Communication and Comovement:

Evidence from Online Stock Forums

January 2015

Abstract

We study the comovement of asset returns caused by communication among investors. We develop an equilibrium model of investor communication and trading and derive a number of testable predictions. We use a novel dataset on an active online stock forum in China to measure investor communication. For each stock, we consider its "related stocks" that are frequently discussed on the sub-forum dedicated to the given stock. We find that there is substantial excess comovement among the returns of a stock and its related stocks. Excess comovement is greater when related stocks are more frequently discussed. Furthermore, the effect of frequent communication on excess comovement is stronger for stocks with higher information asymmetry. Our findings are consistent with the model's predictions and highlight the potential distortive effects of investor communication on the covariance of stock returns.

Keywords: Excess Comovement; Asset Returns; Communication

JEL Classification: G12

1. Introduction

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

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

To motivate our empirical tests, we develop a Grossman-Stiglitz-type (1980) equilibrium model in which investors communicate before trading. The model shows that asset returns can exhibit excess comovement beyond those in fundamental values when investors communicate repeatedly. In the model, investors receive a sequence of signals via communication and update their beliefs before trading. Due to limited cognition, investors do not fully incorporate the consequences of repeated communication in their beliefs.[2] As a result, the model predicts that communication can generate excess comovement in stock prices.

[1] See for example Lee, Shleifer, and Vishny (1991), Pindyck and Rotemberg (1993), and Froot and Dabora (1999).
[2] This assumption is similar to the persuasion bias studied in DeMarzo, Vayanos, and Zwiebel (2003).

The model also predicts that excess comovement in asset returns is positively related to the frequency investors communicate before trading. Intuitively, more frequent communication leads to greater dependence of investors' beliefs on common signals and thus greater comovement. Further, the model predicts that the effect of communication on excess comovement is more pronounced when investors have less accurate beliefs, i.e., for stocks associated with greater information asymmetry. The intuition is that for stocks subject to greater information asymmetry, communication among investors exert a greater influence on investors' beliefs.

We test these predictions using a unique dataset from 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 been developing 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 institutional investors' importance has increased over time, individual investors still dominate in trading. At the end of 2007, individual investors hold 51.3% of the Chinese stock market by value, while institutional investors hold 42.3%, and the government 6.4%.[3] In the Chinese stock market, individual investors frequently exchange information and ideas on online forums.[4] Whereas such communication can help to propagate information and incorporate it into stock prices, it can

[3] 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

[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/).

also potentially lead to distortions in the market through mechanisms such as discussed above.

For any given stock, there is a sub-forum of the online forum devoted to discussion about it. We will 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 the model, we expect the stock returns of the stocks discussed together to have excess comovement. To test this hypothesis, for any target stock, we consider the most frequently discussed stocks (henceforth "most related stocks") on its sub-forum. We construct a *related portfolio* that consists of the five most related stocks to a target stock in each month. We then estimate regressions of target stock returns on the returns of their related portfolios to examine the correlation between these returns. We find that the correlation between a stock's and its related portfolio's returns is highly significant, even after controlling for market returns and industry returns, suggesting that there is excess comovement among these returns. The excess comovement is also economically significant, e.g., a 1% increase in the related portfolio return is associated with a 0.2% 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 the following falsification test. We first create for each target stock a "placebo" portfolio that consists of several placebo stocks randomly selected in the industries of the 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 the regressions to be insignificant, suggesting that the excess comovement we find 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 investors' posts about the top related stocks in the sub-forum for a target stock. We then include the 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 excess comovement between the return of target stock and its related stocks, consistent with the model's prediction.

We then examine the prediction that the effect of communication on return comovement is greater for stocks associated with greater asymmetric information. We use three proxy variables for 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. We find that for more illiquid, smaller, and less covered stocks, the frequency of forum discussion has a greater effect on excess comovement among stocks, consistent with the theoretical prediction.

We conduct a number of robustness tests. First, we carry out a time-series robustness test by conducting our tests separately for two equal sub-periods of our time period. Our results continue to hold for each of the two sub-periods. Second, we include a number of industry, market, and macroeconomic variables in our regressions and find our results to be robust. Third, we use the number of clicks the posts receive (instead of the number of posts) to proxy for the frequency of investor communication and define the portfolio of related stocks. We obtain similar results. Fourth, we control for Fama-French factors in our tests to address the possibility that comovement can arise from style investing and find that our main results to be qualitatively unchanged.

To alleviate the concern about endogeneity in our results, we employ an exogenous variation in the extent of investor communication caused by the most important holiday period in China, the Spring Festival Holidays. We show that communication in online forums in the month that contains the Spring Festival is significantly lower than the months immediately before and after. We reestimate our tests of comovement separately for the festival month and the neighboring months and find that the comovement in the festival month is the lowest, suggesting causality in our main results.

Our paper contributes to the literature that studies excess comovement in asset returns and its relation to investor behavior. To the best of our knowledge, our paper is the first to document excess 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) document 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 together. Green and Hwang (2009) document that after splits stocks comove more with other lower-priced stocks. We complement this literature by using a unique database on individual investors' communication to study the effects of communication and its frequency on excess comovement.

Our paper is also related to the literature on information transmission in social networks and its effects on economic agents' beliefs and behavior (e.g., Hong, Kubik, and Stein (2006), Cohen, Frazzini, and Malloy (2008, 2010)). Similar to this literature, we show that communication among investors can have substantial impact in the financial markets. Finally, 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 together on the same forum.

Finally, our model is related to a stream of theoretical literature that explains the excess comovement of stock prices from different angles. Calvo (1999) proposes a model in which the forced selling of emerging market securities may serve as negative signals for uniformed investors and result 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 experiencing idiosyncratic shocks. Peng and Xiong (2006) show that limited attention of investors can generate comovement in stock returns. Veldcamp (2006) uses the endogenous and costly production of information by investors to explain comovement in asset prices. Our model emphasizes the role of repeated communication in generating price comovement and has the advantage of being directly testable using observable data.

Yang (2013) develops a model that shows that communication in a social network can generate comovement and a concentrated factor structure in asset returns. Unlike his model, the model in this paper does not depend on assumptions about the structure of the social network. Therefore, our model is more parsimonious and the mechanism and intuition are more straightforward.

The rest of the paper is organized as follows. Section 2 develops the model and derives empirical predictions. Section 3 presents our data construction and empirical analysis. Section 4 provides the results of additional robustness tests. Section 5 concludes. All proofs are included in the Appendix.

2. The Model

In this section, we develop a Grossman-Stiglitz-type (1980) model to analyze the effects of communication on comovement in stock prices. The basic structure of our model is similar to that of Veldcamp (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 are idiosyncratic components. Note that without loss of generality, we assume that the coefficients on x to be 1 for both assets. The shocks x and y_i are independent and normally distributed. We assume that investors have identical prior beliefs that

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

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 the optimization of 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 the potential comovement of stock prices, we focus on the case that the signals contain information about the common component x in the asset values. 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. One can think of these signals as posts on a message board or 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 helping him to realize his profits earlier. Indeed, van Bommel (2003) shows that it can be optimal for informed investors with limited investment capacity to release private information with noises to the public.

We begin by assuming that the signals z_j are independent signals, i.e., ϵ_j, $j = 1, 2, ..., N$ are independent of each other, and later consider the possibility that these signals are not independent. By Bayesian updating, we have the following proposition about the beliefs of the agents.

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

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

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 are the asset prices at date zero. Therefore, investors choose their portfolios to solve the following optimization problem

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

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,[5]

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

Note that the covariance of the intrinsic asset values is

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

We have the following proposition that compares the covariances in fundamental values and asset prices.

[5] 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 CARA preferences and asset values with normal distributions, e.g., see Veldcamp (2006) and Banerjee (2011).

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 excess comovement in asset prices beyond those in the fundamental values.

Next, we assume that the signals z_j are not independent from each other and investors still regard them as independent.[6] The motivation is that there are unlikely to be many independent signals about firm values in a short time period. Investors, however, have incomplete information or paid limited attention about the sources of the signals (especially on online forums) and regard them as independent.[7] Our assumption is also similar to the persuasion bias of agents in DeMarzo, Vayanos, and Zwiebel (2003), i.e., people fail to account for the 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

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

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

---

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

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

Proposition 4. i) The covariance of asset prices $Cov(P_1, P_2)$ is always greater than that in the case where investors are fully rational. If $(N^2 - 2N)\sigma_0^2 - \sigma_\epsilon^2 > 0$, then the covariance of asset prices satisfy

$$Cov(P_1, P_2) > Cov(v_1, v_2). \tag{13}$$

ii) The following is always true:

$$\frac{\partial Cov(P_1, P_2)}{\partial N} > 0, \tag{14}$$

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

$$\frac{\partial^2 Cov(P_1, P_2)}{\partial N \partial \sigma_0} > 0. \tag{15}$$

Part (i) of Proposition 4 shows that repeated communication can give rise to excess comovement when investors have limited cognition or exhibit persuasion bias receiving repeated signals. By part (ii), the model predicts that the extent of the excess 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 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).[8] The intuition is that for stocks with greater information asymmetry,

[8] 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 online communications that we study in this paper.

communication among investors have a greater effect on their posterior beliefs and thus exert a larger influence on stock return comovement.

3. Empirical Analysis

3.1 Data and Variables

We collect our data of investor communication records by tracking all the messages posted on an online forum: the *East Money Stock Forum* (http://guba.eastmoney.com/). We choose this forum because it is the earliest stock forum in China and also one of the most active and influential forums. When we search the key words “stock forum” on the most popular search engines in China (Baidu or Google (Hong Kong)), the East Money Stock Forum always ranks as a top outcome. Moreover, the forum is fully compatible with the East Money trading software that 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. Therefore, the East Money Forum provides a relatively representative and comprehensive dataset of communications among investors that can be influential on stock trading and prices.

On the East Money Forum (henceforth the “forum”), there is a sub-forum for every stock on which investors can discuss and exchange information about the given stock. We will 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 (ID: 600005):

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

"Since FAW Automobile (000800) tumbles, 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 excess comovement among the returns of a target stock and its related stocks.

Due to limited availability of the forum data prior to 2008, we study the period from 2008 to 2012 in this paper. To ensure that there is sufficient discussion by investors on the forum, we also focus on the sub-forums devoted to the component stocks in 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 US, 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 them to attract investors' attention. Therefore, there are large numbers of messages on the sub-forums dedicated to these stocks. We use stock returns data from the Resset Database (http://www.resset.cn). During the period from 2008 to 2012, the composition of the SSE 180 index experienced several adjustments and a total of 296 stocks have been included in the index. Our sample of stock return data includes 255,844 stock-trading-day observations for these stocks.

We download investors' messages on the forum using a Perl program. Our program can retrieve from each 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 in the period 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 the messages posted on a target stock's sub-forum during the month. We record the frequency of a related stock being mentioned in these messages and rank the related stocks by such frequencies. We form 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 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.[9] Table A1 in the Appendix shows an example of top

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

five related stocks for one target stock in the SSE index during a six-month period in our sample. In this example, a top related stock is mentioned on the sub-forum from 2 to 15 times each month. 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.

Furthermore, we consider a number of (Chinese) market and macroeconomic factors in our analysis: *Inflation*, the monthly growth rate of 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 yield (Welch and Goyal, 2008); *IPO Activity*, the number of new firms that make initial public offering in a month; *Turnover*, the turnover rate of the stock market; and *Economic Index*, the indicator for economy status calculated by the National Bureau of Statistics of China. Panel A of Table 1 reports the summary statistics of the variables that we use in our empirical tests.

[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, consistent with our hypothesis in Section 2. 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 of sub-forums and their related stocks discussed on these sub-forums. As discussed in Section 2, our model predicts that investors' communication about a group of stocks can generate excess 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 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 can be generated by market-wide stock movement that drive 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 excess 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 23.8. This evidence suggests that there is strong comovement among forum-target stocks and their related stocks. Panel B of Table 2 shows that the average coefficient β_{1m} in Model 2 across all stocks is 0.213, positive and economically significant, with a mean *t*-value of 4.1. On average, a 1% increase in the *MRR* leads to an economically significant 0.21% increase in the daily target stock return. Furthermore, this coefficient is positive and significant at 1% levels for 161, or 68%, out of 296 regressions and insignificant (or negative) only in 86, or 19%, of the regressions. Therefore, after controlling for market-level changes, we find significant excess comovement among forum-target stocks and related stocks.

[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:

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

where $Var(\hat{\beta}_m) = \frac{(\hat{\eta}_m{}'\hat{\eta}_m)}{(T - k)}(X_m{}'X_m)^{-1}$, and $Cov(\hat{\beta}_m, \hat{\beta}_n) = \frac{(\hat{\eta}_m{}'\hat{\eta}_n)}{(T - k)}(X_m{}'X_m)^{-1}(X_m{}'X_n)(X_n{}'X_n)^{-1}$.

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 excess comovement between 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 excess comovement between 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 forum-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 a 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 0.121 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 with 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 1.49. In sum, the results of our placebo tests suggest that the comovement among target stocks and their related stocks are 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 rounds of communication between investors increase, investors update their beliefs about the stocks more, leading to greater comovement among stock returns. Therefore, we expect the excess comovement to be higher for stocks subject to more intense discussion. In this section, we use the frequency that stocks are discussed on sub-forums as a proxy for communication intensity and test this prediction.

We include the (logarithm) of 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:

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

The coefficient of the interaction term between *Log(Freq)* and *MRR* in Model 4 captures the marginal effects of more frequent discussion on the comovement between target stock and related-stock returns.

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 are both positive. The average *t*-statistic for the interaction term is 1.64 and marginally significant. While the coefficients of *MRR* are significant now only in 22% of the regressions, the coefficients of the interaction term are significant in 50% of the regressions. Panel B shows that the overall *t*-statistics of the mean coefficients in these regressions are 6.02 for *MRR* and 15.02 for the interaction term, both significant at the 1% level.

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

[Insert Table 4 Here]

3.5 Information Asymmetry, Communication, and Return Comovement

In this section we examine the relation among information asymmetry, communication, and excess 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 excess 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 and Mendelson, 1986; Amihud, 2002), calculated as follows:

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

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 effect 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 cover 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 of 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 *Log(Freq)×MRR* is increasing (from 0.031 in the bottom quintile to 0.087 in the top quintile) as the illiquidity of stock increases. The difference between the coefficients of *Log(Freq)×MRR* in the top and bottom quintiles is 0.056 and is statistically significant at the 1% level with a *t*-value of 2.66. (Note that since we include the interaction term in these regressions, the coefficients of *MRR* should not be interpreted as the overall excess comovement as before. Therefore, we focus on the interaction term and do not compare the coefficients of *MRR* across the subsamples.) Panel B shows that the coefficient of the interaction term decreases as stock market capitalization increases (from 0.102 in the bottom quintile to 0.04 in the top quintile; the difference is statistically significant with a *t*-value of -3.40). Panel C shows that the difference of the coefficients of *Log(Freq)×MRR* for stocks with the lowest analyst coverage and those with highest analyst coverage is negative but insignificant. Since stocks with higher illiquidity, smaller sizes, 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, consistent with the model's prediction.

4. Robustness Tests

4.1. Time-Series Robustness Tests

In this section, we perform a robustness test by conducting our main regressions in two equal sub-periods of our sample, i.e., the periods from January 2008 to June 2010 and from July 2010 to December 2012. We estimate the regressions of Models 1 to 4 separately for the two sub-periods and report the results in Table 6. For simplicity, we only report 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 sub-periods. In the placebo tests of Model 3, the mean return of a randomly chosen portfolio has a positive insignificant or a negative coefficient in the two sub-periods. In Model 4, the coefficients of the interaction term between *Log(Freq)* and *MRR* for the two sub-periods are both significant at the 1% level. Overall, the results of the above sub-period analyses are consistent with our findings in the previous sections.

[Insert Table 6 Here]

4.2.Industry and Macroeconomic Conditions

In our tests of Models 2 to 4 in the previous sections, we included the market return in the independent variables in order 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 the macroeconomic conditions, we consider various 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:

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

Next, we control for other aggregate factors in 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 the 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 *Economic Index* in the independent variables of the regressions and estimate the following model:

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

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

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

Table 7 (Panel A) reports the results of the regressions in Models 5 to 8. In all specifications, the coefficients of *MRR* are positive and highly significant, suggesting that the excess comovement that we found among stocks discussed together on forums is not due to industry, market, or macroeconomic factors.

[Insert Table 9 Here]

We next include the frequency of discussion, *Log(Freq)*, and its interaction with *MRR* in Models 5 to 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 7 reports the results of regressions for these models. The coefficients of the interaction term *Log(Freq)×MRR* continue to be positive and highly significant in all specifications. This evidence lends more support to the hypothesis that more intensive communication is associated with greater excess comovement in stock returns.

As an additional robustness check, 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 (see Table 8) to our previous findings.

[Insert Table 8 Here]

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 form the portfolio of related stocks each month. We then reestimate 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 9 reports the results, which are consistent with our previous results using the number of messages to proxy investor communication.

[Insert Table 9 Here]

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, i.e., differences of returns with 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 reestimate the models. We obtain the Fama-French 3-factor Alphas as residuals of 3-factor regressions of daily returns over the entire sample period.[10] We report the results of these tests in Table 10.

We find in Panel A of Table 10 that the coefficient of *Excess MRR* is positive and significant at the 1% level in Model 2 augmented with the Fama-French factors. The coefficient of the

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

interaction of *Log(Freq)* with *Excess MRR* is positive and significant at the 1% in the augmented Model 4. In Panel B, we observe similarly that the corresponding coefficients are positive and significant at the 1% levels.[11] These results are in line with our main findings.

[Insert Table 10 Here]

4.4. Lagged Communication and Comovement

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 the communication among investors could instead arise from the excess comovement among the target stock and its related stocks. The results in the cross-sectional tests in Section 3.5 can help to partially alleviate this potential concern about reverse causality.

In this section, 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. Since returns in the current month cannot affect the communication among investors in the previous month, this helps to further address the above concern.

We estimate the regressions in Models 1 to 4 with the above modification and report the results in Table 11. We find that the coefficients on *MRR* in Models 1 and 2 and the coefficients on *Log(Freq)×MRR* in Model 4 continue to be significantly positive. The coefficients on the returns of a randomly selected portfolio in Model 3 remain insignificant. This evidence reiterate

[11] In unreported results, we further control for industry returns and find qualitatively similar results.

our findings that there exists excess comovement among stocks discussed together on online stock forums and that such comovement is stronger when accompanied by more intensive communication among investors.

[Insert Table 11 Here]

4.5. Spring Festival, Communication, and Comovement

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

We consider the Spring Festival or the Chinese New Year, the most important holiday in China. The Spring Festival is the new year in the lunar calendar. Due to mismatches between the lunar and solar calendars, the Spring Festival can fall on different dates in January or February in different years. It is a Chinese tradition to celebrate the Spring Festival with their families for an extended time. In particular, the seven days starting with the Spring Festival are national holidays. Therefore, we expect the discussion in stock online forums during the month that contains the Spring Festival (henceforth, "festival month") to experience a substantial decline because people's attention is diverted elsewhere. Furthermore, it is not clear that the Spring Festival directly affect comovement in stock returns apart from its indirect effect that arises from reduced communication.

We report in Panel A of Table 12 the summary statistics of the number of posts about the top related stocks in each target stock sub-forum during a festival month, the previous month, and the next month. Consistent with expectation, we find that investors post on average 8.5 messages about related stocks in a festival month, compared to 10.8 (11.1) in the previous (next) month, with the differences significant at the 1% levels.

[Insert Table 12 Here]

We estimate the regressions in Model 2 separately for the festival month and the months before and after, and report the results in Panel B of Table 12. The coefficient on the related portfolio return is higher in the previous month (0.259) than in a festival month (0.199), with the difference significant at the 5% level. The coefficient in the next month is also greater than that in the festival month, although the difference is insignificant. This evidence helps to alleviate endogeneity concerns about the relation between communication and comovement.

5. Conclusion

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

We find that there exists substantial excess comovement among the returns of a forum's target stock and its related stocks – stocks that are discussed on the same forum. Excess comovement is greater when related stocks are more frequently discussed. Furthermore, the effect of frequent discussion on excess 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 the Spring Festival month to establish causality in our results. Finally, we find our results to be robust in a host of different specifications, including tests in different sub-periods and tests that control for additional industry, investment style, market, and macroeconomic factors.

Taken together, our evidence sheds light on the effect of investors' communication on their trading behavior and stock prices and can potentially assist investors in better managing their portfolios by understanding the comovement in stock returns.

Appendix

Proof of Proposition 1.

Let $\tau_0 = \sigma_0^{-2}, \tau_\epsilon = \sigma_\epsilon^{-2}$ be the precisions of the prior belief and the noise term in the singals.

The basic Bayesian updating formula implies that

$$\mu_1 = E[x \mid z_1] = \frac{\tau_0}{\tau_0 + \tau_\epsilon}\mu_0 + \frac{\tau_\epsilon}{\tau_0 + \tau_\epsilon} z_1, \qquad (16)$$
$$\tau_1 = \tau_0 + \tau_\epsilon .$$

From (16), 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. Since

$$E[-e^{-\gamma W} \mid I_N] = E[-e^{-\gamma(W_0 + \alpha'(v-P))} \mid I_N]$$
$$= -\exp(-\gamma W_0 - \gamma\alpha'(E[v \mid I_N] - P) + \frac{1}{2}\gamma^2\alpha' Cov(v-P, v-P \mid I_N)\alpha). \qquad (17)$$

Maximizing the above with respect to α , we obtain the investors' optimal portfolio

$$\alpha^* = \gamma^{-1} Cov(v, v \mid I_N)^{-1}(E[v \mid I_N] - P). \qquad (18)$$

The market clearing condition implies that $\alpha^* = S.$ Therefore, we obtain from (18) that

$$P = E[v \mid I_N] - \gamma Cov(v, v \mid I_N) S. \qquad (19)$$

Note that

$$E[v \mid I_N] = E[x + y \mid I_N] = (\mu_N + \mu_{y_1}, \mu_N + \mu_{y_2})', \qquad (20)$$

and

$$Cov(v,v\,|\,I_N) = E[v'v\,|\,I_N] = E\left[\begin{pmatrix} v_1{}^2 & v_1v_2 \\ v_1v_2 & v_2{}^2 \end{pmatrix} |\, I_N\right]$$
$$= \begin{pmatrix} \sigma_N^2 + \sigma_{y_1}^2 & \sigma_N^2 \\ \sigma_N^2 & \sigma_N^2 + \sigma_{y_2}^2 \end{pmatrix}. \tag{21}$$

Therefore, (8) follows from (19), (20), and (21).

Proof of Proposition 3.

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

$$\sigma_0^2 > \left(\frac{N\sigma_\epsilon^{-2}}{\sigma_0^{-2} + N\sigma_\epsilon^{-2}}\right)^2 (\sigma_0^2 + \frac{1}{N}\sigma_\epsilon^2) = \frac{N^2\sigma_0^4}{(\sigma_\epsilon^2 + N\sigma_0{}^2)^2}(\sigma_0^2 + \frac{1}{N}\sigma_\epsilon^2). \tag{22}$$

Or

$$\sigma_0^2(\sigma_\epsilon^2 + N\sigma_0{}^2)^2 > N^2\sigma_0^4(\sigma_0^2 + \frac{1}{N}\sigma_\epsilon^2). \tag{23}$$

Now the LHS minus the RHS of (23) is equal to

$$\sigma_0^2(\sigma_\epsilon^2 + N\sigma_0{}^2)^2 - N^2\sigma_0^4(\sigma_0^2 + \frac{1}{N}\sigma_\epsilon^2) = \sigma_0^2\sigma_\epsilon^4 + N\sigma_0{}^4\sigma_\epsilon^2 > 0. \tag{24}$$

Therefore, (23) holds.

Proof of Proposition 4.

i) The first statement follows directly from (9) and (12). By (12) and (10),

$$Cov(P_1,P_2) - Cov(v_1,v_2) = \left(\frac{N\sigma_\epsilon^{-2}}{\sigma_0^{-2} + N\sigma_\epsilon^{-2}}\right)^2 (\sigma_0^2 + \sigma_\epsilon^2) - \sigma_0^2$$
$$= \left(\frac{N\sigma_\epsilon^{-2}}{\sigma_0^{-2} + N\sigma_\epsilon^{-2}}\right)^2 (\sigma_0^2 + \sigma_\epsilon^2) - \sigma_0^2 = \frac{N^2\sigma_0^4(\sigma_0^2 + \sigma_\epsilon^2) - \sigma_0^2(N\sigma_0^2 + \sigma_\epsilon^2)^2}{(N\sigma_0^2 + \sigma_\epsilon^2)^2} \tag{25}$$
$$= \frac{\sigma_0^2\sigma_\epsilon^2((N^2 - 2N)\sigma_0^2 - \sigma_\epsilon^2)}{(N\sigma_0^2 + \sigma_\epsilon^2)^2} > 0.$$

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

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

$$Cov(P_1, P_2) = \frac{N^2 q^2 (q+1)}{\left(Nq+1\right)^2}\sigma_\epsilon^2 - q\sigma_\epsilon^2 + \sigma_0^2. \tag{26}$$

Therefore, by calculation,

$$\frac{\partial Cov}{\partial N} = \frac{2Nq^2(q+1)}{(Nq+1)^3}\sigma_\epsilon^2, \tag{27}$$

and

$$\frac{\partial^2 Cov}{\partial N \partial q} = \frac{3q^2 + q(2-Nq)}{(Nq+1)^4} > 0. \tag{28}$$

where the last equation follows from the fact that $2 - Nq = 2 - N\sigma_0^2 / \sigma_\epsilon^2 > 0$. (15) then follows from (28) and the chain rule.

Table A1
The Top Five Related Stocks of 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 most 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.

Year and Month	Top 5 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 & Resources	10
	600019	Baoshan Iron & Steel	7
	002224	Sanlux	6
	000005	Shenzhen Fountain Corporation	5
	000819	Yueyang Xingchang Petro-Chemical	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
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

Year and Month	Top 5 Related Stocks ID	Firm Name	Number of Posts
November, 2008	600782	Xinyu Iron & 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 & Steel	6
	002271	Beijing Oriental Yuhong Waterproof Technology	4
	002267	Shaanxi Provincial Natural Gas	3
	600637	Bestv New Media	3

Table A2. Definition of Variables

This table provides detailed descriptions of the variables that we use in our empirical analysis.

Variables	Definition
Return Variables	
Ret	The daily stock return
MRR	The average daily return of the top 5 related stocks of each target stock
MKTRet	The market daily average weighted return
INDRet	The daily average weighted return of all stocks of the same industry as our target stock. We use the industry sector definitions b China Securities Regulatory Commission (CSRC)
RANDRet	The mean return of the 5 randomly chosen stocks for each target stock
Other Variables	
Freq	The total number of the times that related stocks are mentioned on the forum for a target stock in a month
Inflation	The monthly growth rate of CPI (Consumer Price Index)
GDP Growth	The monthly growth rate of GDP (Gross Domestic Product), interpolated from quarterly data.
Term Spread	The difference between the long term yield (10-year) and the short term yield (3-month) on National debt (Welch and Goyal, 2008)
IPO Activity	The number of new firms that make Initial Public Offering in a month
Log(Turnover)	The log value of the value-weighted monthly turnover rate of all stocks in the market
Economic Index	The indicator for economy status calculated by the National Bureau of Statistics of China

References

Avramov, D., Chordia, T., and Jostova, G., 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.

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

van Bommel, J., 2003. Rumors. *Journal of Finance* 58, 1499-1519.

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

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

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

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

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

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

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

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

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

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

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

Hong, H., Kubik, J. D., and Stein, J. C., 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 Pritsker, M., 2002. A rational expectations model of financial contagion. *Journal of Finance* 57, 769-799.

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

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

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

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

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

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

Pirinsky, C. A. and Wang, Q., 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.

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

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

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

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 the summary statistics and correlations of the variables that we use in our empirical analysis. Our sample period is from January 2008 to November 2012. All of the 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 the 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 the 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	Num. 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 the 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 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	Num. 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 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 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*	Log(*Freq*) ×*MRR*	Log(*Freq*)	*MKTRet*	Constant
Average coefficient	0.067	0.069	0.000	0.734	-0.001
Mean *t*-value	0.575	1.637	0.102	12.120	-0.383
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 Log(*Freq*)×*MRR*	
	Range of *p*-values	Num. of Stocks	%	Num. of Stocks	%
	* [0.05,0.1)	20	6.8	23	7.8
Significant	** [0.01,0.05)	21	7.1	43	14.5
	*** (0,0.01)	25	8.4	81	27.4
Insignificant	*p* > 0.1	230	77.7	149	50.3
Total		296	100.0	296	100.0

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

	Model 4
MRR	0.067
	(6.02)
Log(*Freq*)×*MRR*	0.069
	(15.02)
Log(*Freq*)	0.000
	(2.32)
MKTRet	0.734
	(94.50)
Constant	-0.0010
	(-5.11)
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 and market capitalization, and the number of analysts who cover 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 the 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
Log(Freq)×MRR	0.0310	0.067	0.067	0.062	0.087
	(2.282)	(5.587)	(5.651)	(4.434)	(7.419)
Log(Freq)	0.000156	0.0000681	0.000418	0.000583	0.000854
	(1.386)	(0.974)	(6.157)	(7.832)	(9.351)
MRR	0.202	0.086	0.046	0.087	0.017
	(5.673)	(3.070)	(1.716)	(2.917)	(0.689)
MKTRet	0.622922	0.763	0.799	0.782	0.761
	(36.381)	(60.054)	(69.568)	(64.270)	(61.003)
Constant	-0.00131	-0.001	-0.00161	-0.00175	-0.00166
	(-4.696)	(-7.671)	(-17.079)	(-17.879)	(-13.854)
Diff. of Coeff. of *Log(Freq)×MRR* (Top – Bottom Quintile)					0.056
t-statistic					(2.66)

Panel B: Subsample Analysis: Quintiles by Market Capitalization

Mean Coefficients	Bottom Quintile	2nd Quintile	3rd Quintile	4th Quintile	Top Quintile
Log(Freq)×MRR	0.102	0.060	0.059	0.058	0.040
	(10.240)	(5.185)	(5.036)	(4.709)	(3.507)
Log(Freq)	0.00104	0.000845	0.0000149	0.000288	-0.00051
	(11.787)	(11.199)	(0.204)	(3.621)	(-4.085)
MRR	0.009	0.075	0.087	0.093	0.134
	(0.432)	(3.066)	(3.196)	(3.069)	(4.743)
MKTRet	0.802	0.790	0.783	0.710	0.604
	(69.712)	(68.212)	(65.632)	(54.067)	(37.624)
Constant	-0.00216	-0.00228	-0.00058	-0.00161	0.000472
	(-16.937)	(-22.598)	(-5.504)	(-10.644)	(1.523)
Diff. of Coeff. of *Log(Freq)×MRR* (Top – Bottom Quintile)					-0.062
t-statistic					(-3.40)

Panel C: Subsample Analysis: Quintiles by Analyst Coverage

Mean Coefficients	Bottom Quintile	2nd Quintile	3rd Quintile	4th Quintile	Top Quintile
Log(Freq)×MRR	0.070	0.055	0.061	0.060	0.063
	(7.89)	(5.33)	(5.02)	(5.71)	(4.59)
Log(Freq)	0.001	0.000	0.001	0.000	-0.001
	(7.28)	(6.19)	(11.92)	(0.34)	(-4.68)
MRR	0.068	0.059	0.071	0.081	0.081
	(3.64)	(2.52)	(2.77)	(3.20)	(2.52)
MKTRet	0.820	0.837	0.758	0.696	0.636
	(71.55)	(78.41)	(67.63)	(55.27)	(40.93)
Constant	-0.002	-0.002	-0.003	0.000	0.001
	(-13.90)	(-15.42)	(-23.04)	(-2.57)	(3.57)
Diff. of Coeff. of *Log(Freq)×MRR* (Top – Bottom Quintile)					-0.007
t-statistic					(-0.321)

Table 6
Time-Series Robustness Tests: Sub-period Analyses

This table reports the results of regressions in Models 1 to 4 conducted for two sub-periods in our sample. Panels A and B report the results for the sub-period January 2008 to June 2010 and the sub-period 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: Sub-period analysis: January 2008 to June 2010

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
MRR	0.760	0.236		0.096
	(84.838)	(32.683)		(4.276)
RANDRet			0.010	
			(1.411)	
Log(Freq)× MRR				0.062
				(6.859)
Freq				0.000332
				(2.516)
MKTRet		0.732	0.960	0.724
		(67.206)	(88.866)	(59.506)
Constant	-0.00103	-0.00083	-0.00066	-0.00148
	(-3.859)	(-4.909)	(-3.659)	(-4.938)
No. of Stocks	289	289	289	289

Panel B: Sub-period analysis: July 2010 to December 2012

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
MRR	0.633	0.179		0.067
	(57.590)	(27.237)		(4.859)
RANDRet			-0.013	
			(-2.175)	
Log(Freq)× MRR				0.059
				(9.277)
Freq				0.00032
				(2.460)
MKTRet		0.733	0.927	0.724
		(60.021)	(76.889)	(59.505)
Constant	-0.0005	-0.00036	-0.00026	-0.00097
	(-1.978)	(-2.374)	(-1.638)	(-3.659)
No. of Stocks	294	294	294	294

Table 7
Robustness Tests: Industry, Market, and Macroeconomic Controls

This table reports the results of regressions of Models 5 to 8 that 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 the 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		-0.0000087		0.0000448
		(-0.13)		(0.72)
Log(Turnover)		0.000065		0.000108
		(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			-0.0001003	-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)
Log(Freq)×MRR	0.059	0.069	0.069	0.059
	(15.68)	(15.05)	(14.91)	(15.62)
Log(Freq)	0.00027	0.00030	0.00035	0.00040
	(4.11)	(2.04)	(2.64)	(2.69)
MRR	0.031	0.067	0.068	0.030
	(3.37)	(5.95)	(6.04)	(3.34)
MKTRet	0.291	0.736	0.734	0.293
	(13.30)	(94.61)	(94.59)	(13.41)
INDRet	0.522			0.521
	(25.04)			(25.00)
IPO Activity		-9.55E-06		0.000047
		(-0.14)		(0.74)
Log(Turnover)		-0.00015		-0.00024
		(-0.61)		(-0.90)
Term Spread		-0.072		-0.014
		(-2.37)		(-0.38)
Inflation			0.017	0.004
			(0.65)	(0.18)
GDP Growth			0.00057	0.00046
			(0.94)	(0.24)
Economic Index			-0.01	-0.01
			(-41.45)	(-4.96)
Constant	-0.001	0.001	0.013	0.005
	(-7.95)	(3.40)	(61.55)	(35.53)
No. of Stocks	296	296	296	296

Table 8
Stock Return Comovement and Communication:
Removing Same-Industry Related Stocks

This table reports the results of 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 average coefficients and the overall *t*-statistics calculated using the methodology in Chordia et al. (2000) of the stock-by-stock regressions are reported.

	(1) Model 1	(2) Model 2	(3) Model 4
MRR	0.582	0.069	-0.011
	(75.802)	(18.093)	(-0.975)
MKTRet		0.885	0.884
		(119.315)	(192.644)
Log(Freq)× *MRR*			0.037
			(4.918)
Freq			0.000339
			(4.045)
Constant	-0.00053	-0.00051	-0.00117
	(-2.124)	(-4.302)	(-6.369)
No. of Stocks	296	296	296

Table 9
Comovement and Communication:
Using Numbers of Clicks to Proxy for Communication

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

	(1) Model 1	(2) Model 2	(3) Model 4
MRR	0.712	0.261	-0.176
	(114.851)	(63.768)	(-7.307)
MKTRet		0.692	0.696
		(98.068)	(98.864)
Log(Clicks)× *MRR*			0.052
			(18.376)
Clicks			0.0003185
			(12.956)
Constant	-0.0008227	-0.000643	-0.003314
	(-4.665)	(-5.917)	(-17.221)
No. of Stocks	296	296	296

Table 10
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, analogous to the factor models. In Panel B, the Fama-French 3-factor alpha is used as the dependent variable. Fama-French factors and the 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	0.041
	(14.479)	(4.646)
Log(Freq) × *Excess MRR*		0.063
		(16.964)
Log(Freq)		0.0003492
		(6.591)
Excess MKTRet	0.864	0.859
	(26.490)	(153.839)
SMB	-0.471	-0.472
	(-4.826)	(-36.851)
HML	0.168	0.165
	(1.352)	(10.199)
Constant	-0.000303	-0.000987
	(-0.588)	(-9.125)
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	-0.043
	(19.311)	(-5.201)
Log(Freq) × *MRR*		0.047
		(13.031)
Log(Freq)		0.0003713
		(5.840)
Constant	0.0000303	-0.0006975
	(0.400)	(-5.648)
No. of Stocks	296	296

Table 11
Stock Return Comovement and Lagged Communication

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

	(1)	(2)	(3)
MRR	0.737	0.226	0.094
	(113.719)	(44.002)	(7.983)
MKTRet		0.721	0.715
		(89.606)	(88.462)
Log(Freq) × *MRR*			0.062
			(12.737)
Log(Freq)			-0.00046
			(-5.382)
Constant	-0.000025	-0.00038	0.000634
	(-0.151)	(-3.247)	(3.378)
No. of Stocks	296	296	296

Table 12

Spring Festival, Communication, and Comovement

This table compares the relation between communication and comovement in the month that contains Spring Festival (the festival month; month *t*) and the months before (*t* – 1) and 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 regressions in Model 2 in the different months. The average coefficients and the overall *t*-statistics calculated using the methodology in Chordia et al. (2000) of the stock-by-stock regressions are reported.

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: Excess 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	-0.0003193	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