JSPS Grant-in-Aid for Scientific Research (S)
Central Bank Communication Design

working paper series
No.010 (August 2019)

Detecting Stock Market Bubbles Based on the Cross-Sectional Dispersion of Stock Prices

Takayuki Mizuno
Takaaki Ohnishi
Tsutomu Watanabe

Research Project on Central Bank Communication
702 Faculty of Economics, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: +81-3-5841-5595 E-mail: watlab@e.u-tokyo.ac.jp
http://www.centralbank.e.u-tokyo.ac.jp/en/

Noname manuscript No.
(will be inserted by the editor)

Detecting stock market bubbles based on the cross-sectional dispersion of stock prices

Takayuki Mizuno · **Takaaki Ohnishi** ·
Tsutomu Watanabe

Received: date / Accepted: date

Abstract A statistical method is proposed for detecting stock market bubbles that occur when speculative funds concentrate on a small set of stocks. The bubble is defined by stock price diverging from the fundamentals. A firm's financial standing is certainly a key fundamental attribute of that firm. The law of one price would dictate that firms of similar financial standing share similar fundamentals. We investigate the variation in market capitalization normalized by fundamentals that is estimated by Lasso regression of a firm's financial standing. The market capitalization distribution has a substantially heavier upper tail during bubble periods, namely, the market capitalization gap opens up in a small subset of firms with similar fundamentals. This phenomenon suggests that speculative funds concentrate in this subset. We demonstrated that this phenomenon could have been used to detect the dot-com bubble of 1998-2000 in different stock exchanges.

Keywords Stock market · Financial bubble · Nowcast · Power law

T. Mizuno
National Institute of Informatics, Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan.
Department of Informatics, SOKENDAI (The Graduate University for Advanced Studies), Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan.
The Canon Institute for Global Studies, Marunouchi, Chiyoda-ku, Tokyo, 100-6511, Japan.
E-mail: mizuno@nii.ac.jp

T. Ohnishi
Graduate School of Information Science and Technology, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan.
The Canon Institute for Global Studies, Marunouchi, Chiyoda-ku, Tokyo, 100-6511, Japan.

T. Watanabe
Graduate School of Economics, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan.
The Canon Institute for Global Studies, Marunouchi, Chiyoda-ku, Tokyo, 100-6511, Japan.

1 Introduction

It is common knowledge in macroeconomics that, as Federal Reserve Board Chairman Alan Greenspan said in 2002, "...it is very difficult to identify a bubble until after the fact; that is, when its bursting confirms its existence." In other words, before a bubble bursts, there is no way to establish whether the economy is in a bubble or not. In economics, a stock bubble is defined as a state in which speculative investment flows into a firm in excess of the firm's fundamentals, so the market capitalization (= stock price $\times$ number of shares issued) becomes excessively high compared to the fundamentals. Unfortunately, it is exceedingly difficult to precisely measure a firm's fundamentals and this has made it nearly impossible to detect a stock bubble by simply measuring the divergence between fundamentals and market capitalization [1–3]. On the other hand, we empirically know that market capitalization and PBR (= market capitalization / net assets) of some stocks increase during bubble periods [4–7]. However, they are also buoyed by rising fundamentals, so it is not always possible to figure out if increases can be attributed to an emerging bubble.

Recently it was reported that real estate bubbles can be detected from increased variability in the prices of houses with similar attributes: the location of the house, the size and floor plan of the house, and so on [8–10]. Similar houses can be regarded as having similar fundamental values. Therefore, increased variation in price among houses with similar attributes signify that speculative money beyond the fundamentals is flowing into a subset in the housing market triggering a real estate bubble.

By applying the real estate bubble detection approach to stock markets, this paper will propose a statistical method of detecting stock market bubbles from growing disparities in market capitalization between firms that are otherwise similar in industrial sector, size, and other attributes. There are similar firms in a stock market. The law of one price would dictate that the fundamentals of these firms are comparable. In other words, if the market capitalization of just few firms among similar firms is exorbitantly high, this suggests that the stocks of those firms are overvalued and that a bubble has formed.

A firm's financial standing is certainly a key fundamental attribute of that firm. Our approach will be to first identify the financial variables that most closely reflect fundamentals using random forest method. We investigate the variation in market capitalization normalized by fundamentals that is estimated by Lasso regression of a firm's financial standing. We will then observe the normalized market capitalization on each stock market from year to year. In non-bubble periods, the normalized market capitalization distribution is thin because market capitalization remains close to the firms' fundamentals. But if a bubble emerges, speculative funds flow into the stock of a small number of firms and this causes the upper tail of distribution to become fatter than during non-bubble periods. In this paper, by observing the upper tail of distribution we will detect the dot-com bubble of 1998-2000 on nine of the leading global exchanges: NASDAQ, NYSE, London SE, Tokyo SE, Paris SE, Korea SE, Shanghai SE, Hong Kong SE, and Taiwan SE. Market capitalization and other financial data were obtained from Thomson Reuters Quantitative Analytics. Classification of firms by industrial sector is according to Thomson Reuters two-digit SIC codes.

The rest of the paper is organized as follows. In Section 2, we identify the financial variables and weighting factors that most strongly reflect firms' fundamentals of different industries on the exchanges based on market capitalization and key financial variables during non-bubble periods. In Section 3, fundamentals roughly are estimated using Lasso regression of the key financial variables. We observe differences in market capitalization that is normalized by fundamentals. Section 4 demonstrates that stock market bubbles can be nowcast by observing the distribution of normalized market capitalization. And finally, Section 5 is a summary of the paper.

2 Financial variables that reflect fundamentals

In definition of bubble, market capitalization remains close to the firms' fundamentals during non-bubble periods. In order to identify the financial variables that most closely reflect fundamentals, we focus on market capitalization in the years 1997 and 2004 when there were manifestly no stock market bubbles on world exchanges. We employ a regression tree and random forests to rank the importance of key financial variables —total assets, net assets, total revenue, operating income, net income, operating cash flow, and number of employees —that reflect the fundamentals.

We set logarithmic market capitalization and key financial variables during non-bubble periods to explain variables and explanatory variables. Figure 1 shows a regression tree for electronics firms listed on NASDAQ in 2004. Branching of the tree enable us to visualize net assets as a most important explanatory variable. Random forests are an ensemble learning method for regression that operates by constructing a forest or a multitude of regression trees [11]. Random forests quantitatively rank the importance of the explanatory variables. Figure 2(a) shows a time series of the importance of financial variables for electronics firms listed on NASDAQ from 1995 to 2005. One can see that net assets are critical in almost all of the years covered. The importance of net assets are around 0.5 in 1997 and 2004. Turning to Figure 2(b), this shows the same time series for retailer firms listed on NYSE from 2004 to 2014. For this sector, one can immediately see that both operating and net incomes are predominantly reflected in market capitalization for these firms.

3 Variation in market capitalization that is normalized by fundamentals

We construct Lasso model to estimate firms' fundamentals from financial variables by assuming that the market capitalization is close to the fundamentals during non-bubble periods. In general, Random forests are more accurate

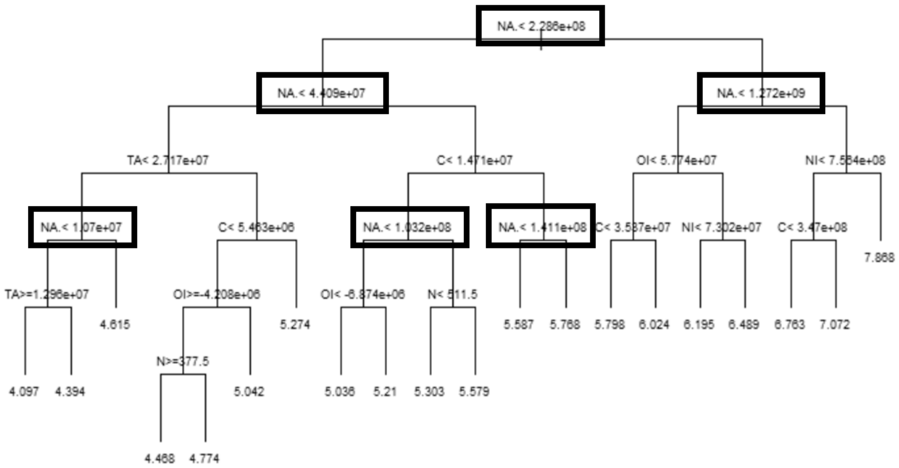

Fig. 1 Regression tree for electronics firms listed on NASDAQ in 2004. The non-explanatory variable is logarithmic market capitalization. Explanatory variables are (TA) total assets, (NA) net assets, (R) total revenue, (OI) operating income, (NI) net income, (C) operating cash flow, and (N) number of employees. Boxes represent branches subject to net assets.

than Lasso regression in identifying important variables. However, the Random forests cannot extrapolate large fundamentals during bubble periods because large financial variables do not appear during non-bubble periods. We chose Lasso model to solve this extrapolation problem in this session.

Assuming that the relationship between fundamentals F and financial variables $\boldsymbol{X}$ follows a power function, we propose the regression as follows,

$$\begin{aligned}\log F = w_{A_T} logX_{A_T} + w_{A_N} logX_{A_N} + w_R logX_R + w_{I_O} logX_{I_O} \\ + w_{I_N} logX_{I_N} + w_C logX_C + w_N logX_N + \epsilon\end{aligned} \quad (1)$$

where all of logarithmic variables are standardized, and the coefficients $\boldsymbol{w}$ depend on the market and industry. X_{A_T}, X_{A_N}, X_R, X_{I_O}, X_{I_N}, X_C, and X_N mean total assets, net assets, total revenue, operating income, net income, operating cash flow, and number of employees, respectively. In Lasso regression, we search the coefficients $\boldsymbol{w}$ that minimize the cost function as follows,

$$S_\lambda(\boldsymbol{w}) = ||logF - \boldsymbol{w} log\boldsymbol{X}||^2 + \lambda \sum_i |w_i| \quad (2)$$

while using the cross-validation. The hyperparameter λ, which maximizes the accuracy of Lasso model predictions (coefficient of determination), is chosen using the cross-validation.

Next, we estimate the coefficients and accuracy of Lasso regression for the electronics firms on NASDAQ and the retailer firms on NYSE during the non-bubble period in 2004. The $logF$ is set to the standardized logarithmic market capitalization. Data is randomly allocated 80% for learning and 20% for testing. Figure 3 displays the relationship between the hyperparameter λ

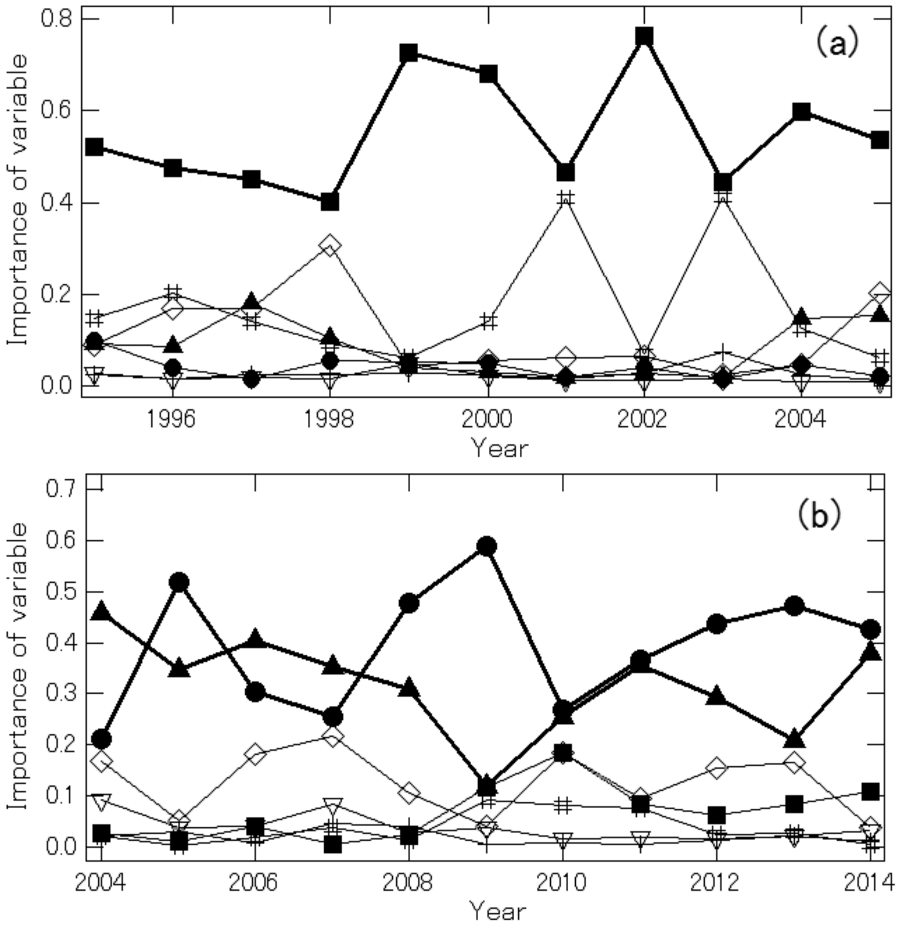

Fig. 2 Importance of financial variables: (a) electronics firms from 1995 to 2005 on NASDAQ, and (b) retailer firms from 2004 to 2014 on the NYSE. (#), (Black square), (+), (Black triangle), (•), (◇), and (∇) represent total assets, net assets, total revenue, operating income, net income, operating cash flow, and number of employees, respectively.

and coefficients $\boldsymbol{w}$. The trend of the coefficients $\boldsymbol{w}$ is consistent with the importance of the random forest model. We obtain the optimal hyperparameter that maximizes the coefficient of determination.

We use the coefficients $\boldsymbol{w}$ during the non-bubble period in 2004 to estimate the fundamentals in different years. We normalize the market capitalization M at time t with the fundamentals F calculated by the Lasso regression (1) of financial variables at time t.

$$M'(t) = M(t)/F(t) \tag{3}$$

Figure 4 shows the cumulative distributions of normalized market capitalization $M'(t)$ on NASDAQ. The distribution function has always a fat-tail and

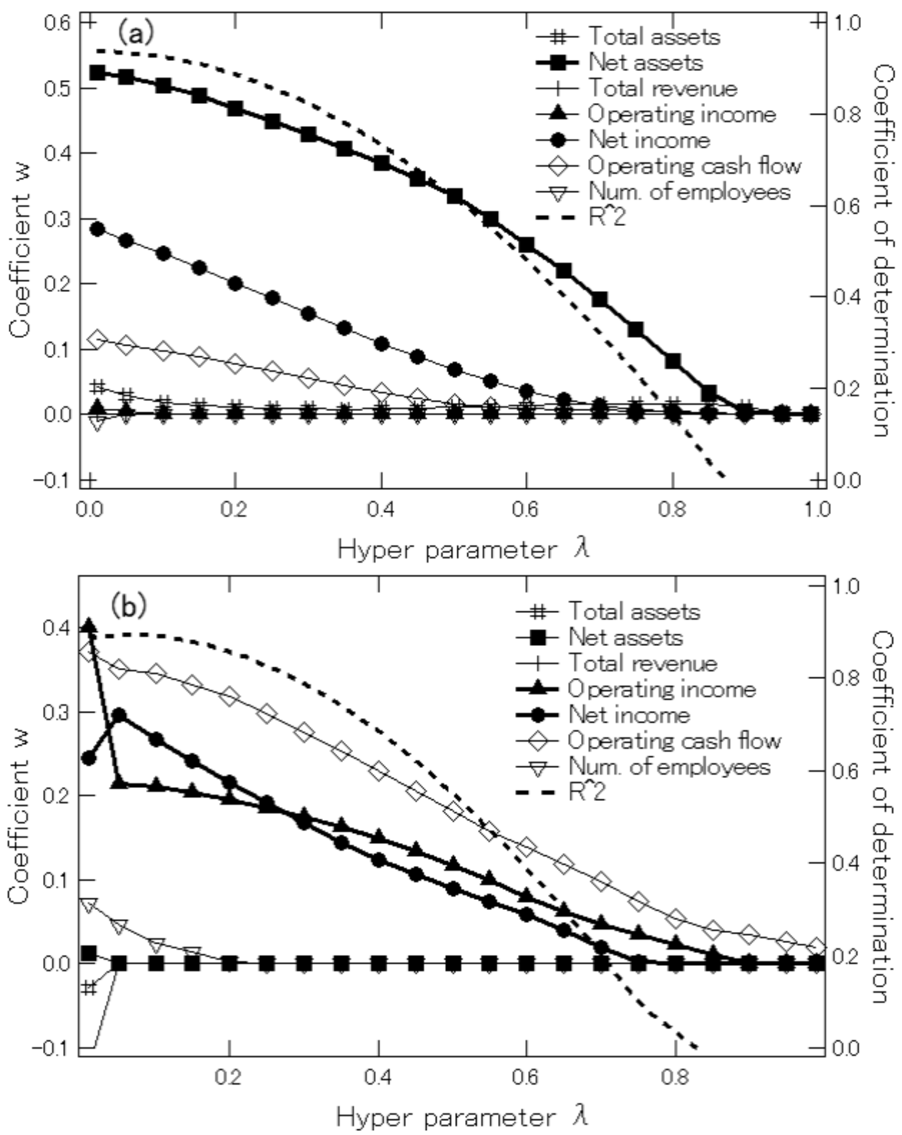

Fig. 3 Coefficient of financial variables in 2004: (a) electronics firms on NASDAQ, and (b) retailer firms on the NYSE. (#), (Black square), (+), (Black triangle), (•), (◇), and (∇) represent total assets, net assets, total revenue, operating income, net income, operating cash flow, and number of employees, respectively. The dashed lines show coefficient of determination.

depends on time t. When the tail is approximated by a power function, the slopes, β_t, are 1.5, 1.0, 1.5, and 1.3 for t =1997, 1999, 2004, and 2014, respectively. Such time dependence is also observed in other stock markets.

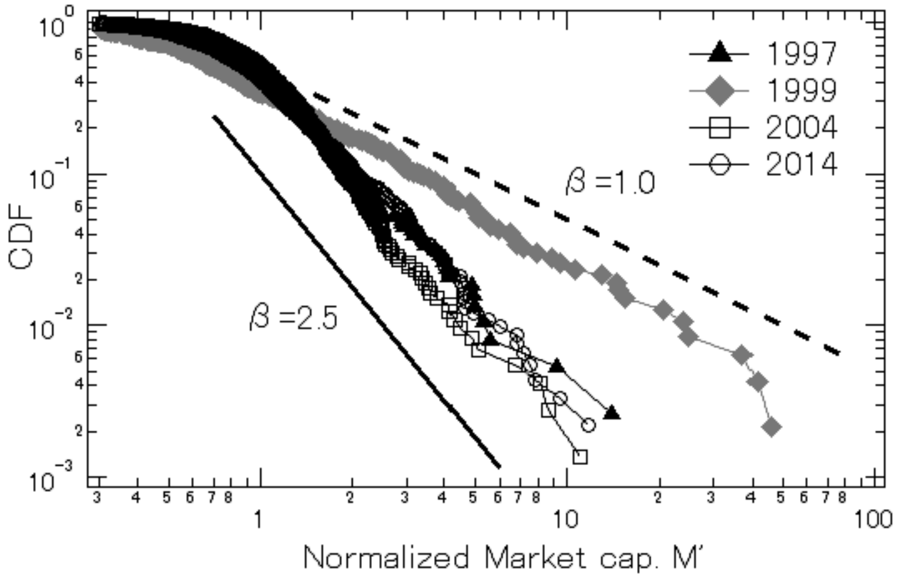

Fig. 4 Cumulative distributions of normalized market capitalization on NASDAQ. (Black triangle), (Black diamond), (Square), and (○) plots represent distributions for 1997, 1999, 2004, and 2014, respectively. The solid and dashed lines show power low slopes, $\beta = 2.5$ and 1.0, respectively.

4 Detecting stock market bubbles: the Dot-com bubble

The dot-com bubble reached its peak toward the end of 1999 as the NASDAQ composite index closed-in on the 5000 mark. Before the bubble in 1997 and after the bubble burst in 2004, the NASDAQ composite index hovered around 2000. Despite the collapse of Lehman Brothers in 2008, NASDAQ has come back and as of 2014 was again approaching 5000. Figure 4 shows distributions of normalized market capitalization in 1997, 1999, 2004, and 2014. The upper tail of distribution grew fat during the dot-com bubble but after the bubble burst, the distribution fell back to the same level as before the bubble. The fat upper tail is not present in 2014. This means that the stock prices increased in line with fundamentals. Market capitalization of firms in 2014 is firmly based on fundamentals.

Figure 5 shows time series of distribution's slopes, β_t, for the nine stock exchanges: NASDAQ (1995-2005), NYSE (1995-2005), London SE (1995-2005), Tokyo SE (1999-2005), Paris SE (1998-2005), Korea SE (1999-2005), Shanghai SE (1999-2005), Hong Kong SE (2000-2005), Taiwan SE (2000-2005). The slopes diminished during the dot-com bubble on all of the exchanges, then immediately steepened again after the bubble collapsed. Essentially, this means that market capitalization increases without any accompanying increase in fundamentals for a subset of stocks during bubble periods. Such an aberration

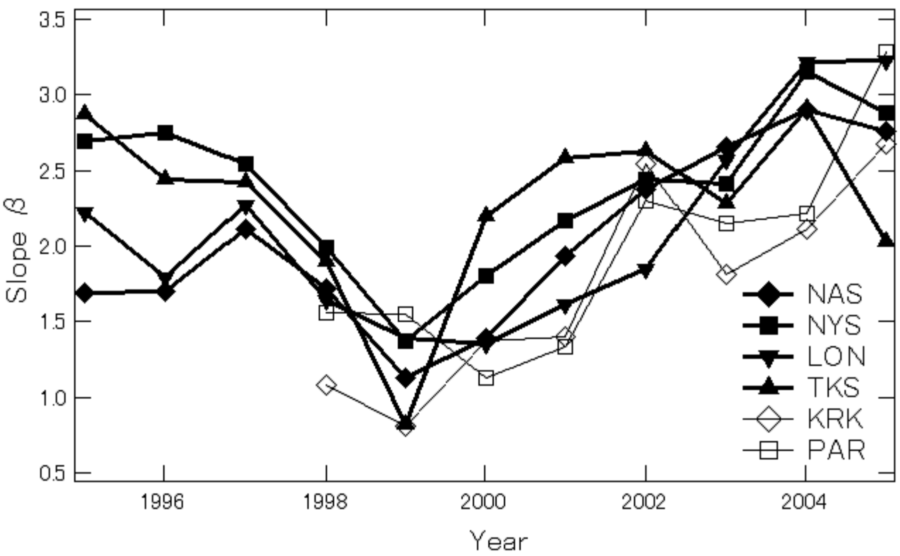

Fig. 5 Distribution's slopes, β_t, of normalized market capitalization. (Black diamond), (Black square), (•), (◇), (Square), (○), (△), (#), and, (+) represent slopes for NASDAQ, NYSE, London SE, Korea SE, Shanghai SE, Tokyo SE, Paris SE, Hong Kong SE, and Taiwan SE, respectively.

can be readily detected by closely monitoring variations in normalized market capitalizations.

5 Conclusions

The defining characteristic of stock market bubbles is divergent between market capitalization (asset prices) and the fundamentals. In accordance with the definition of bubble, "market capitalization remains close to the firms' fundamentals during non-bubble periods", we identified firm's financial variables that most closely reflect fundamentals using machine learning. We observed considerable variability of market capitalization normalized by fundamentals that was estimated by Lasso regression model of financial variables. There is significantly greater divergence between market capitalization and fundamentals during bubbles. We have demonstrated that this phenomenon could have been used to detect the dot-com bubble of 1998-2000 across several stock markets. While it is difficult to determine if an individual firm or stock is caught up in a bubble, we can detect when a stock market is in the midst of a bubble.

We should note that there are types of bubbles that cannot be detected by the approach outlined here. The method we describe only works when identifying the type of bubble in which a concentration of speculative money flows into a small set of stocks. For example, our method would not work in a situation where speculative funds flow evenly into all stocks so that market capitalization diverges uniformly from fundamentals. In order to detect the formation of this kind of bubble, one would have to have knowledge of speculative money flows between financial markets. This is precisely the challenge that we intend to pursue next: detection of bubbles that arise from the concentration of speculative funds flowing into specific financial markets.

Acknowledgements This work was supported by JSPS KAKENHI Grant Numbers 18H05217, 16H05904 and the Center of Advanced Research in Finance (CARF), University of Tokyo.

References

1. Brunnermeier MK (2016) Bubbles. In: Jones G (ed) Banking Crises. Palgrave Macmillan UK, pp 28-36
2. Palan S (2013) A review of bubbles and crashes in experimental asset markets. Jurnal of Economic Surveys 27:570-588
3. Arshanapalli B, Nelson WB (2016) Testing for Stock Price Bubbles: A Review of Econometric Tools. The International Journal of Business and Finance Research 10: 29-42
4. Mizuno T, Ohnishi T, Watanabe T (2016) Power laws in market capitalization during the Dot-com and Shanghai bubble periods. Evolutionary and Institutional Economics Review 13:445-454
5. Kaizoji T (2006) Power laws and market crashes: Empirical laws on bursting bubbles. Progress of Theoretical Physics Supplement 162:165-172
6. Kaizoji T (2006) Precursors of Market Crashes: Empirical laws of the Japan's internet bubbles. European Physical Journal B 50:123-127
7. Kaizoji T, Miyano M (2016) Why does the power law for stock price hold? Chaos, Solitons & Fractals 88:19-23
8. Ohnishi T, Mizuno T, Shimizu C, Watanabe T (2012) Power laws in real estate prices during bubble periods. International Journal of Modern Physics:Conference Series 16:61-81
9. Ohnishi T, Mizuno T, Watanabe T (2019) House Price Dispersion in Boom-Bust Cycles: Evidence from Tokyo. Submitted to Japanese Economic Review.
10. Tay DJ, Chou CI, Li SP, Tee SY, Cheong SA (2016) Bubbles Are Departures from Equilibrium Housing Markets: Evidence from Singapore and Taiwan. PLoS ONE 11:e0166004
11. Breiman L (2001) Random Forests. Machine Learning 45:5-32