Correlation Analysis Between Stock Prices And Four Financial Indexes for Some Listed Companies of Mainland China

**Abstract.**

Stock prices forecasting is vital for making informed investment decisions. This paper study on which financial index have a reasonably high correlation with stock price during the past thirteen years. The four financial indexes include revenue, net profit, the earning per share, the main business income per stock. The results show that, in pharmaceutical industry, stock prices have a relatively high correlation with these indexes, which provide a valuable reference to investors when they make a long-term investment decisions.

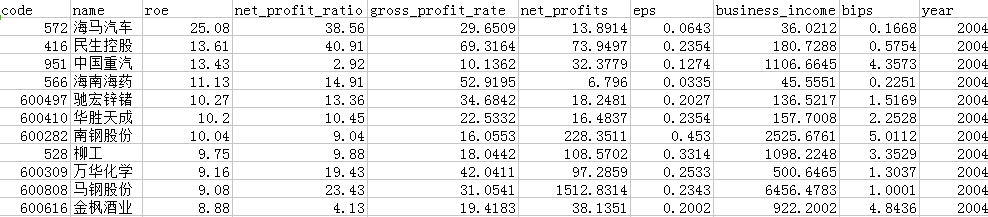
**Keywords**: Stock prices, Correlation Analysis, Financial data

**Introduction**

Stock prices prediction is an important challenge in financial industry, however, most focus on financial time series prediction. Generally speaking , stock prices in short term have a large volatility which cause a high risk for investors and accuracy reduction in predication. There are much stock investment theories which can be categorized into value investing strategies, growth strategies and so on. This paper didn’t discuss which was much better or how to invest have a better performance but just made a correlation analysis between stock prices and financial performance in the past thirteen years for different listed companies. In the long term, we discovered the real quality of a listed company like competitiveness, earning stability, growth speed, etc..On the other hand, we also found some companies had little been affected by market volatility even in stock-market crash. This paper provided a quantitative view of stock market performance and earning data in the long term, which was believed to have a high reliance than some abstract value analysis.

**Data description**

All data were collected and processed in python and some third-party libraries. The financial data we used were all collected from Sina financial and economics through a python library called TuShare, which is a free and open-sourced finance and economics python library. This library achieving data collection, data cleaning and data saving offered financial analysts easy-used financial data. This paper just used a part of profit data of every year like fig.1 shows. As we can see, the profit table includes ROE, Net margin, Gross margin, Net profit, The earning per share, Revenue, The revenue per share of different company from 2004 to 2016. However, we just used Net profit, Revenue, The earning per share and The revenue per share for they were growing over time.

 Fig 1. A part of the profit data getting from the Sina financial and economics.

**Data processing**

All collecting data should be pre-processed before being analyzed because of some missing data, repeated data, false data. All steps of pre-processing are showed as follow figure.

Fig 2. The process of pre-processing

The first step was data cleaning in order to remove missing data, repeated data, false data and after that remaining companies count up to 636. The next step was normalization by features in order to compare them at the same scale.

After the above two steps, we needed to calculate the correlation coefficient between normalized prices and normalized features like Net profit, Revenue, The earning per share and The revenue per share. After normalization, we had to extracted the mean prices and the above four indexes of every year for every company. However,another processing had to be said was that prices would be adjusted for divide and dividends, so there were two kinds of comparing, one was split-adjusted share prices comparing with Net profit and Revenue because they were not affected by divide and dividends, the other one was no-adjusted share prices comparing with The earning per share and The revenue per share because they would be affected by divide and dividends.

The processed data are shown as follows.

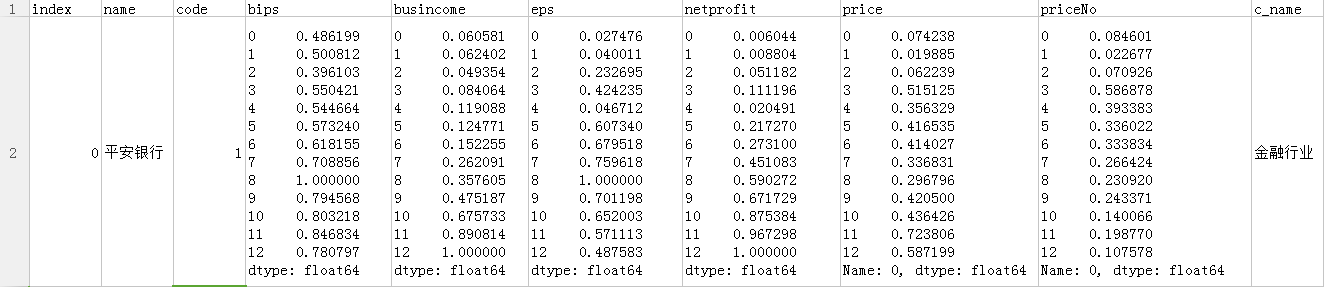


Fig 3. One example of the processed data, all have been normalized into between 0 and 1. The column called ‘price’ is price adjusted for divide and dividends and ‘priceNo’ is price not adjusted, the other are four financial indexes.

Then we needed to calculate the correlation coefficient, which was a number that quantifies a type of correlation and dependence, meaning statistical relationships between two or more values in fundamental statistics. A high value (approaching +1.00) is a strong direct relationship, a low negative value (approaching -1.00) is a strong inverse relationship, and values near 0.00 indicate little, if any, relationship. The formula can be shown as follows.

 (1)

We calculated the correlation coefficient between column ‘price’ and column ‘busincome’, ’netprofit’ respectively and between column ‘priceNo’ and column ‘eps’ ,column ‘bips’ respectively. The results are shown in the next part.

**Results**

The final results can be shown as follows. Fig .4 Shows the top 25 industries of whose correlation coefficient between business income and no-adjusted share prices are larger than 0.7. Fig .5 Shows the top 25 industries of whose correlation coefficient between net profit and no-adjusted share prices are larger than 0.7. Fig.6 shows the top 25 industries of whose correlation coefficient between the earning per share and split-adjusted share prices are larger than 0.7. Fig .7 Shows the top 25 industries of whose correlation coefficient between the revenue per share and split-adjusted share prices are larger than 0.

Fig 4. Top 25 industries of whose correlation coefficient between business income and no-adjusted share prices are larger than 0.7 .

Fig 5. Top 25 industries of whose correlation coefficient between net profit and no-adjusted share prices are larger than 0.7.

Fig 6. Top 25 industries of whose correlation coefficient between the earning per share and split-adjusted share prices are larger than 0.7.

Fig 7. Top 25 industries of whose correlation coefficient between the revenue per share and split-adjusted share prices are larger than 0.7.

**Conclusion**

From the result last part, we found that there was a high correlation coefficient in pharmaceutical industry in these four indexes, which means that the stock prices were much decided by the financial performance than other industries. After that, in the automobile manufacturing industry and the brewing industry, there are also relatively high correlation coefficient which indicated that in these industries stock prices were also stable and not affected by market sentiment easily.

**Reference**

[1] Abraham, A., Nath, B. & Mahanti, P. K. (2001), Hybrid intelligent systems for stock market analysis, in ‘Proceedings of the International Conference on Computational Science-Part II’, Springer-Verlag, London, UK, pp. 337–345.  
[2] Abraham, A., Philip, N. S. & Saratchandran, P. (2003), ‘Modeling chaotic behavior of stock indices using intelligent paradigms’, Neural, Parallel Sci. Comput. 11(1 & 2), 143–160.  
[3] Armano, G., Marchesi, M. & Murru, A. (2005), ‘A hybrid genetic-neural architecture for stock indexes forecasting’, Information Sciences 170(1), 3–33.  
[4] Bekiros, S. D. & Georgoutsos, D. A. (2008), ‘Direction-of-change forecasting using a volatilitybased recurrent neural network’, Journal of Forecasting 27(5), 407–417.  
[5] Chen, A.-S., Leung, M. T. & Daouk, H. (2003), ‘Application of neural networks to an emerging financial market: forecasting and trading the taiwan stock index’, Comput. Oper. Res. 30(6), 901–923.  
[6] Chen, Q.-A. & Li, C.-D. (2006), ‘Comparison of forecasting performance of ar, star and ann models on the chinese stock market index’, Advances in Neural Networks 3973, 464–470.

[7] Huarng, K. & Yu, H.-K. (2005), ‘A type 2 fuzzy time series model for stock index forecasting’, *Physica A: Statistical Mechanics and its Applications* 353, 445–462.  
[8] Jaruszewicz, M. & Mandziuk, J. (2004), ‘One day prediction of nikkei index considering information from other stock markets’, *International Conference on Artificial Intelligence and Soft Computing* 3070, 1130–1135.  
[9] Jia, G., Chen, Y. & Wu, P. (2008), ‘Menn method applications for stock market forecasting’, *Advances in Neural Networks* 5263, 30–39.  
[10] Kim, K.-J. (2004), ‘Artificial neural networks with feature transformation based on domain knowledge for the prediction of stock index futures’, *Intelligent Systems in Accounting, Finance & Management* 12(3), 167–176.