordia et al. (2000).

Panel A. Using Fama-French factor returns as controls

	Dependent Variable: *Excess Ret*	
	(1)	(2)
Excess MRR	0.173	0.184
	(14.48)	(44.77)
Freq × *Excess MRR*		0.074
		(9.53)
Freq		0.001
		(16.68)
Excess MKTRet	0.864	0.869
	(26.49)	(161.15)
SMB	-0.471	-0.470
	(-4.83)	(-37.80)
HML	0.168	0.169
	(1.35)	(10.77)
Constant	-0.000303	8.94E-07
	(-0.59)	(0.01)
No. of Stocks	296	296

Panel B. Using Fama-French 3-factor alphas as dependent variables

	Dependent Variable: *FF 3-factor alpha*	
	(1)	(2)
MRR	0.057	0.071
	(19.31)	(20.29)
Freq × *MRR*		0.071
		(16.88)
Freq		0.001
		(8.36)
Constant	0.00003	0.0003
	(0.400)	(3.88)
No. of Stocks	296	296

Table 10

Robustness Tests: Industry, Market, and Macroeconomic Controls

This table reports the results of the regressions of Models 5 through 8, which include various industry, market, and macroeconomic control variables. All variables are defined in Table A2 in the Appendix. Cross-sectional averages of coefficients from time series regressions are reported with *t*-statistics in parentheses. The *t*-statistics are calculated following Chordia et al. (2000). Panels A and B report robustness tests for Models 2 and 4, respectively.

Panel A. Robustness tests of Model 2

	(1)	(2)	(3)	(4)
MRR	0.156	0.213	0.213	0.156
	(42.20)	(45.22)	(45.23)	(42.16)
MKTRet	0.291	0.741	0.740	0.294
	(13.36)	(95.87)	(95.78)	(13.48)
INDRet	0.526			0.524
	(25.27)			(25.21)
IPO Activity		-8.7E-6		4.5E-5
		(-0.13)		(0.72)
Log(Turnover)		6.5E-5		1.1E-4
		(0.31)		(0.45)
Term Spread		-0.074		-0.012
		(-2.50)		(-0.38)
Inflation			0.014	-0.004
			(0.53)	(-0.20)
GDP Growth			-1.0E-4	-0.002
			(-0.18)	(-1.03)
Economic Index			-0.014	-0.007
			(-60.33)	(-8.72)
Constant	-0.000542	0.000656	0.014	0.007
	(-6.65)	(3.23)	(65.61)	(44.60)
No. of Stocks	296	296	296	296

Panel B. Robustness tests of Model 4

	(1)	(2)	(3)	(4)
Freq×MRR	0.064	0.077	0.076	0.064
	(15.83)	(17.34)	(17.26)	(15.73)
Freq	0.001	0.001	0.001	0.001
	(8.90)	(7.47)	(7.05)	(6.40)
MRR	0.164	0.224	0.224	0.164
	(39.43)	(43.75)	(43.66)	(39.32)
MKTRet	0.300	0.745	0.744	0.301
	(13.73)	(96.06)	(96.00)	(13.78)
INDRet	0.521			0.521
	(25.03)			(25.00)
IPO Activity		-0.002		0.002
		(-0.29)		(0.36)
Log(Turnover)		4.13E-5		5.05E-5
		(0.20)		(0.18)
Term Spread		-0.087		-0.0351
		(-2.90)		(-0.75)
Inflation			0.012	0.007
			(0.48)	(0.27)
GDP Growth			-2.6E-5	1.2E-4
			(-0.04)	(0.05)
Economic Index			-0.014	-0.005
			(-59.03)	(-3.90)
Constant	-0.00024	0.00141	0.01397	0.00526
	(-2.77)	(6.95)	(66.84)	(35.19)
No. of Stocks	296	296	296	296

Table 11
Communication and Comovement in Absolute Returns

This table reports the results of analyses where we replace all return variables by their absolute values. The variables are defined in Table A2 of the Appendix. The *t*-statistics are calculated using following Chordia et al. (2000).

	Dependent Variable: |*Ret*|		
	(1)	(2)	(3)
|*MRR*|	0.528	0.523	0.487
	(80.20)	(80.42)	(67.27)
|*MKTRet*|	-0.081	-0.078	-0.168
	(-9.99)	(-9.61)	(-5.57)
Freq×|*MRR*|		0.013	0.013
		(4.23)	(4.06)
Freq		0.003	0.002
		(24.33)	(20.70)
|*INDRet*|			0.080
			(2.80)
IPO Activity			-0.00051
			(-4.74)
Log(Turnover)			0.004
			(9.79)
Term Spread			-0.022
			(-0.43)
Inflation			0.179
			(4.66)
GDP Growth			0.00341
			(1.52)
Economic Index			-0.039
			(-28.00)
Constant	0.011	0.012	0.041
	(60.04)	(62.94)	(155.00)
No. of Stocks	296	296	296

Table 12
Relative Communication Intensity and Stock Return Comovement

This table reports the results of the stock-by-stock regressions in which we control for the overall discussion intensity in a target forum. *Relative_Freq* is the monthly frequency of discussions of related stocks in a target forum normalized by the total number of discussions of the target stock in the same sub-forum. *Top_Discussed_PortRet* is the equally weighted return of top-decile target stocks ranked by the total number of discussions of these stocks in their own sub-forums each month.

	(1)	(2)	(3)
Relative_Freq×MRR	5.528		5.132
	(3.77)		(3.67)
Relative_Freq	-0.033		0.021
	(-1.14)		(0.31)
MRR	0.225	0.149	0.150
	(46.56)	(37.72)	(42.13)
Freq×MRR		0.062	
		(15.50)	
Freq		0.001	
		(6.58)	
Top_Discussed_PortRet		0.225	0.227
		(23.25)	(23.78)
MKTRet	0.735	0.035	0.022
	(95.08)	(1.48)	(0.97)
INDRet		0.605	0.608
		(30.42)	(30.72)
IPO Activity		-1E-4	1E-4
		(-0.21)	(0.13)
Log(Turnover)		5E-6	-2E-5
		(0.02)	(-0.11)
Term Spread		-0.01442	0.007674
		(-0.33)	(0.23)
Inflation		0.0073	-0.00057
		(0.31)	(-0.03)
GDP Growth		0.00117	-0.0002
		(0.50)	(-0.12)
Economic Index		-0.00633	-0.00718
		(-5.22)	(-7.82)
Constant	-0.000615	0.006183	0.006599
	(-5.27)	(47.94)	(51.77)
No. of Stocks	296	296	296

Table 13
Forum Outage, Communication, and Comovement: Robustness Tests

This table reports the results of tests that include additional variables in those of Table 6. *Top_Discussed_PortRet* is the equally weighted return of top-decile target stocks ranked by the total number of discussions of these stocks in their own sub-forums each month. Other variables are defined in Table A2 of the Appendix. Panel A compares the results of the stock-by-stock regressions in the outage month (June 2010) to the months next to it. Panel B reports the results using the placebo sample of May, June, and July of 2009 and 2011.

Panel A: Forum outage

Month	(1) May 2010	(2) June 2010 (Outage month)	(3) July 2010
MRR	0.097	0.072	0.128
	(2.55)	(1.72)	(3.66)
MKTRet	-0.335	-0.189	-0.263
	(-1.20)	(-1.05)	(-0.82)
INDRet	0.920	0.784	0.964
	(4.13)	(5.17)	(3.61)
Top_Discussed_PortRet	0.373	0.340	0.178
	(3.43)	(4.11)	(2.05)
Constant	-0.00011	-0.00041	-0.00034
	(-0.16)	(-0.79)	(-0.81)
No. of Stocks	275	275	275
Diff. of Coeff. of MRR with (2)	0.025		0.057
t-statistic	(2.45)		(6.61)

Panel B: Placebo tests

Month	(1) May 2009 and May 2011	(2) June 2009 and June 2011	(3) July 2009 and July 2011
MRR	0.165	0.198	0.198
	(8.67)	(10.86)	(10.72)
MKTRet	-0.269	-0.026	0.039
	(-1.97)	(-0.29)	(0.51)
INDRet	0.930	0.640	0.631
	(7.28)	(7.91)	(11.02)
Top_Discussed_PortRet	0.115	0.184	0.131
	(1.97)	(4.79)	(2.90)
Constant	-0.00079	-0.00045	-0.00088
	(-2.38)	(-1.55)	(-3.14)
No. of Stocks	275	275	275
Diff. of Coeff. of MRR with (2)	0.033		0.000
t-statistic	(1.071)		(-0.012)

Table 14
Communication, Comovement, and News

This table reports the results of the stock-by-stock regressions that extend those in Models 1, 2, and 8, in which we control for news about target and related firms. *Log(News_Target)* and *Log(News_Related)* are the natural logarithm of the daily numbers of news articles about the target firm and the top five related firms, respectively. *Top_Discussed_PortRet* is the equally weighted return of top-decile target stocks ranked by the total number of discussions of these stocks in their own sub-forums each month. Other variables are defined in Table A2 of the Appendix. The *t*-statistics are calculated following Chordia et al. (2000).

Panel A. Communication and Comovement: Controlling for News

	(1)	(2)	(3)
MRR	0.212	0.155	0.139
	(45.00)	(41.92)	(40.32)
MKTRet	0.740	0.283	0.139
	(95.78)	(13.02)	(40.32)
Log(News_Target)	0.007	0.007	0.007
	(36.11)	(37.23)	(37.15)
Log(News_Related)	-5.36E-5	-4.54E-5	-5.64E-6
	(-1.20)	(-1.23)	(-0.16)
INDRet		0.534	0.615
		(25.81)	(31.17)
Top_Discussed_PortRet			0.227
			(23.78)
IPO Activity			1.24E-5
			(0.23)
Log(Turnover)			7.49E-5
			(0.36)
Term Spread			-0.00324
			(-0.11)
Inflation			7.01E-5
			(0.003)
GDP Growth			-5.38E-4
			(-0.32)
Economic Index			-0.00585
			(-7.76)
Constant	-0.001	-0.001	-0.001
	(-8.89)	(-11.66)	(-11.66)
No. of Stocks	294	294	294

Panel B. Communication Intensity and Comovement: Controlling for News

	(1)	(2)	(3)
MRR	0.221	0.161	0.147
	(40.52)	(36.20)	(34.57)
MKTRet	0.743	0.291	0.031
	(95.84)	(13.43)	(1.34)
Freq × *MRR*	0.074	0.061	0.059
	(16.54)	(14.94)	(14.56)
Log(News_Related) × *MRR*	0.00141	0.00095	0.00035
	(0.86)	(0.67)	(0.26)
Freq	0.00067	0.00066	0.00064
	(6.74)	(7.09)	(5.10)
Log(News_Related)	-8.02E-5	-7.14E-5	-3.70E-5
	(-1.83)	(-1.95)	(-1.04)
Log(News_Target)	0.00667	0.00646	0.00644
	(35.27)	(36.31)	(36.26)
Top_Discussed_PortRet			0.223
			(23.18)
INDRet		0.529	0.609
		(25.57)	(30.73)
IPO Activity			-6.16E-6
			(-0.11)
Log(Turnover)			4.61E-5
			(0.18)
Term Spread			-0.022
			(-0.48)
Inflation			0.00955
			(0.40)
GDP Growth			0.000974
			(0.38)
Economic Index			-0.00344
			(-2.67)
Constant	-0.000766	-0.000713	0.00280
	(-5.97)	(-7.71)	(21.75)
No. of Stocks	294	294	294

Internet Appendix to "Communication and Comovement: Evidence from Online Stock Forums"

Part I: Proof of Propositions

Proof of Proposition 1

Let $\tau_0 = \sigma_0^{-2}, \tau_{\text{蝌}} = \sigma^{-2}$ be the precisions of the prior belief and the noise term in the signals. The basic Bayesian updating formula implies that

$$\begin{aligned} \mu_1 &= E[x \mid z_1] = \frac{\tau_0}{\tau_0 + \tau_{\text{蝌}}} \mu_0 + \frac{\tau_{\text{ò}}}{\tau_0 + \tau} z_1, \\ \tau_1 &= \tau_0 + \tau_{\text{ò}} . \end{aligned} \tag{15}$$

From (15), it is easy to show by induction that (6) holds.

Proof of Proposition 2

For simplicity, we use the vector notations below, i.e.,

$\alpha = (\alpha_1, \alpha_2)', \ v = (v_1, v_2)', P = (P_1, P_2)',$ etc. Let I_N be the information set of investors after receiving all N signals, then

$$\begin{aligned} &E[-e^{-\gamma W} \mid I_N] = E[-e^{-\gamma(W_0 + \alpha'(v-P))} \mid I_N] \\ &= -\exp(-\gamma W_0 - \gamma\alpha'(E[v \mid I_N] - P) + \frac{1}{2}\gamma^2 \alpha' Cov(v-P, v-P \mid I_N)\alpha). \end{aligned} \tag{16}$$

Maximizing the above with respect to $\alpha,$ we obtain the investors' optimal portfolio:

$$\alpha^* = \gamma^{-1} Cov(v, v \mid I_N)^{-1}(E[v \mid I_N] - P). \tag{17}$$

The market clearing condition implies that $\alpha^* = S.$ Therefore, we obtain from (17) that

$$P = E[v \mid I_N] - \gamma Cov(v, v \mid I_N) S. \tag{18}$$

Note that

$$E[v \mid I_N] = E[x + y \mid I_N] = (\mu_N + \mu_{y_1}, \mu_N + \mu_{y_2})', \tag{19}$$

and

$$Cov(v,v \mid I_N) = E[v'v \mid I_N] = E\left[\begin{pmatrix} {v_1}^2 & v_1 v_2 \\ v_1 v_2 & {v_2}^2 \end{pmatrix} \mid I_N \right]$$
$$= \begin{pmatrix} \sigma_N^2 + \sigma_{y_1}^2 & \sigma_N^2 \\ \sigma_N^2 & \sigma_N^2 + \sigma_{y_2}^2 \end{pmatrix}. \tag{20}$$

Therefore, (8) follows from (18), (19), and (20).

Proof of Proposition 3

By (9) and (10), it suffices to show that

$$\sigma_0^2 > \left(\frac{N\sigma_\varepsilon^{-2}}{{\sigma_0}^{-2} + N\sigma_\varepsilon^{-2}} \right)^2 (\sigma_0^2 + \frac{1}{N}\sigma_\varepsilon^2) = \frac{N^2\sigma_0^4}{(\sigma^2 + N{\sigma_0}^2)^2}(\sigma_0^2 + \frac{1}{N}\sigma^2), \tag{21}$$

or

$$\sigma_0^2(\sigma_\varepsilon^2 + N{\sigma_0}^2)^2 > N^2\sigma_0^4(\sigma_0^2 + \frac{1}{N}\sigma^2). \tag{22}$$

Now the left hand side minus the right hand side of (22) is equal to:

$$\sigma_0^2(\sigma_\varepsilon^2 + N{\sigma_0}^2)^2 - N^2\sigma_0^4(\sigma_0^2 + \frac{1}{N}\sigma^2) = \sigma_0^2\sigma_\varepsilon^4 + N{\sigma_0}^4\sigma^2 > 0. \tag{23}$$

Therefore, (22) holds.

Proof of Proposition 4.

i) The statement follows directly from (9) and (12).

ii) Since the fraction $\frac{N\sigma_\varepsilon^{-2}}{{\sigma_0}^{-2} + N\sigma_\varepsilon^{-2}}$ increases in *N*, it follows that $\frac{\partial Cov}{\partial N} > 0.$

iii) Denoting $q = \sigma_0^2 / \sigma_\varepsilon^2$, using (12), the covariance can be rewritten as

$$Cov(P_1, P_2) = \frac{N^2q^2(q+1)}{(Nq+1)^2}\sigma_\varepsilon^2 - q\sigma^2 + \sigma_0^2. \tag{24}$$

Therefore, by calculation,

$$\frac{\partial Cov}{\partial N} = \frac{2Nq^2(q+1)}{(Nq+1)^3}\sigma_\varepsilon^2, \tag{25}$$

and

$$\frac{\partial^2 Cov}{\partial N \partial q} = \frac{3q^2 + q(2 - Nq)}{(Nq+1)^4} > 0, \qquad (26)$$

where the last equation follows from the fact that $2 - Nq = 2 - N\sigma_0^2 / \sigma_{\grave{o}}^2 > 0$. Equation (14) then follows from (26) and the chain rule.

Part II: Supplementary Tables

Table IA.1
Spring Festival, Communication, and Comovement

This table compares the relationship between communication and comovement in the month that contains the Spring Festival (the festival month; month *t*) as well as the month before (*t* – 1) and the month after (*t* + 1). Panel A reports the summary statistics of the number of posts about related stocks in target stock sub-forums in each month. Panel B reports the results of the stock-by-stock regressions in Model 2 in the various months. The average coefficients and overall *t*-statistics are calculated following Chordia et al. (2000).

Panel A: Monthly numbers of posts around the festival month

	Mean	Std. Dev.	Median	Obs.	Mean Diff. with (2)	*t*-stat.
(1) Month *t* – 1	11.757	10.812	8	1,047	1.187	(4.540)
(2) Festival Month (*t*)	10.120	8.532	7	1,047		
(3) Month *t* + 1	13.045	11.139	10	1,047	3.125	(12.215)

Panel B: Comovement around the festival month

	(1) Month *t* – 1	(2) Festival Month (*t*)	(3) Month *t* + 1
MRR	0.259	0.199	0.209
	(17.799)	(9.497)	(13.318)
MKTRet	0.658	0.709	0.762
	(45.259)	(33.785)	(48.527)
Constant	-0.000319	0.0000212	-0.000944
	(-2.356)	(0.046)	(-2.927)
Diff. of Coeff. of *MRR* with (2)	0.060		0.010
t-statistic	2.432		0.456

Table IA.2
Stock Return Comovement and Communication:
Removing Same-industry Related Stocks

This table reports the results of the stock-by-stock regressions in Models 1, 2, and 4 under the alternative specification in which we remove related stocks that are in the same industry as the target stock. The *t*-statistics are calculated following Chordia et al. (2000).

	(1) Model 1	(2) Model 2	(3) Model 4
MRR	0.582	0.069	0.075
	(75.80)	(18.09)	(17.54)
MKTRet		0.885	0.887
		(119.32)	(119.51)
Freq × *MRR*			0.045
			(10.37)
Freq			0.001
			(9.53)
Constant	-0.00053	-0.00051	-0.00018
	(-2.12)	(-4.30)	(-1.44)
No. of Stocks	296	296	296

Table IA.3
Stock Return Comovement and Lagged Communication

This table reports the results of the stock-by-stock regressions in Models 1, 2, and 4, where we define the related portfolio of a target stock using the discussions from the previous month. The *t*-statistics are calculated following Chordia et al. (2000).

	(1) Model 1	(2) Model 2	(3) Model 4
MRR	0.737	0.226	0.241
	(113.72)	(44.00)	(43.20)
MKTRet		0.721	0.721
		(89.61)	(89.58)
Freq × *MRR*			0.065
			(14.21)
Freq			-0.001
			(-5.66)
Constant	-0.000025	-0.00038	-0.00046
	(-0.15)	(-3.25)	(-3.82)
No. of Stocks	296	296	296

Table IA.4
Communication, Comovement, and News: Alternative Measure of News

This table reports the results of the stock-by-stock regressions that extend those in Models 1, 2, and 8, in which we control for news about target and related firms. *News_Target_Norm* and *News_Related_Norm* are the daily numbers of news articles about the target firm and the top five related firms, respectively, normalized by the total number of news about all firms in the Chinese stock market. *Top_Discussed_PortRet* is the equally weighted return of top-decile target stocks ranked by the total number of discussions of these stocks in their own sub-forums each month. Other variables are defined in Table A2 of the Appendix. The *t*-statistics are calculated following Chordia et al. (2000).

Panel A. Communication and Comovement: Controlling for News

	(1)	(2)	(3)
MRR	0.211	0.154	0.138
	(44.85)	(41.74)	(40.24)
MKTRet	0.741	0.284	0.021
	(95.94)	(13.09)	(0.89)
News_Target_Norm	0.421	0.410	0.407
	(28.05)	(28.69)	(28.67)
News_Related_Norm	0.011	0.010	0.011
	(4.48)	(4.62)	(5.03)
INDRet		0.534	0.614
		(25.75)	(31.13)
Top_Discussed_PortRet			0.227
			(23.69)
IPO Activity			1.89E-5
			(0.35)
Log(Turnover)			-7.42E-5
			(-0.35)
Term Spread			0.00997
			(0.34)
Inflation			-0.00537
			(-0.27)
GDP Growth			-0.00127
			(-0.77)
Economic Index			-0.0102
			(-13.70)
Constant	-0.001	-0.001	0.010
	(-8.92)	(-11.73)	(74.75)
No. of Stocks	294	294	294

Panel B. Communication Intensity and Comovement: Controlling for News

	(1)	(2)	(3)
MRR	0.220	0.160	0.145
	(52.49)	(38.33)	(36.64)
MKTRet	0.744	0.293	0.032
	(129.86)	(13.48)	(1.39)
Freq × *MRR*	0.072	0.059	0.057
	(17.58)	(14.52)	(14.19)
News_Related_Norm × *MRR*	-0.056	-0.059	-0.075
	(-0.28)	(-0.29)	(-0.38)
Freq	0.001	0.001	0.001
	(6.87)	(6.81)	(4.83)
News_Related_Norm	0.014	0.015	0.016
	(2.68)	(2.72)	(2.91)
News_Target_Norm	0.415	0.528	0.401
	(28.93)	(36.81)	(28.13)
INDRet		0.403	0.609
		(19.43)	(30.66)
Top_Discussed_PortRet			0.223
			(23.16)
IPO Activity			2.0E-6
			(0.04)
Log(Turnover)			-0.000133
			(-0.53)
Term Spread			-0.015
			(-0.32)
Inflation			0.00614
			(0.26)
GDP Growth			0.000705
			(0.28)
Economic Index			-0.00743
			(-5.88)
Constant	-0.0007513	-0.0007003	0.0073668
	(-8.48)	(-7.89)	(57.18)
No. of Stocks	294	294	294