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

Encai Zhang, Prof.Chen
Yuan Ze University

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 index include revenue, net profit, the earning per share, the main business income per stock. The result shows 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.

code	name	roe	net_profit_ratio	gross_profit_rate	net_profits	eps	business_income	bips	year
572	海马汽车	25.08	38.56	29.6509	13.8914	0.0643	36.0212	0.1668	2004
416	民生控股	13.61	40.91	69.3164	73.9497	0.2354	180.7288	0.5754	2004
951	中国重汽	13.43	2.92	10.1362	32.3779	0.1274	1106.6645	4.3573	2004
566	海南海药	11.13	14.91	52.9195	6.796	0.0335	45.5551	0.2251	2004
600497	驰宏锌锗	10.27	13.36	34.6842	18.2481	0.2027	136.5217	1.5169	2004
600410	华胜天成	10.2	10.45	22.5332	16.4837	0.2354	157.7008	2.2528	2004
600282	南钢股份	10.04	9.04	16.0553	228.3511	0.453	2525.6761	5.0112	2004
528	柳工	9.75	9.88	18.0442	108.5702	0.3314	1098.2248	3.3529	2004
600309	万华化学	9.16	19.43	42.0411	97.2859	0.2533	500.6465	1.3037	2004
600808	马钢股份	9.08	23.43	31.0541	1512.8314	0.2343	6456.4783	1.0001	2004
600616	金枫酒业	8.88	4.13	19.4183	38.1351	0.2002	922.2002	4.8436	2004

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.

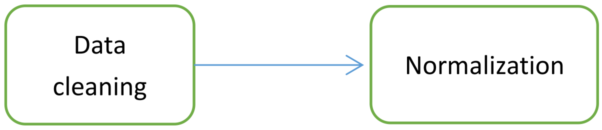

Fig 2. The process of pre-processing

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.

1	index	name	code	bips	busincome	eps	netprofit	price	priceNo	c_name
2	0	平安银行	1	0 0.486199	0 0.060581	0 0.027476	0 0.006044	0 0.074238	0 0.084601	金融行业
				1 0.500812	1 0.062402	1 0.040011	1 0.008804	1 0.019885	1 0.022677	
				2 0.396103	2 0.049354	2 0.232695	2 0.051182	2 0.062239	2 0.070926	
				3 0.550421	3 0.084064	3 0.424235	3 0.111196	3 0.515125	3 0.586878	
				4 0.544664	4 0.119088	4 0.046712	4 0.020491	4 0.356329	4 0.393383	
				5 0.573240	5 0.124771	5 0.607340	5 0.217270	5 0.416535	5 0.336022	
				6 0.618155	6 0.152255	6 0.679518	6 0.273100	6 0.414027	6 0.333834	
				7 0.708856	7 0.262091	7 0.759618	7 0.451083	7 0.336831	7 0.266424	
				8 1.000000	8 0.357605	8 1.000000	8 0.590272	8 0.296796	8 0.230920	
				9 0.794568	9 0.475187	9 0.701198	9 0.671729	9 0.420500	9 0.243371	
				10 0.803218	10 0.675733	10 0.652003	10 0.875384	10 0.436426	10 0.140066	
				11 0.846834	11 0.890814	11 0.571113	11 0.967298	11 0.723806	11 0.198770	
				12 0.780797	12 1.000000	12 0.487583	12 1.000000	12 0.587199	12 0.107578	
				dtype: float64	dtype: float64	dtype: float64	dtype: float64	Name: 0, dtype: float64	Name: 0, dtype: float64	

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.

$$\rho_{xy} = \frac{Cov(r_x, r_y)}{\sigma_x \sigma_y} \tag{1}$$

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.

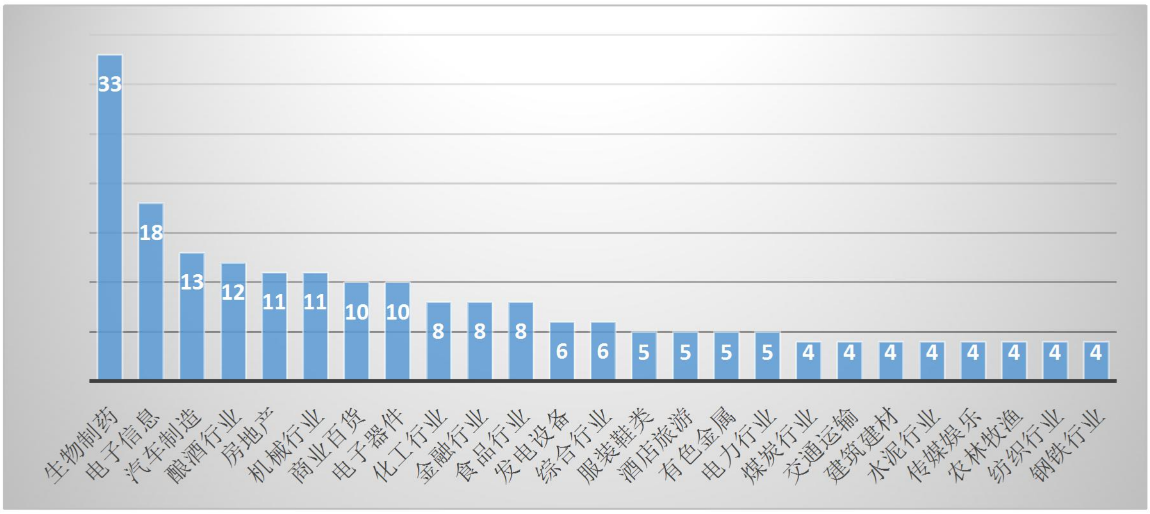

Fig 4. Top 25 industries of whose correlation coefficient between business income and no-adjusted share prices are larger than 0.7 .

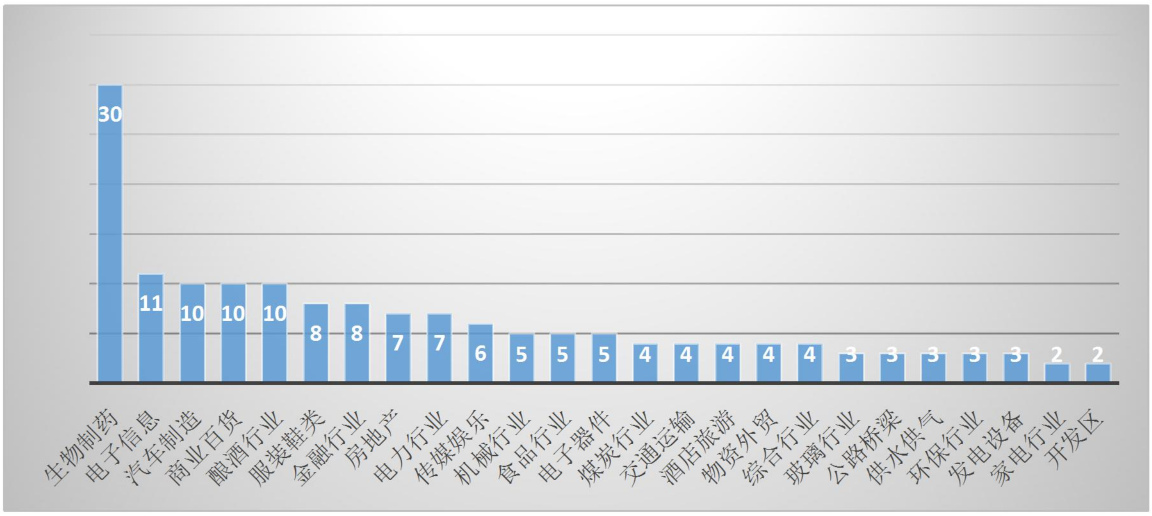

Fig 5. Top 25 industries of whose correlation coefficient between net profit and no-adjusted share prices are larger than 0.7.

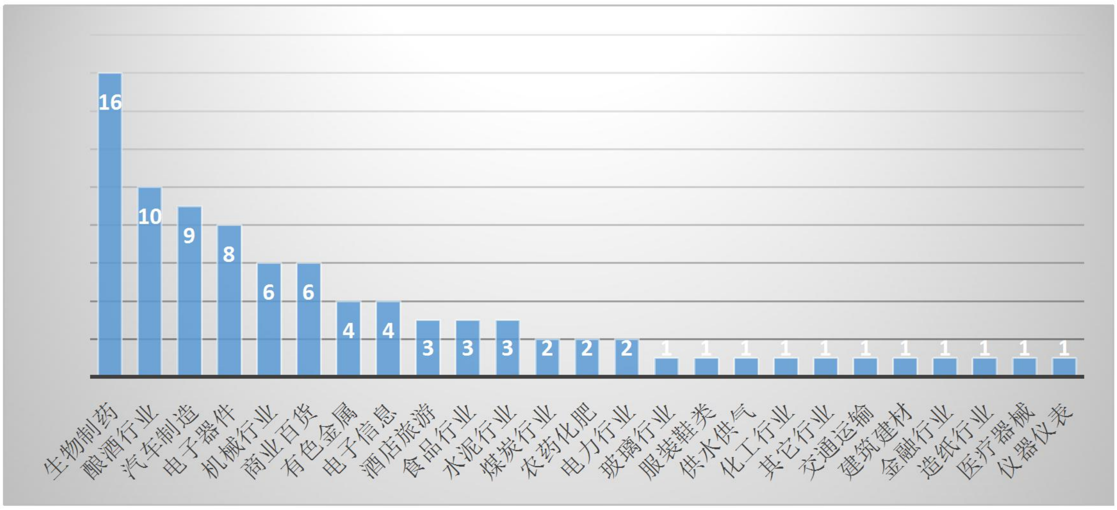

Fig 6. Top 25 industries of whose correlation coefficient between the earning per share and split-adjusted share prices are larger than 0.7.

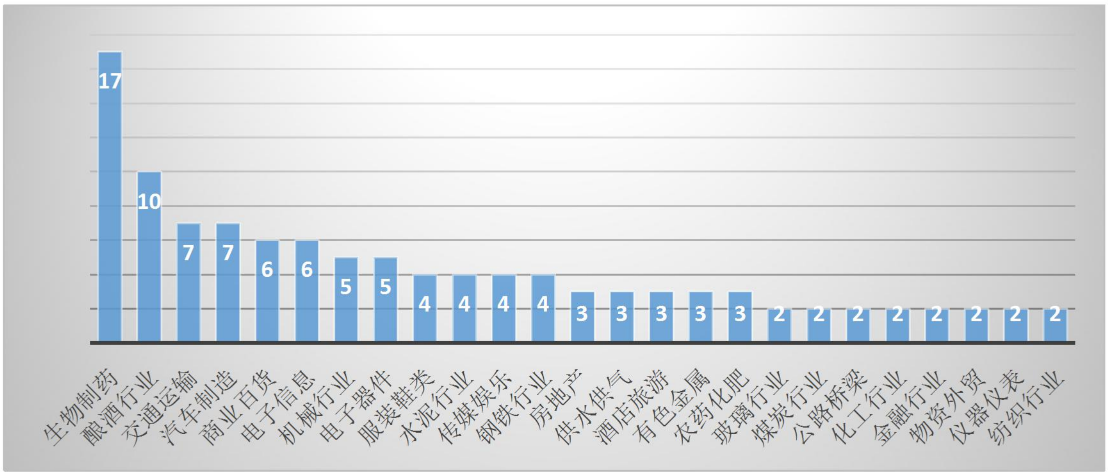

Fig 7. Top 25 industries of whose correlation coefficient between the revenue per share and split-adjusted share prices are larger than 0.7.

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.