Identifying Herding effect in Chinese stock market by High-frequency data

Yunfei Hou
Institute of Complexity Science
and Big Data Technology
Guangxi University, Guangxi, China
Email: 739958730@qq.com

Jianbo Gao
Institute of Complexity Science
and Big Data Technology
Guangxi University, Guangxi, China
Email: jbgao.pmb@gmail.com

Fangli Fan
Institute of Complexity Science
and Big Data Technology
Guangxi University, Guangxi, China
Email: 514744840@qq.com

Feiyan Liu
Institute of Complexity Science
and Big Data Technology
Guangxi University, Guangxi, China
Email: liufeiyan1984@gmail.com

Changqing Song
State Key Laboratory of Earth Surface Processes and Resource Ecology
Center for Geodata and Analysis, Faculty of Geographical Science
Beijing Normal University, Beijing, China
Email: songcq@bnu.edu.cn

***Abstract*—Herding behavior is thought to often occur during market frenzy, stock crashes, financial crises, as well as strong bull markets. The issue has been gaining increasing attention in recent years, in the hope that timely detection of herding behavior can be used to implement effective means to mitigate them, thus to make the market more rational. So far, herding behavior has been mainly studied using low-frequency data with methods such as LSV, PCM, CH, CKK, and HS. Such studies can only report whether herding behavior exists in a long time span, such as a few months to even a few years, and thus essentially renders all those studies irrelevant to the design of any policies for curbing herding behavior. To achieve the latter goal, it is important to realize that herding behavior is a dynamic process that may only last for a short time span, such as a few minutes. This dictates that to timely detect the herding behavior in a stock market, high frequency data must be used. Guided by this rationale, we show that computation of mutual information and cross correlation coefficient from high frequency data can indeed effectively identify herding behavior from Chinese stock markets.**

I. Introduction

The terrifying Chinese stock market crash in 2015 has made regulators and scholars increasingly realized that the significant herding behavior during the period has greatly exacerbated market volatility. Compared with mature markets in developed countries, empirical evidence has been found that there are more herding behaviors in the emerging markets such as Chinese stock market. [9]. Since herding behavior greatly aggravates market irrationality, it is important to timely detect it in markets, especially in emerging unmature ones.

A number of methods have indeed be proposed to detect herding behavior. One of the earliest approaches is the LSV method developed by Lakonishok, Shleifer, and Cishny [2], which defines the herding behavior as the average trend of the fund managers to buy or sell certain stocks at the same time. The method is the first to measure the herding behavior on fund managers and thus has been widely used. Unfortunately, the LSV method suffers a number of drawbacks, such as the chosen stocks are not properly weighted by volumes. In order to improve the LSV method, Wermers [3] has proposed an improved method called PCM (Portfolio-change measure) to quantify mutual fund herding. However the PCM method requires a large amount of data so that it is difficult to apply in empirical research. A more widely used method is called CSSD (Cross-sectional standard deviation), proposed by Christie and Huang [4], which tests the degree of similarity in the behavior of the market participants by quantifying return dispersion. Being a pioneering approach, the method is also called CH, to honor its creators. However, it is not very sensitive, and essentially can only detect herding behavior in extreme market situations. To overcome the shortcoming of CSSD, Chang et al. [5] have proposed a variant called CSAD (Cross-sectional absolute deviation) to test the herding. There are lots of empirical researches with CSAD model. The CSAD model is one of the most mainstream approaches to quantitatively identify the herding effect. Demirer and Kutan [6] conclude that there are no herding effects in Chinese stock market with CSAD model. Tan et al. [7] point out there exist herding behaviors in Chinese stock market both in the rising and falling periods and the herding effects are more significant when the market in a period of high volatility. Paulo Lao, Harminder Singh [8] have studied the herding of Chinese and Indian stock markets. They suggest that herding behavior exists in both of these two emerging markets, relative speaking, a higher prevalence of herding behavior has been detected in Chinese stock market. In Chinese stock market, herding behavior is greater when the market is decreasing and the trading volume is high. Chiang et al. [9] have improve the CSAD model (CKK), taking into account the asymmetry in the rising and decreasing market. They get such a conclusion in Chinese stock market, which is there be herding effect whether the SH A stock market is up or down and there only exist herding when the SH B stock market is down. Rongbao Gu, and Kexue Jiang [10] take a step further than Chiang et al.

978-1-5386-2366-4/17/$31.00

They not only study the asymmetry of the herding effect in Shenzhen stock market, but also use the approach of moving the window to study the dynamic evolution of herding. They respectively select the length of three-year and four-year as the window to estimate the herding, concluding that there always exist herding behavior in Shenzhen stock market. Obviously, this conclusion is inaccurate because it is impossible that the herding effect always exist. Most likely, they choose a too long window size to estimate the herding. So, how long the window size should we select ?

Herding behavior is a dynamic process and it may only occur for a few tens of minutes or even shorter. The desired approach for studying herding behavior has to tell when it occurs. For this purpose, high frequency data has to be used and we need a more precise indicator to identify the herding effect.

II. Data and Methods

A. Data

We analyze both the low-frequency (daily) as well as high-frequency (minutely) composite index of the Shanghai and Shenzhen markets and individual stock price, from November 11, 2013 to August 8, 2016. The data include 943 companies in the Shanghai market and 1524 in the Shenzhen market. In China, the trading time of a trading day for both markets is from 9:30 to 11:30 in the morning and 13:00 to 15:00 in the afternoon from Monday to Friday. Thus there are 240 data points for each per day.

B. Methods

To infer herding behavior from high-frequency data, we will need the concepts of mutual information and cross correlation coefficient. Before we explain them, we will describe the CSAD model, which is based on low-frequency data, so that we can make comparison between our approach and the classic CSAD method.

1) CSAD model: The method was initially proposed by Chang et al. [5]. The basic idea is to infer herding behavior by computing cross-sectional absolute deviations. It consists of the following steps. Firstly, it computes returns of individual stocks,

$$R_t = \ln(P_t) - \ln(P_{t-1}) \tag{1}$$

Then it computes cross-sectional absolute deviations (CSAD),

$$CSAD_t = \frac{1}{N}\sum_{i=1}^{N} |R_{i,t} - R_{m,t}| \tag{2}$$

where N is the number of available securiries, $R_{i,t}$ is the individual stock return on firm i at time t, $R_{m,t}$ is the average return of the equal-weighted market portfolio at time t. Finally, it infers whether there exists herding behavior based on a regression analysis of CSAD. The basic equation for the regression analysis is

$$CSAD_t = \alpha + \beta_1|R_{m,t}| + \beta_2|R_{m,t}|^2 + \varepsilon_t \tag{3}$$

where $|R_{m,t}|^2$ is used to emphasize that it is the square of $|R_{m,t}|$, and β_1 and β_2 are two coefficients.

The rational asset pricing model predicts that the return dispersion linearly increases with market return. The CSAD method interprets a special type of deviation from this rational asset pricing model's prediction as caused by the herding behavior. This is specified by the nonlinear term in Eq. (3). When the coefficient of $|R_{m,t}|^2$, β_2, is significantly negative , it is considered that herding behavior has occurred.

It is often useful to consider the herding behavior in the up and down markets separately, as the two markets are quite asymmetric. This can be achieved by considering two equations,

$$CSAD_t^{(U)} = \alpha + \beta_1^{(U)}|R_{m,t}^{(U)}| + \beta_2^{(U)}|R_{m,t}^{(U)}|^2 + \varepsilon_t \tag{4}$$

$$CSAD_t^{(D)} = \alpha + \beta_1^{(D)}|R_{m,t}^{(D)}| + \beta_2^{(D)}|R_{m,t}^{(D)}|^2 + \varepsilon_t \tag{5}$$

where $|R_{m,t}^{(U)}|$ and $|R_{m,t}^{(D)}|$ are the absolute values of an equally-weighted market portfolio on day t when the market is up and down, respectively.

2) Mutual information: Before defining mutual information, it is helpful to first consider how information entropy is defined for individual stocks everyday. For this purpose, it is important to recall that the stock data analyzed here are high-frequency. Because of this, we can define stock return over a short period of time, which we choose to be 1 minute here. Then we have 239 stock return data for each stock everyday. From these data points we can estimate its discrete probabilities. Denote them by $\{P_i, i = 1, \cdots, n\}$, then information entropy is defined as

$$H(X) = \sum_{i}^{n} P_i \ln P_i \tag{6}$$

We can now define mutual information between the return of an individual stock and the market return. Denote them by X and Y, respectively. $H(X, Y)$ is the joint information entropy based on the joint probability for X and Y,

$$H(X, Y) = \sum_{(x,y)} P(x, y) \ln P(x, y) \tag{7}$$

Mutual information quantifies how much X and Y are mutually dependent on each other. It is given by

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \tag{8}$$

In terms of probabilities, it is given by

$$I(X; Y) = \sum_{x\epsilon X}\sum_{y\epsilon Y} P(x, y) \ln \frac{P(x, y)}{P(x)P(y)} \tag{9}$$

Mutual information provides an excellent means of quantifying the shared information between X and Y. Here, we employ it to quantify the degree of similarity between the return of individual stocks and the market return. Let $I(X_i; Y)$ denote the mutual information between stock X_i and the market return Y, we obtain the synthesized mutual information by averaging over all $I(X_i; Y)$, with the weights w_i equal to the value of firm i divided by the total value of the market,

$$I_{synthesized} = \sum_{i=1}^{N} I(X_i; Y) w_i \tag{10}$$

3) Cross correlation coefficient: Cross correlation coefficient can be used to measure the linear correlation between two random variables. It is defined as

$$R_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \tag{11}$$

where x and y are the values for the random variables X and Y respectively. Following the definition for synthesized mutual information, we can obtain the synthesized cross correlation coefficient as

$$R_{synthesized} = \sum_{i=1}^{N} R(X_i; Y) w_i \tag{12}$$

Since herding behavior amounts to significant increase in the correlations between all available securities' returns and the market return, sharp increases in these synthesized mutual information and correlation coefficient measures can be used to effectively indicate emergence of herding behavior.

III. Empirical results

A. A dilemma with the CSAD model

To motivate the use of high-frequency data for detecting herding behavior, we first present a dilemma with the CSAD model. By choosing different window sizes, the CSAD method may change a no-herding to a herding behavior. Such a dilemma cannot be solved within the framework of CSAD, since there is no theory to determine how the window size should be chosen.

TABLE I
Descriptive Statistics of CSAD

CSAD	Mean	Std	Skewness	Kurtosis	P value
SH	0.0173	0.0074	1.8198	7.0318	0.00
SZ	0.0194	0.0076	1.7258	6.2826	0.00

To apply the CSAD method, we have used the low-frequency daily stock data. The relevant descriptive statistics are summarized in table 1. The P values of the CSAD are less than 0.5, implying that the CSADs in Shanghai and Shenzhen market largely obey normal distribution, and hence the CSAD approach could be used. To better see the relationship between CSAD and the market returns, we have plotted them out in Fig. 1, where each point represents one day, and the total duration is two years, from the beginning of 2014 to the end of 2015. We then have examined whether herding behavior existed in the Shanghai market or not by considering the stock data in 2014 alone, 2015 alone, and both years' data combined. The basic results are summarized in the following equations:

$$CSAD_t^{(2014)} = 0.0128 + 0.05\left|R_{m,t}^{(2014)}\right| + 4.19^{(**)}\left|R_{m,t}^{(2014)}\right|^2 \tag{13}$$

$$CSAD_t^{(2015)} = 0.0148 + 0.39\left|R_{m,t}^{(2015)}\right| - 2.12^{(**)}\left|R_{m,t}^{(2015)}\right|^2 \tag{14}$$

$$CSAD_t^{(2yr)} = 0.1371 + 0.3965\left|R_{m,t}^{(2yr)}\right| - 1.84^{(**)}\left|R_{m,t}^{(2yr)}\right|^2 \tag{15}$$

where *, **, *** indicate that the parameters are significant at the confidence level of 10%, 5% and 1%, respectively.

Recall that the CSAD approach prescribes that when the coefficient for the term $\left|R_{m,t}\right|^2$ is significantly negative, then there existed herding behavior in the market during the time period examined. Now, from Eq. (13) to (15), we observe that this coefficient is positive in 2014, negative in 2015, and negative when the joint 2-year time interval from the beginning of 2014 to the end of 2015. Thus, we conclude that there was no herding behavior in 2014, but there existed herding behavior in 2015 and from the beginning of 2014 to the end of 2015. That there was no herding behavior in 2014 but there was when the year 2014 and 2015 were combined is clearly a dilemma.

There are other difficulties with CSAD. For ease of later discussions, we point out one more before ending this subsection. It is related to the asymmetry of herding behavior, that is, the down market is more likely to have herding behavior than the up market. This is often associated with the slow rise in up markets but sharp drop in down markets. One such observation was made by Paulo Lao, Harminder Singh [8] when studying the Chinese stock market. To illustrate the idea, we use the 2-year data from the beginning of 2014 to the end of 2015. The basic results are summarized by the following two equations.

$$CSAD_t^{(U)} = 0.0141 + 0.0845\left|R_{m,t}^{(U)}\right| + 4.1814\left|R_{m,t}^{(U)}\right|^2 \tag{16}$$

$$CSAD_t^{(D)} = 0.0127 + 0.65\left|R_{m,t}^{(D)}\right| - 5.23^{(***)}\left|R_{m,t}^{(D)}\right|^2 \tag{17}$$

Indeed, CSAD indicates that the herding behavior existed in the down market but not the up market. However, we will show with high-frequency data analysis that up market may also have herding behavior.

B. Detecting herding behavior using high-frequency data

In the last subsection, we pointed out a dilemma with the CSAD method. In fact, even without that kind of dilemma, CSAD is still of little use for helping quash herding behavior in a market, since it can only tell whether herding existed in a long window chosen for analysis, while herding usually only occurred in a very short period of time, such as a few minutes or a few tens of minutes in a trading day. To timely detect herding behavior, one has no choice but to use high-frequency data. However, even with high-frequency data, the window size chosen for analysis with CSAD still has to be large. This is mandated by the part of regression analysis in CSAD. To timely detect herding behavior, in this subsection, we compute the mutual information and cross correlation coefficient on everyday and explore whether their temporal variations may offer desired solutions.

We have computed mutual information and cross correlation coefficients between all the individual stocks and the market returns on every day. Since the number of mutual information and cross correlation coefficients, being the same as the number of listed companies, is large enough, we have also computed the probability density distribution (PDF) of mutual information and cross correlation coefficients on every day. Examples of these PDF curves are shown in Fig. 2, for the periods from November 11, 2013 to September 1, 2014 (as blue curves), and from June 1, 2015 to September 1, 2015 (as red curves). The PDFs for mutual information and cross correlation coefficients in the first period are on the left side of that for the second period, clearly indicating that correlations in the market were much stronger in the second turbulent period than the first fairly stationary quiet period.

We have also computed the synthesized mutual information and cross correlation coefficient on every day. Their temporal variations are shown in Fig. 3 as green and red curves, respectively. To ease comparison with the index data, we have also plotted the re-scaled index data in the figure. Clearly, the whole period encompassing the end of 2013 and 2014, synthesized mutual information and cross correlation coefficients are small compared to those in other periods. Therefore, we have to conclude that during this period, herding behavior did not exist in the Shanghai market, as there was nothing unusual that had happened in the market to cause herding behavior. However, the synthesized mutual information and cross correlation coefficients in year 2015 are generally large, signifying that this was a year full of "drama" in the market. While these straightforward calculations have enriched the indications from Equations (13) and (14), they have also got rid of the dilemma from the CSAD method, since now the herding behavior is associated with days, but not a long interval. In fact, as one can readily see, detection of herding behavior based on computation of mutual information and cross correlation coefficient with high-frequency data can be extended to time intervals much shorter than a day.

We have further examined special days with large synthesized mutual information and cross correlation coefficients. A few examples are highlighted in Fig. 3 as circles when $R_{synthesized}$ is larger than 0.5. In fact, on those few special days, the market returns were highly varied, with amplitude of variation close to or large than 10% (however, the daily returns for some of them were still small; which provides another imperative reason to detect herding behavior using high-frequency instead of low-frequency data). It is important to emphasize here that some of those special days were in the up market, therefore, herding behavior did exist in the up markets on days when the market gained huge returns. This is in stark contrast with the result from CSAD analysis, which stated based on Equation. (16) that herding behavior did not exist in the up market.

Finally, we have examined the relation between the synthesized mutual information and cross correlation coefficients.

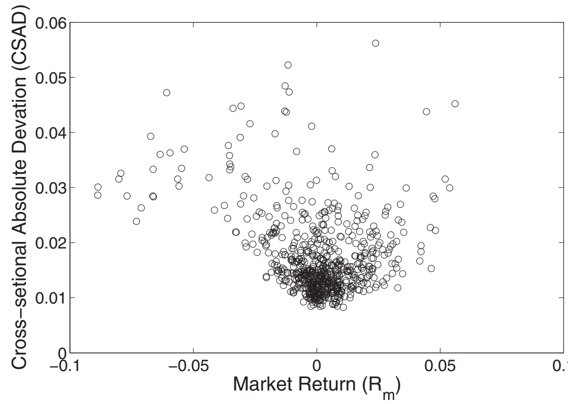

Fig. 1. The relationship between CSAD and market return of SH.

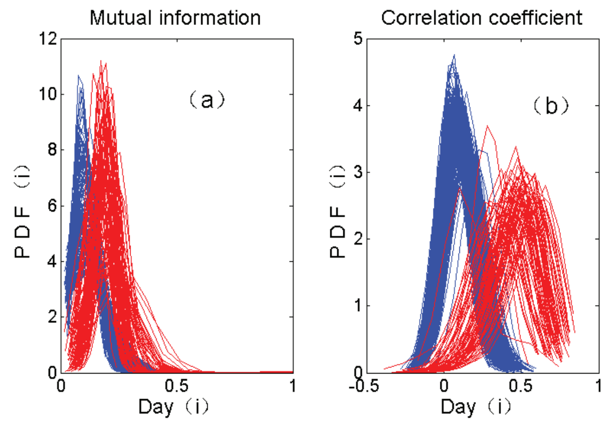

Fig. 2. Distributions of mutual information and cross correlation coefficients for two periods. Blue curves were for the period from November 11, 2013 to September 1, 2014 and the red curves were for the period from June 1, 2015 to September 1, 2015.

They appear to be proportional to each other in most of the situations, as shown in Fig. 4. However, caution should be exercised when one tries to replace one by the other, since there were days that they could be quite different. This is the reason that R^2 in Fig. 4 is smaller than 1. Also, the distributions for the mutual information and cross correlation coefficients are different, as shown in Fig. 2.

IV. Conclusion

One of the more well-defined and significant irrational behaviors in a market is the herding behavior. To suppress market irrationality, it is important to robustly and timely detect herding behavior in a market. So far herding behavior has been mainly studied using low-frequency data and with methods such as CSAD. Such studies can only report whether herding behavior exists in a long time span, such as a few months to even a few years, and thus cannot help with the design of any policies for curbing herding behavior. Equally seriously, methods including CSAD can often give inconsistent results when different window sizes are chosen. One example

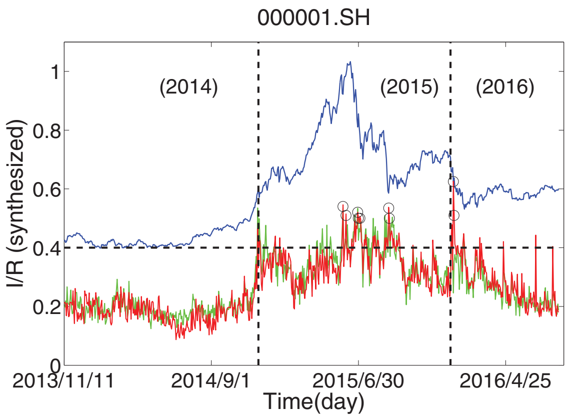

Fig. 3. Daily variation of the synthesized mutual information (green) and cross correlation coefficients (red). For ease of comparison, the composite index was re-scaled and also plotted as the blue curve.

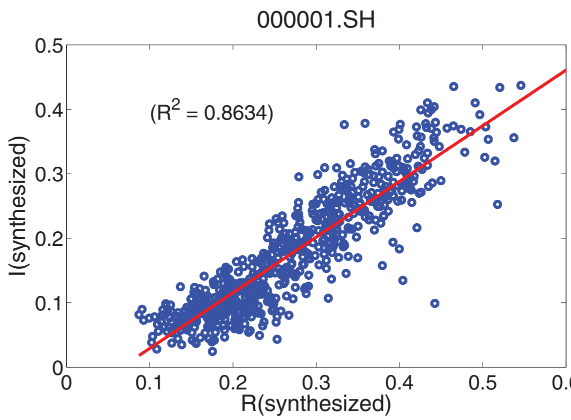

Fig. 4. The relationship between $I_{synthesized}$ and $R_{synthesized}$.

was illustrated in this paper, where herding behavior was claimed to exist or not exist when different window sizes were used. The drawbacks of the popular approaches currently used, i.e., low-frequency data plus methods such as CSAD, motivate us to explore entirely different approaches, by focusing on high-frequency data. By computing mutual information and cross correlation coefficients, we indeed are able to determine whether herding behavior exists in a market on a particular day or not.

This work is exploratory in nature. In future studies, we will further examine how mutual information and cross correlation coefficient can be best combined to offer a single convenient score to detect herding behavior in a short period of time, including as short a a few minutes.

ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China under Grant No. 71661002 and No. 41671532 and Innovation Project of Guangxi Graduate Education No.YCSW2017012.

REFERENCES

[1] ChiangT. C. and D. Z. Zheng, An empirical analysis of herd behavior in global stock markets. Journal of Banking and Finance, 2010.
[2] Lakonishok J, Shleifer A, Vishny R W. The impact of institutional trading on stock prices. Journal of financial economics, 23-43, 1992.
[3] Wermers R. Mutual fund herding and the impact on stock prices. The Journal of Finance, 581-622, 1999.
[4] Christie W. G.Huang R. D. Following the Pied Piper: Do Individual Returns Herd Around the Market. Financial Analysts Journal51(4)31-37, 1995.
[5] Chang Eric C, Cheng Joseph W and Khorana Ajay. An Examination of Herd Behaviorin Equity Markets: An Empirical Perspective. Journal of Banking and Finance, 24:1651-1679, 2000.
[6] Demirer R.Kutan A. M. Does Herding Behavior Exist in Chinese Stock Markets. Journal of International Financial Markets, Institutions and Money, 16(2):123-422006.
[7] Tan, L, Chiang, T. C, Mason, J, Nelling, Edward. Herding behavior in Chinese stock markets: an examination of A and B shares. Pacific-Basin Finance Journal, 61-77, 2008.
[8] Paulo Lao, Harminder SinghHerding behaviour in the Chinese and Indian stock markets, Global Finance Journal, Journal of Asian Econimics 2011.
[9] ChiangT. C, J. D. Li and L. Tan, Empirical investigation of herding behavior in Chinese stock markets: Evidence from quantile regression analysis, Global Finance Journal, 2010.
[10] Rongbao Gu, Kexue Jiang. Herding Behavior and its Dynamics in Shenzhen Stock MarketłłłBased on a New CCK Model. Journal of Asian Ecinomics, 2012.
[11] Welch I. Herding among security analysts. Journal of Financial economics, 58(3): 369-396, 2000.
[12] Bernhardt D, Campello M, Kutsoati E. Who herds. Journal of Financial Economics, 80(3): 657-675, 2006.
[13] Hwang S, Salmon M H. A new measure of herding and empirical evidence. 2001.
[14] Hwang S, Salmon M. Market stress and herding. Journal of Empirical Finance, 11(4): 585-616, 2004.
[15] Dhaene J, Linders D, Schoutens W, et al. The herd behavior index: a new measure for the implied degree of co-movement in stock markets. Insurance: Mathematics and economics, 50(3): 357-370, 2012.
[16] Lee W, Ahn J Y. Financial interpretation of herd behavior index and its statistical estimation. Journal of the Korean Statistical Society, 44(2): 295-311, 2015
[17] Conlon T, Ruskin H J, Crane M. Cross-correlation dynamics in financial time series. Physica A: Statistical Mechanics and its Applications, 388(5): 705-714, 2009.
[18] Preis T, Kenett D Y, Stanley H E, et al. Quantifying the behavior of stock correlations under market stress. Scientific reports, 2012.
[19] Ozoguz A, Rebello M J. Information, competition, and investment sensitivity to peer stock prices. Competition, and Investment Sensitivity to Peer Stock Prices 2013.