The Influence of Registration System Reform Based on Computer Big Data on IPO Mispricing-evidence From China's Growth Enterprise Market

Xinbo Zhang1,*
1Xiamen University Malaysia Campus, Xiamen, China
*Corresponding author e-mail: 2415603072@qq.com

Abstract. This paper studies and predicts the influence of registration system reform on IPO mispricing, based on computer big data. This paper takes a quantitative method, using an ARIMA-EGARCH model to carry out data analysis and make predictions. After obtaining the results, I made some suggestions on how to carry out the reform to reduce IPO mispricing.

Keywords: IPO Mispricing, ARIMA-EGARCH, Big Data Analysis, China Financial Reform

1. Introduction
In the stock market, the laws and regulations closely related to the information asymmetry are the information disclosure system. The information disclosure system refers to the legal system in which securities issuing companies disclose all the real information related to their securities in accordance with the law in the links of securities issuance and circulation, so as to provide investors with reference for investment judgment. Here we introduce two aspects of Chinese stock market information disclosure system.

1.1. The difference of information disclosure system between GEM and Main-Board Market
Firstly, the documents required to be disclosed in GEM are not much different from those in the Main-Board Market. But they are focusing on the different point compared to documents in the Main-Board Market. Because the companies listed in this sector are mainly high-tech growth companies, with the technical content as the core feature. As a result, they place more emphasis on the detailed and comprehensive information about the technical content in information disclosure.

In addition, according to Shenzhen Stock Exchange required, the listed companies will be subject to on-site inspection, which refers to the supervision on the production, operation and management sites of listed companies, affiliated enterprises and institutions by consulting and copying documents and materials, viewing physical objects, talking and inquiring, as well as the standardized operation of the inspection objects such as information disclosure and corporate governance (Shenzhen Stock Exchange GEM Stock Listing Requirements, November 2018, p. 7). This clause is also not mentioned in the Main-Board Market listing requirements.
1.2. *The difference of information disclosure system between two IPO systems*

The key point of the registration-based IPO system reform is on information disclosure. Compared with the information disclosure system currently implemented in China, it will bring further improvement in information disclosure[4]. At present, the CSRC promotes reform by implementing the responsibilities of intermediaries (sponsors) such as securities companies, requiring sponsors to accurately grasp the information disclosure material such as the listing application documents of the issuing company, and conduct a comprehensive inspection and verification. This action will further improve the information disclosure quality of listing companies.

2. Literature review

2.1. *EMH & Influence of information asymmetric on volatility of stock price*

EMH can market efficiencies have always been hot topics among researchers. Fama (1997) claimed that market efficiency still the challenge from the literature on long-term return anomalies. Consistent with the market efficiency prediction that apparent anomalies can be due to methodology, most long-term return anomalies tend to disappear with reasonable changes in technique. Liu and Wu(2001) found that Chinese IPO market has a relatively high initial rate of return, and the level of initial yield is less correlated with the public information before and after the IPO[5]. Furthermore, they claimed that the pricing of new shares, the number of listed companies and the number of shares issued are controlled by the government, resulting in a greater demand for new shares than supply. Kong, Xiao and Liu(2010) advocated that the asymmetric information has a significant negative impact on the investment sensitivity to stock price; and the asymmetric information has a significant positive impact on the stock price sensitivity to investment in Chinese stock market[6].

For further test, Yu and Yao (2013) chose the first batch of 28 IPO stocks and the small and medium-sized board composite index of the GEM of 2009~2012 as samples. The result proved that the effective market hypothesis does not exist in the current GEM secondary market. The high IPO price is the main reason for the long-term weakness of the GEM market[7-8]. Titan (2015) argued the EMH is simple in theory but was proved to be very difficult to test and have a precise result. The EMH is simple in theory but was proved to be very difficult to test and have a precise result. Because there is no consensus among economists regarding any of the three forms of EMH, some researches and well-known scientists issued the hypothesis that the reason the EMH is not validated by models is that the models themselves are biased and may provide erroneous results.

2.2. *Opinions about China’s registration reform*

Mainstream scholars believe that registration system reform brings opportunities as well as great challenges. For instance, the rapid expansion of the stock market hits the market; The supervision system is not perfect and the illegal cost is low; Shell resource value decreases, while the delisting rate greatly increases; Because of the uneven professional level, medium and small investors profit difficulty increases. Li Mei and Lu Yang (2014), from the perspective of game theory, established a dynamic game model of incomplete information, based on the behaviour of the proposed listed company, investors and regulatory authorities in the process of IPO pricing under the registration system[9]. According to the research results, it is necessary to clarify the regulatory authorities' pre-audit, in-process supervision and post-issue law enforcement in the IPO process, and increase the reasonable punishment scheme. Through establishing the credit guarantee mechanism and the credit network platform of the capital market, balance the discourse power of the proposed listed company, investors and regulatory authorities in the process of IPO pricing under the registration system[9]. According to the research results, it is necessary to clarify the regulatory authorities' pre-audit, in-process supervision and post-issue law enforcement in the IPO process, and increase the reasonable punishment scheme. Through establishing the credit guarantee mechanism and the credit network platform of the capital market, balance the discourse power of the proposed listed company and investors to establish a mature and effective new issue market. Wang Wenjuan(2016), through comparative analysis and case study, analysed the current situation and problems of the information disclosure in China. The result shows, under the background of the registration system reform, it can be optimized by standardizing the delisting system, strengthening the disclosure of listed companies, improving the protection system of investors' rights and interests, and emphasizing intermediary liability system[10].
3. Methodology

3.1. Methodology introduction

Time series modelling is a dynamic research area that has attracted the attention of researchers over the last few decades. Here I use the ARIMA model raised by Box and Jenkins. Mathematically, the ARIMA \((p,d,q)\) lag polynomials is represented as:

\[
\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) \left(1 - L\right)^d y_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right) \epsilon_t
\]

a. \(p, d\) and \(q\) are non-negative integer, referring to the order of the autoregressive, integrated, and moving average parts of the model respectively.

b. \(\phi_i \theta_i\) are parameters of AR and \(\theta_i \theta_i\) are parameters of MA.

c. The random variable \(\epsilon_t\) is white noise.

ARIMA model is widely used in forecasting various time series, such as S&P 500 index volatility, energy production and consumption, government securities, tourist arrivals, exchange rates, for its efficient capability to generate short-term forecasts[11]. (Radha S, 2006)

In real life, financial time series have a volatility clustering effect. For instance, time series data of exchange rate change rate generally fluctuates violently in one period and relatively gently in another period, showing the phenomenon of "leptokurtosis, fat tail and volatility clustering". In this case, ARMA model and simple random walk model cannot fit such data well, which led to the use of ARCH based models. It is originally proposed by Engle.R (1983). After Modifying and developing by Bollerslev, T (1986), the GARCH model is formed and now widely used. On the basis of that, Nelson & Cao (1991) proposed the EGARCH model, in order to measure the asymmetry of yield volatility. (Liu Shuling, 2008).

The mathematical formulation of EGARCH \((p, q)\) model is given below:

\[
\log \sigma_t^2 = \omega + \sum_{k=1}^{q} \beta_k g(Z_{t-k}) + \sum_{k=1}^{p} \alpha_k \log \sigma_{t-k}^2
\]

a. \((Z_t) = \theta Z_t + \lambda(|Z_t| - E(|Z_t|)).

b. \(\sigma_t^2\) is the conditional variance, \(\omega, \beta, \alpha, \theta, \lambda\) are coefficients.

c. \(Z_t\) is a standard normal variable or come from a generalized error distribution.

GARCH and EGARCH models are extensively used in forecasting volatility of stock returns. However, they are rarely used to analyze other financial time series. The GARCH model is most suitable to deal with the volatility clustering effect, while the EGARCH model is best for capturing the asymmetric effect and forecasting volatility. Kearns and Pagan (1993) examined monthly volatility of the Australian stock market over the period 1875-1987 and fitted ARCH, GARCH and EGARCH models to the data. It was found that the asymmetric EGARCH model outperformed the other models for forecasting the volatility of the returns. (Radha S, 2006)[12]

3.2. General statement of logic

Since there is difference between requirements of information disclosure of main board IPO and GEM IPO, I would like to examine whether this information disclosure difference causes the difference in volatility of the stock price of the companies just gone public. As stated in literature review part, insufficient information disclosure can be a reason for IPO mispricing. My hypothesis is, since information disclosure requirements for GEM IPOs are higher, IPO stocks in GEM market will be less mispriced. By training the model using stock price index of my chosen portfolio 971 trading days after IPO and test the models using data of first 300 days after IPO, I compare the AIC and standard error of regression to determine whether the model fits the data of stock price of first 300 days after IPO well[13]. If the model can fit the data well, I conclude the post IPO early – period mispricing is insignificant, because the model is trained in a relatively stable period and it can fit the test data well, which means the test data itself is relatively stable, and vice versa. I will also use EGARCH models to
test whether information leverage effect exists in both markets, thus provide suggestions to the information disclosure requirements to the registration reform.

3.3. Identifying playing field and data pre-processing

I choose 20 stocks that Int public in 2010. I select 10 stocks from the Main Board of Shanghai Stock Exchange (SSE), 10 stocks from the Second Board of Shenzhen Stock Exchange (SHE). I present their name in the following table:

| Code   | Company                        | Board | Code   | Company                        | Board   |
|--------|--------------------------------|-------|--------|--------------------------------|---------|
| 601801 | Anhui Xinhua Media Co Ltd      | Main  | 300050 | Dingli Communications Corp Ltd | Second  |
| 601877 | Zhejiang Chint Electrics Co Ltd | Main  | 300142 | Walvax BioTech                 | Second  |
| 601179 | China XD Electric Co Ltd       | Main  | 300110 | Huaren Pharmaceutical Co Ltd   | Second  |
| 601158 | Chongqing Water Group Co Ltd   | Main  | 300133 | Zhejiang Huace Film TV Co      | Second  |
| 601098 | China South Publishing & Media Group Co Ltd | Main | 300104 | Leshi Internet Info Tech       | Second  |
| 601000 | Tangshan Port Group Co Ltd     | Main  | 300048 | Hiconics Drive Tech Co         | Second  |
| 601688 | Huatai Securities Co Ltd       | Main  | 300094 | Zhanjiang Guolian Aquatic Products | Second |
| 601369 | XiAn Shaangu Power Co Ltd      | Main  | 300106 | Xingjiang 1st Animal Hmebandry | Second  |
| 601718 | Jihua Group Corp Ltd           | Main  | 300123 | Sunbird Yacht                  | Second  |
| 601288 | Agricultural Bank of China Ltd Class A | Main | 300053 | Zhuhai Orbita Control Eng      | Second  |

I extracted the daily price data from Yahoo Finance. The data set contains daily Open, High, Low, Close (OHLC) prices for each stock, and I choose the daily close price as the proxy for the stock price each day. Then, I calculated a market capitalization weighted index for the stocks[14].

3.4. Unit root test

Here is the KPSS test result for first-differenced stock index of Main board (dMAIN) and first-differenced stock index of Second Board (dSECOND). The result all showed stationary. So the parameter d in ARIMA model is 1.

**Figure 1.** The KPSS test result for first-differenced stock index of Main board (dMAIN) and first-differenced stock index of Second Board (dSECOND)

3.5. ARIMA model building

The general form of the ARIMA(p,1,q) model is:

$$\Delta Y_t = \alpha_1 \Delta Y_{t-1} + \alpha_2 \Delta Y_{t-2} + \cdots + \alpha_p \Delta Y_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

(3)
Now, I proceed to fit MAIN into ARIMA model. I need to determine the AR(p) and MA(q) by looking at autocorrelation function (acf) and partial autocorrelation function (pacf). And here is how it looks like:

**Figure 2.** The KPSS test result for first-differenced stock index of Main board (dMAIN) and first-differenced stock index of Second Board (dSECOND)

The acf & pacf plot of MAIN reinforce my conclusion before: MAIN is non-stationary time series. The acf & pacf of dMAIN shows that p = 3 and q = 3. Then, I start to examine the best combination of (p,d,q) by trial & error. I try to fit MAIN into ARIMA(3,1,3), ARIMA(2,1,3), ARIMA(2,1,2), ARIMA(2,1,1), ARIMA(2,1,1) with a drift. Finally, I find that ARIMA(2,1,1) with drift has the lowest $AIC_c$, and I decide to use this model. And here is the model:

**Figure 3.** The model

The residual plot of ARIMA(2,1,1) is stationary.

Then, I repeat the same procedure on SECOND. First, I look at acf & pacf plot for SECOND & dSECOND.

From acf & pacf plot of dSECOND, I decide to try p = 2, q = 2; I fit SECOND into ARIMA(2,1,2), ARIMA(2,1,1), and find that ARIMA(2,1,1) has the lowest $AIC_c$. I decide to use ARIMA(2,1,1). And here is how the model and residual looks like:\[15\]:

```
Series: dMAIN
ARIMA(2,1,1) with drift

Coefficients:
.ar1  ar2  ma1  drift
    -0.6935  0.0026  0.493  -0.0033
s.e.   0.1080   0.0355  0.189    0.0022

sigma2 estimated as 0.00835:  log likelihood=-1282.85
AIC=-2202.64  AICC=-2291.83

Training set error measures:
          ME     RMSE      MAE      MPE     MAPE      MASE      ACF1
Training set 0.00876004 0.049741433 0.03032965 1.2233683 0.9901203
```
The residuals from ARIMA(2,1,1) look stationary as well.

Figure 4. The model and residua

3.6. EGARCH model building

I use the ARIMA model built in 3.3 as the mean equation for my EGARCH model. The aim of building EGARCH model is to examine whether the information leverage effect exists: if the third coefficient of variance equation of EGARCH model is significant, I can accept the hypothesis that the information leverage effect exists, which means either good news have more significant effect on the portfolio index than bad news, or vice versa. I use the following as variance equation of EGARCH model:

$$\log(\sigma_t^2) = \omega + \delta_1 \frac{z_t^2}{\sigma_{t-1}^2} + \delta_2 \frac{z_{t-1}^2}{\sigma_{t-1}^2} + \beta \log(\sigma_{t-1}^2)$$

(4)

4. Results

4.1. Results for ARIMA Model

As stated before, I have determined to use ARIMA (2, 1, 1) model. Here is the training result:

![Figure 6. The model and residua](image-url)

Following are results for testing the model:
I can see clearly from the result that the GEM stocks have higher degree of mispricing after IPO. The training model has an AIC of 1.38 and standard error of regression 0.48, while the testing model has an AIC of 0.26 and standard error of regression 0.27. The model fits the testing data worse, which means the testing data has a larger degree of volatility\[16\].

However, I did see the model fits main board stocks better. The training model has an AIC of \(-2.59\) and standard error of regression 0.07, while the testing model has an AIC of \(-2.2\) and standard error of regression 0.09. Therefore, the result indicates the testing data has lower volatility.

In conclusion for this part, the GEM market has a high degree of stock mispricing after IPO, while main board market has lower degree of mispricing.

4.2. Result for EGARCH model

The result for EGARCH model is really surprising. Both results show a positive but significant \(C\) (7), which means good news have more significant effect on the stocks than bad news. Normally the
condition should be bad news have more significant effect[17]. Therefore, the regulatory should pay attention to this phenomenon, and the suggestions given by me are provided next section.

5. Conclusion and Suggestion

5.1. Conclusion
It is important to understand why my conclusion shows that such a stock market reform cannot improve IPO mispricing in China, I think the most important reason is too many individual investors.

Chinese stock market is an immature and inefficient market. One of the most characteristic is that the individual investors own a huge part of the market. According to Chinese Stock Market Investors Report, the percentage of individual investors is 99.77%. And 77% of them are the small and medium sized- investors whose funds are less than 500 thousand RMB. Individual investors are irrational when they make decisions and don’t have the concept of stop loss and control risk. Many individual investors are loss aversion. When bad news occurs, they are unwilling to give up because they cannot tolerate the loss occurs.

5.2. Suggestion about reformation of registration-based IPO system
Based on my result and discussion, I doubt that whether Chinese stock market is mature enough for reformation. To help the reformers, I will give some suggestions.

1. Chinese financial institution market needs to be improved to attract more and more individual investors with their trust. In this way the number of individual investors can be reduced and institutional investors can be increased. With less individual investors and more institutional investor, Chinese stock market will become more efficient and mature. The negative effect on reformation can be alleviated.

2. Government should encourage securities company take care of their individual investors through propaganda and education. For example, they can set up some classes related to the strategies and financial concept about the stock market on the internet, which will help reduce the individual investors’ irrationality.

3. In many situations, institutional investors can get insider information, which will increase the information asymmetry. Thus, the act about increasing transparency should be introduced. The act should require the equal information disclosure to individual investors and institutional investors. This is also another way to reduce the speculation opportunities of institutional investors and protect the interest of individual investors.

References
[1] Fama (1997). Market efficiency, long-term returns, and behavioral finance. [online] Available at: https://www.researchgate.net/publication/228532383_Market_efficiency_long-term_returns_and_behavioral_finance, Financial and legal information, 27(6), 28-32.
[2] Kong, Xiao and Liu (2010). [online] Available at: https://www.researchgate.net/publication/242345332_Asymmetric_information_firm_investment_and_stock_prices
[3] Li Mei, Lu Yang, (2014), Research on the optimization of IPO pricing under registration system, based on the perspective of game theory, Page 88.
[4] LIU and WU (2001). [online] Available at: https://www.ixueshu.com/document/0dcfb8445399855c318947a18e7f9386.html
[5] Liu Shuling, In Tao, Ge Jun, (2008), For ecasting and the Choice of RMB Exchange Rate Model, Base on ARIMA Model and GARCH Model, Page 91-93
[6] Radha S, M. Thenmozhi, (2006), Forecasting short term interest rates using ARMA, ARMA-GARCH and ARMA-EGARCH models, Page 1-4
[7] Shenzhen Stock Exchange. (November, 2018). Shenzhen Stock Exchange GEM Stock Listing Requirements. Retrieved 20 October, 2019, from: http://www.szse.cn/lawrules/index.html
[10] Titan (2015). The Efficient Market Hypothesis: Review of Specialized Literature and Empirical Research. [online] Available at: https://www.sciencedirect.com/science/article/pii/S2212567115014161

[11] Wang Wenjuan, (2016), Research on information disclosure of Chinese listed companies under the registration system reform, Page 37-42.

[12] YU and YAO (2013). [online] Available at: https://www.ixueshu.com/document/a9f827adad76426f318947a18c7f9386.html

[13] Engle, R. F. (1983). Estimates of the Variance of US Inflation Based upon the ARCH Model. Journal of Money, Credit and Banking, 15(3), 286-301.

[14] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), 307-327.

[15] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica: Journal of the Econometric Society, 347-370.

[16] Kearns, P., & Pagan, A. R. (1993). Australian stock market volatility: 1875–1987. Economic Record, 69(2), 163-178.