Can Stock Market Bubbles in China Be Predicted?

Qiyun Li

Claremont Graduate University

Problem Description

With the emerging of new and complicated time-series methods and machine learning techniques that allow us to better diagnose and understand stock market bubbles to be able to predict and make an early warning on a stock market bubble burst. This project examined the multiple bubble periods on two main stock market indexes in China and then using machine learning classification models to predict the next bubble periods.

Two main questions that this project wants to answer 1) How accurate the GSADF test to define the bubble periods in China stock markets?; 2) Can we use complexity variables like entropy to generate an early warning signal to predict stock market bubble?

The data is weekly from Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZE), which are two official stocks market of China and were established on December 1990 and July 1991 respectively. The SSE has 1520 observations and SZSE has 1307 observations.

Figure 1:The SSE Composite Index

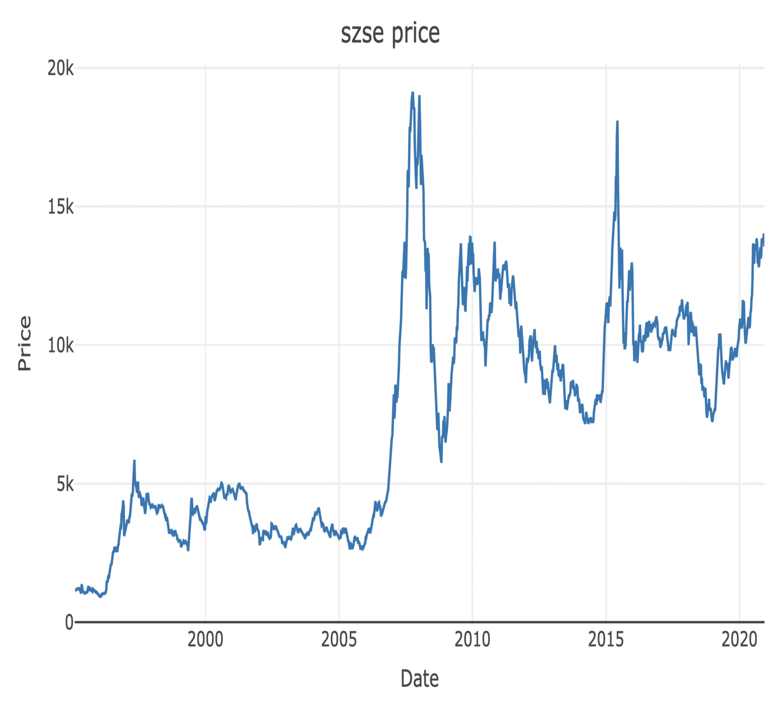


Figure 2:The SZSE Component Index

Theoretical model background

In this project, I used the sup augmented Dickey-Fuller test(SADF) that developed by Phillips et al. (2011) and the generalized sup augmented Dickey-Fuller test (GSADF) by Phillips et al. (2013) to identify bubbles in Shanghai and Shenzhen stock markets. The GSADF is a generalized version of the SADF. Both methods rely on a recursive right-tailed ADF unit root test to detect exuberance and collapse of bubbles. The SADF test is only able to identify a single bubble because of the procedure is using a forward expanding window so that it is fixed starting point. But the GSADF test given user minimum window size and try to use all possible subsamples to test exuberance. GSADF is good for test multiple bubble periods with non-linear structure. Another pros of the GSADF is that it allows date-stamping to find the exact date of past bubbles.

The SADF Test

$$SADF(r_0) = supADF_{r_2}^{0} \tag{1}$$

and has a limit distribution given by,

$$sup\frac{\int_0^{r_2} WdW}{(\int_0^{r_2} W^2)^{1/2}} \tag{2}$$

Rejection of null hypothesis of a unit root, the SADF exceeds the right-tailed critical value from its limit distribution.

The GSADF Test

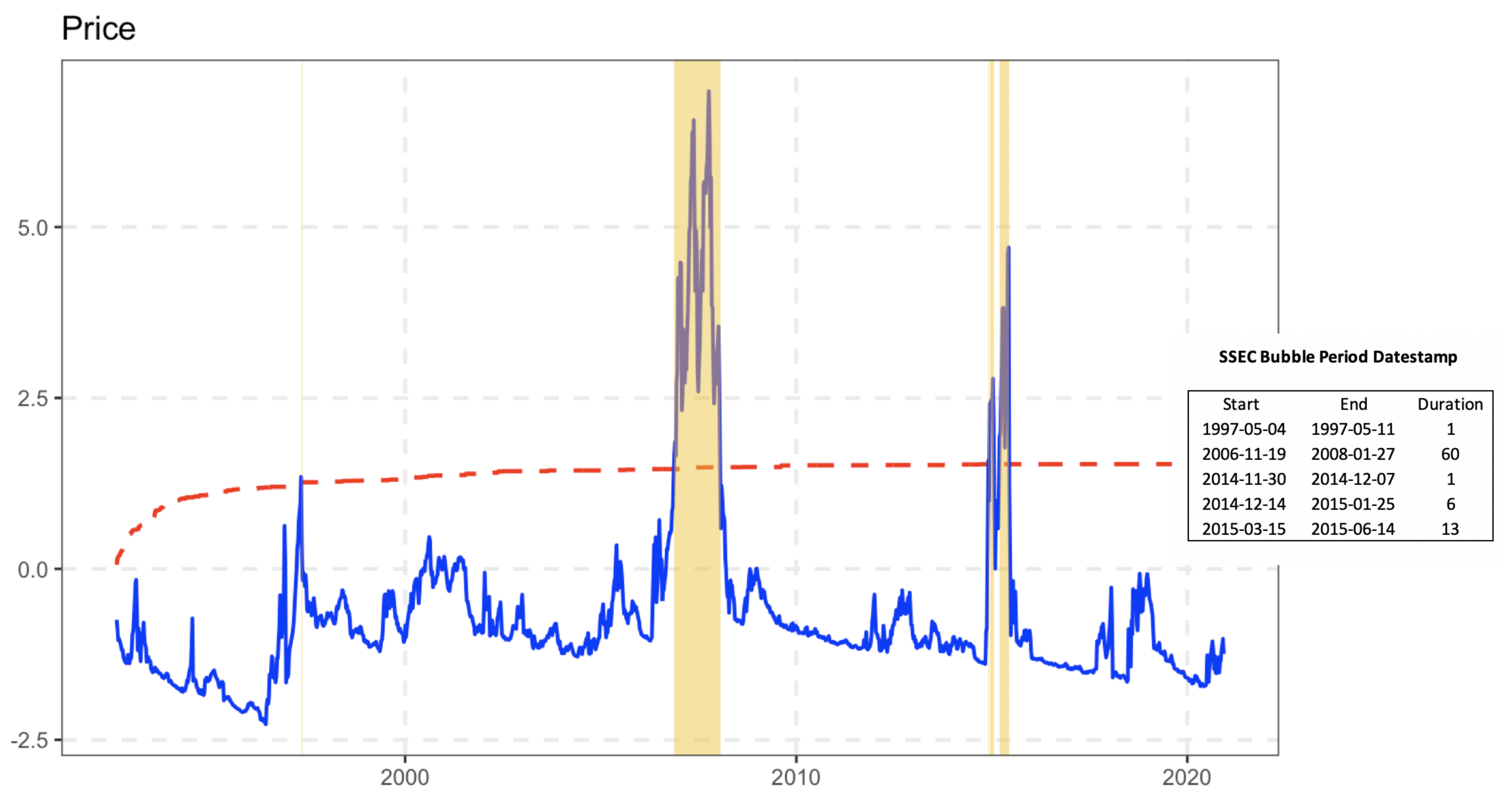
$$GSADF(r_0) = supADF_{r_2}^{r_1} \tag{3}$$

and has a limit distribution given by,

$$sup\frac{\frac{1}{2}r_w[W(r_2)^2 - W(r_1)^2 - r_w] - \int_{r_1}^{r_2} W(r)dr[W(r_2) - W(r_1)]}{r_2^{1/2}[r_w \int_{r_1}^{r_2} W(r)^2 dr - [\int_{r_1}^{r_2} W(r)dr]^2}]^{1/2} \tag{4}$$

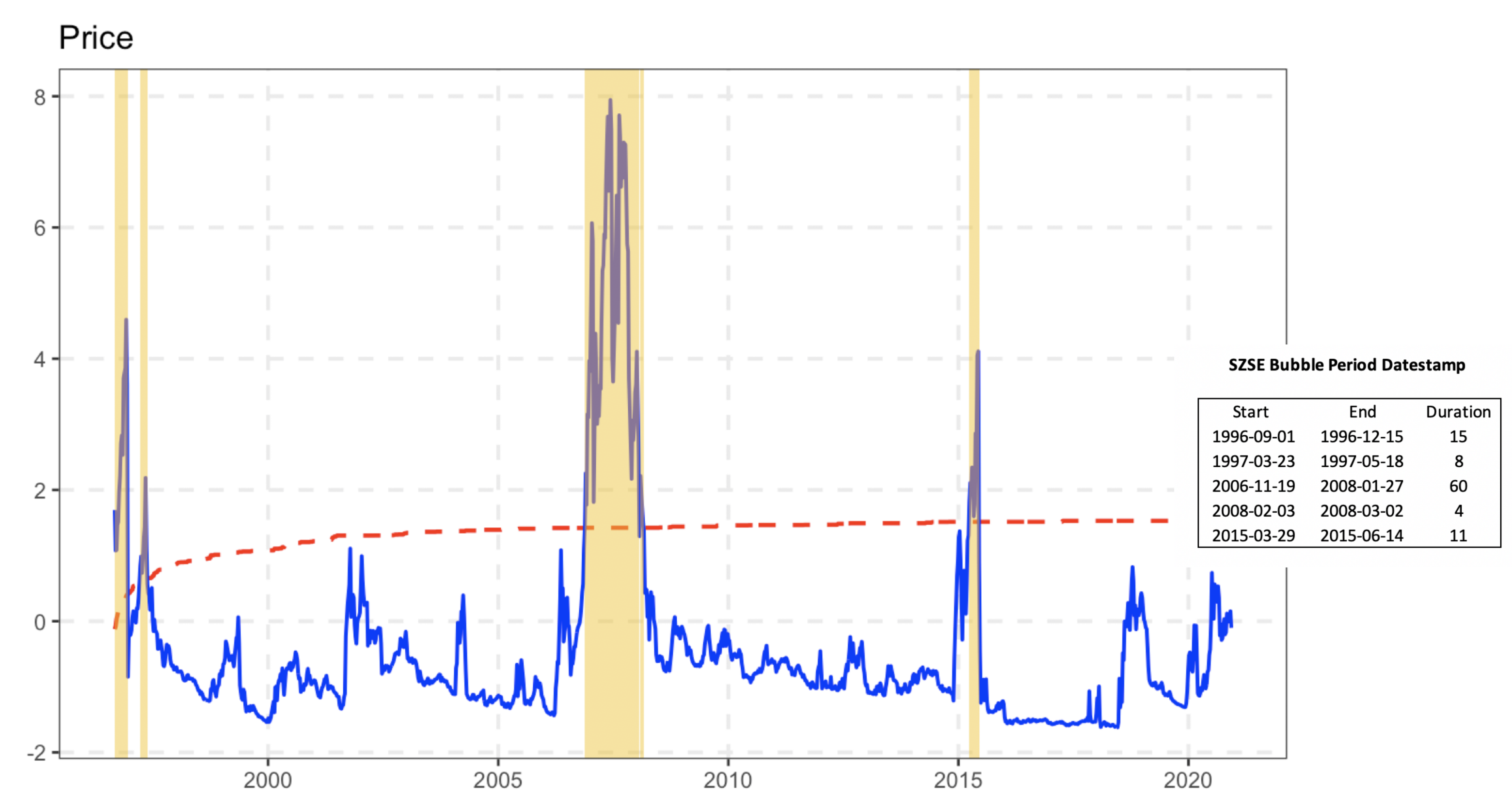
r_w is the size of the window. Same as the SADF test, reject null of a unit root that the test statistic exceeds the right- tail critical value from its limit distribution.

Bubble Diagnosis



Start	End	Duration
1997-05-04	1997-05-11	1
2006-11-19	2008-01-27	60
2014-11-30	2014-12-07	1
2014-12-14	2015-01-25	6
2015-03-15	2015-06-14	13

Figure 3:SSEC Bubble Period



Start	End	Duration
1996-09-01	1996-12-15	15
1997-03-23	1997-05-18	8
2006-11-19	2008-01-27	60
2008-02-03	2008-03-02	4
2015-03-29	2015-06-14	11

Figure 4:SZSE Bubble Period

The GSADF test helped define two main bubble period in China's stock markets. In Shanghai stock exchange index, the first was the subprime mortgage crisis period, which start from 2006-11-19 to 2008-01-27. The second collapsing period was from late 2014 and end in 2015-06-14. Shenzhen stock exchange index has similar bubble period, but the second bubble period start early 2015 and the duration is 11 weeks. Because of the two markets have the similar bubble pattern, I only used SSEC series to predict bubble in the following section and the SZSE should follow the similar result.

Bubble Prediction - Features

I calculated the lag 1 of autocorrelation, standard deviation, permutation entropy, sample entropy, the correlation between two entropies, and price. Scheffer et al. (2009) discovered autocorrelation and standard deviation could be the early-warning signals for system regime shifts. Entropy has been used to describe the level of randomness, disorder or complexity. Hou et al. (2017) showed permutation entropy has strong negative relationship with market crashes depend on the research on SSEC and SZSE.

Features	Correlation with bubble in SSEC
autocorrelation lag1	0.13
standard deviation lag1	0.52
permutation entropy lag1	-0.1
sample entropy lag1	0.06
correlation between permutation entropy and sample entropy lag1	0.08
price lag1	0.44

Bubble Prediction - Results

Results of different models with using 10-fold cross-validation

Classifier	Accuracy	Precision	Recall	AUC	Kappa
Logistic	0.945	0.983	0.958	0.827	0.533
Decision Tree	0.936	0.974	0.958	0.740	0.420
Random Forecast	0.943	0.972	0.968	0.720	0.428
Naïve Bayesian	0.936	0.988	0.945	0.860	0.523
SVM	0.961	0.981	0.977	0.810	0.604

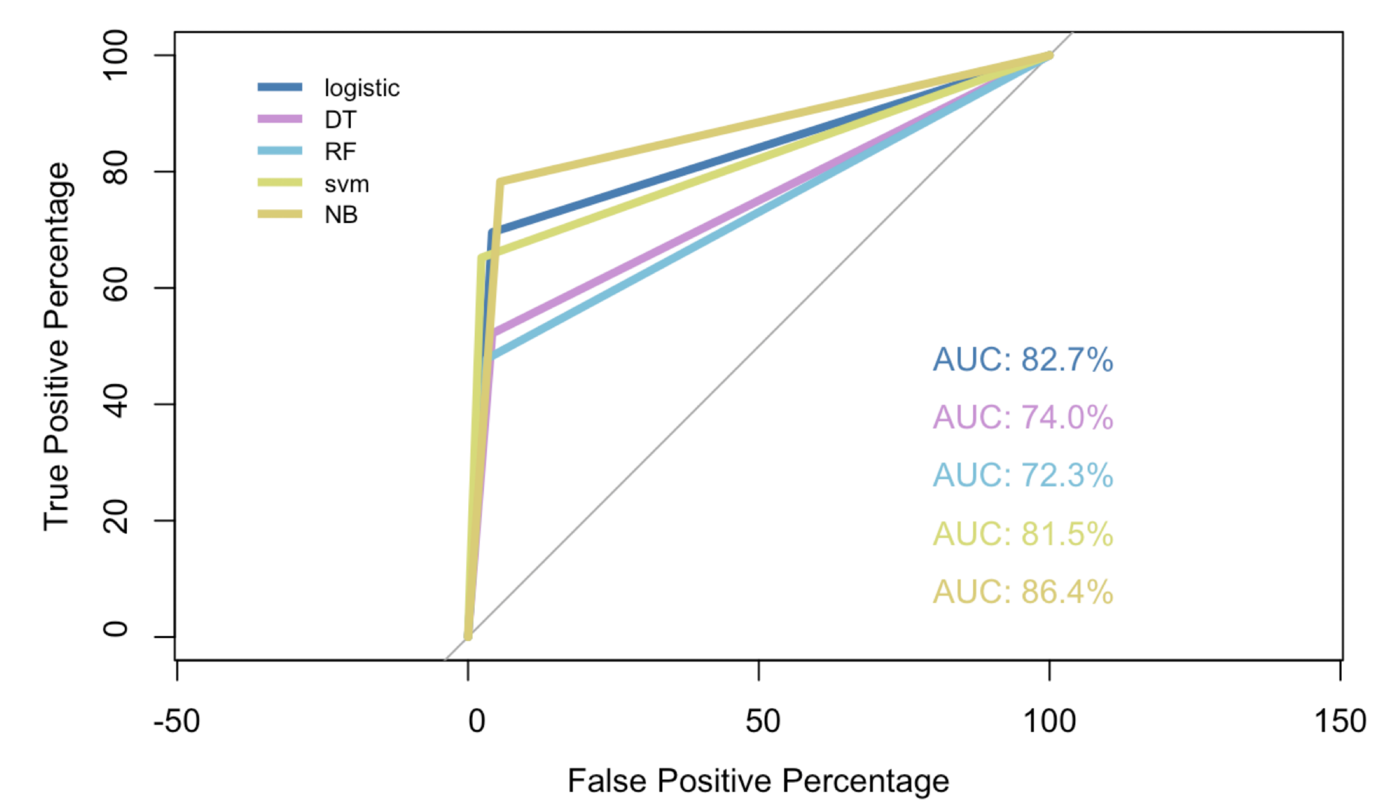


Figure 5:AUC Plot in SSEC

From the five classification models, Naïve Bayesian has the best performance than others. The precision is 0.988 and AUC is 0.86. According to variable importance, price, standard deviation, and permutation entropy has a high variable importance, which could prove that entropy and other complexity variables can be useful in bubble prediction.

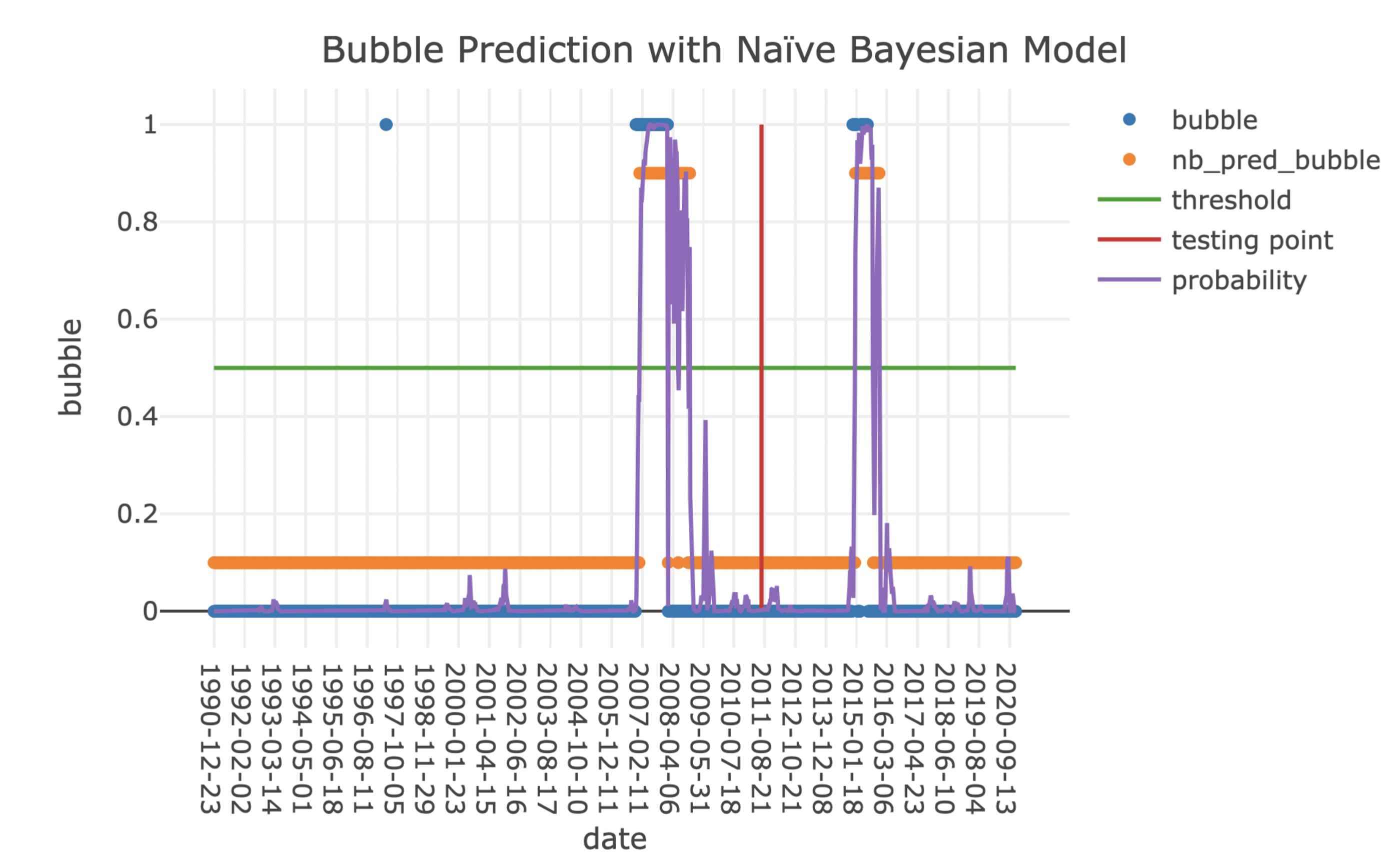


Figure 6:Naive Bayesian Model Prediction Results

Key References

[1] Yangru Wu Peter C.B. Phillips and Jun Yu. Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? *International Economic Review*, 52:201--226, 02 2011.

[2] Shi S. P. Phillips, P. C. B. and J. Yu. Testing for multiple bubbles: Historical episodes of exuberance and collapse in the sp 500. *Economics Statistics Working Paper Series No. 4, July, Singapore: Singapore Management University.*

[3] Bascompte J. Brock W. A. Brovkin V. Carpenter S. R. Dakos V. ... Sugihara G. Scheffer, M. Early-warning signals for critical transitions. *Nature*, 461:53--59.

[4] Jianbo Gao Changxiu Cheng Yunfei Hou, Feiyan Liu and Changqing Song. Characterizing complexity changes in chinese stock markets by permutation entropy. *Entropy*, 19:514.