An Effect of Financial Indicators on Share Price of largest US companies

Jiří Pešík[1]*, Petra Taušl Procházková[2]

[1]IUniversity of West Bohemia, Faculty of Economics, Department of Business Administration and Management, Pilsen, Czech Republic

[2] University of West Bohemia, Faculty of Economics, Department of Business Administration and Management, Pilsen, Czech Republic

Abstract: There is fierce competition among various approaches for a share price determination, as technical, fundamental, and behavioural analyses try to explain the stock market phenomenon. This paper aims to contribute to the fundamental analysis and compare the importance of various financial indicators for investors. It uses financial data of one hundred largest companies from the New York Stock Exchange (NYSE) and the National Association of Securities Dealers Automated Quotations (NASDAQ) and determines how significantly the financial indicators influence the price of the share, using a panel regression model as the main statistical approach. In the second part of the paper, a machine learning approach is used to cluster shares into groups using the financial indicators and is studied how much the structure of the clusters matches the sectoral structure of the companies and how miscellaneous are the companies from a financial perspective.

Keywords: fundamental analysis, machine learning, share price

JEL classification: G10

1. Introduction

An explanation and prediction of prices of shares have been a widely discussed and researched topic for many years. Technical, fundamental, and behavioural analysis are the main approaches to explaining share market movements (Bustos & Pomares-Quimbaya, 2020). This paper uses the fundamental analysis as the scientific approach and focuses on the description of the relationship between financial indicators and share price, using data from one hundred largest companies traded on the New York Stock Exchange (NYSE) and the National Association of Securities Dealers Automated Quotations (NASDAQ). The main goal of the article is to determine, whether a stock price is affected by the values of financial indicators of the company, to identify indicators whose effect is statistically significant and to explore differences among industries. The paper aims to provide information on which financial indicators should a potential investor focused on in his/her investor decision.

The study of financial indicators has a long-term tradition, as the first paper has been published in 1957 (Collins, 1957). Since then, many other papers have been published, as can be seen in the literature review. Despite extensive research efforts in this area, there is no recent paper focused on the NYSE and NASDAQ. As these stock exchanges are among the most important ones in the world, this paper aims to fill this gap in the current research and study the effect of financial indicators on the prices of the 100 largest companies traded on the NYSE and NASDAQ.

2. Literature Review

As a first step of the literature review, the keywords were determined. As the main point of interest are financial indicators and their effect on share prices, keywords “financial indicators” and “share price” has been identified and searched for in titles, abstracts and keywords. Another keyword “stock

* Corresponding author’s email: pesikj@kpm.zcu.cz

price", which is commonly used with the identical meaning, has been added, with an "OR" relationship with "share price". Articles Web of Science and Scopus have been used as main sources. The found articles were evaluated based on their abstracts and a set of articles has been created to represent various stock markets and approaches. The results are presented in Table 1.

It can be seen that regression is the dominant approach in the analysis of the relationship between financial indicators and share prices, followed by neural networks. Cointegration, Vector Error Correction Model (VECM), Confirmatory Factor Analysis (CFA), Variance Inflation Factor (VIF), Principal Component Factor Analysis (PCFA). Panel regression is the most common regression model.

Table 1 List of papers analysing the relationship between financial indicators and share price

Years included	Country, companies	Method	Factor influential to the share price	Paper
1995-2004	Greece, 101	Regression	7 ratio indicators identified as influential	(Dimitropoulos & Asteriou, 2009)
2004-2013	20 emerging countries	PCFA, regression	Four influential factors identified by PCFA, plus size of the company and macro factors	(Takamatsu & Lopes Fávero, 2019)
2006-2015	Poland, 32	Cointegration, VECM, regression	Relationships differ among industries; rentability, liquidity and FL are mostly influential	(Ligocká, 2019)
2006-2018	Lithuania, 4	Regression	The relevance of financial indicators differs among companies	(Roldugin & Roldugin, 2018)
2010-2018	Jordan, 57	VIF, regression	7 indicators identified as influential	(Abdallah et al., 2022)
2011-2018	Indonesia, 24	CFA	Dividend policy, profitability and solvability are influential	(Sholichah et al., 2021)
2012-2020	China, Ctrip.com	Neural network	Asset growth rate, TAT, and interest coverage are the most influential	(Chen et al., 2021)
2013-2018	V4, 55	Regression	Only total equity influential in all V4 countries	(Aliu et al., 2021)
2014-2019	India, 12	Regression	P/E and EPS are influential, prices are generally unstable	(Sampathkumar et al., 2021)
2016-2020	China, 200	Regression	EPS, CF per share, profit per share and intrinsic value are influential	(Xu, 2021)
not given	China	Neural network	Net profit rate growth, ROE and net CF are the most influential	(Gao et al., 2022)

The papers imply there is a relationship between share price and financial indicators, however, different indicators seem to be influential in different markets. The review encourages studying the relationship between price and financial indicators for various markets separately, as this relationship tends to be different for different markets.

3. Data and Methodology

The data used for this paper are provided by the studied companies in their annual reports. These data were gained through a paid service Stock Analysis On Net (Stock Analysis on Net, 2022), which aggregates the publicly available data as Excel files, including the financial indicators values. The sector of each company was determined using Yahoo Finance (Yahoo, 2022). The statistical approach is primarily based on (Cipra, 2008).

As can be seen in the literature review, there is a vast number of financial indicators that need to be taken into consideration. As a panel regression (and regression in general) requires independent variables to be uncorrelated, a two-step approach has been applied. In the first step, a smaller sample of 15 companies has been selected and a correlation matrix has been created for all 30 financial indicators available at the service. A representative indicator has been selected from each group of highly correlated indicators, using correlation with the share price and the criteria. Six representative indicators were selected, and their definitions are as follows.

The first indicator is the current ratio (CR), which can be calculated using the formula:

$$CR = \frac{CA}{CL} \tag{1}$$

where CA represents current assets and CL represents current liabilities.

The second indicator is debt to assets (DA), which can be calculated using the formula:

$$DA = \frac{TD}{TA} \tag{2}$$

where TD represents total debt and TA represents total assets.

The third indicator is financial leverage (FL), which can be calculated using the formula:

$$FL = \frac{TA}{SE} \tag{3}$$

where SE represents shareholders' equity.

The fourth indicator is operating profit margin (OPM), which can be calculated using the formula:

$$OPM = 100\frac{OI}{R} \tag{4}$$

where OI represents operating income and R revenue.

The fifth indicator is receivables turnover (RT), which can be calculated using the formula:

$$RT = \frac{NS}{AC} \tag{5}$$

where NS represents net sales and AC accounts receivable (net of allowances).

The last indicator is total assets turnover (TAT), which can be calculated using the formula:

$$TAT = \frac{NS}{TA}. \tag{6}$$

Most of the cited articles use the price as a dependent variable. However, two arguments against the price were identified: the price does not affect a book value of a share and the dataset of prices of analysed companies is not normally distributed and could not be transformed to the normal distribution using the transformation described below. For these reasons, a price-to-book ratio PB was selected, as it reflects both market and book value of a share and can be calculated using the formula:

$$PB = \frac{P}{BVPS}. \tag{7}$$

where P is the share price, more specifically closing price on the day when the financial report was published, and $BVPS$ is book value per share, which is defined as a total shareholder's equity divided by several shares of common stock outstanding.

3.1. Panel Regression Model

Data pre-processing was done using Python modules Pandas, SciPy and statsmodels. During the data processing, the following steps were taken. Outliers were removed using the Interquartile Range (IQR). Normality assumption has been tested for data without outliers using the D'Angostino normality test (D'Agostino, 1971). As the normality hypothesis was rejected for some indicators, data were transformed to a normal distribution using the Yeo-Johnson transformation (Yeo & Johnson, 2000).

Hausman test of endogeneity was performed as a next step to determine, whether a fixed-effects or random-effects model should be used. The fixed-effect model determines individual effects of unobserved, independent variables as constant over time, whereas the random-effect model determines individual effects of unobserved, independent variables as random variables over time. The p-value of the test was smaller than 0.001 which means the fixed-effect model should be used.

For the given dataset, the regression model can be defined by the formula

$$PB_{i,t} = \alpha_j + \gamma_1 CR_{j,t} + \gamma_2 DA_{j,t} + \gamma_3 FL_{j,t} + \gamma_4 OPM_{j,t} + \gamma_5 RT_{j,t} + \gamma_6 TAT_{j,t} + \varepsilon_{j,t} \tag{8}$$

where $\varepsilon_{j,t} \sim iid(0, \sigma^2)$.

The model overall was proved to be statistically significant, with the p-value of the F-test less than 0.001, however, the value of the coefficient of the determination is not very high (0.35). The p-value of the t-test of significance was below 0.05 and they can be considered statistically significant: γ_3, γ_5 and γ_6. Consequently, financial leverage, receivables turnover and total assets turnover were proved to affect the price of the share, whereas current ratio, debt to assets and operating profit margin were not.

The low value of the coefficient of determination may be caused by different effects of financial indicators in various industries, as can be found in (Ligocká, 2019). To investigate this possibility, the healthcare sector has been selected as a sample sector, as it contains the biggest number of companies. The model for 20 companies was also determined as statistically significant, with a notably higher value of the coefficient of determination of 0.71. Furthermore, debt to assets and financial leverage were determined as statistically significant using a 0.05 level of significance.

A classical regression model was created as a next step to investigate the differences among industries. The information about the sector was added to the model using one-hot encoding, omitting industries with fewer than 5 companies. Out of 7 industry coefficients, 5 were evaluated as statistically significant. More specifically, communication services and energy have negative coefficients and consumer defensive, healthcare and technology have positive coefficients, i.e., the companies in the communication services, and energy sector would be valued lower than companies with the same financial indicators, but doing business in the consumer defensive, healthcare, and technology sectors.

3.2. Cluster model

The main outcome of the regression model is a difference among the evaluated industries. The purpose of the second part of the paper is to evaluate the homogeneity of companies inside the individual industries. The cluster model splits data samples into homogenous groups. The cluster method is a non-supervised machine learning method, i.e., it requires no "correct" solution. K-means algorithm was selected for the model. This method creates a set of points across the value domain, each representing the centre of each group. Each point is classified into a group based on the closest centroid.

The cluster model has been created using the Python module scikit-learn. Two models were constructed: a model with reduced dimensionality and a non-reduced dimensionality model. Data was transformed for both models using a standard scaler, which standardized features by removing the mean and scaling to unit variance. For the first model t-distributed stochastic neighbour embedding (TSNE) method was used to reduce the dimensionality of data to 2 (van der Maaten & Hinton, 2008). An optimal combination of perplexity and number of clusters was determined as 6 and 10, using the silhouette index as a criteria function. The results can be seen in Figure 1. The silhouette index is 0.5267.

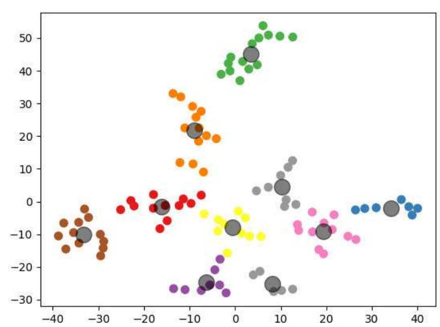

Figure 1 Result of the cluster model

As can be seen in Table 2, companies from most industries are spread across multiple clusters, which indicates the analysed industries consists of diversified companies from a financial perspective. Three industries are not presented in the table as there were fewer than 5 companies present in the data sample.

Table 2 Count of companies in individual clusters

Sector/Cluster number	0	1	2	3	4	5	6	7	8	9
Communication Services					1		3	1	2	3
Consumer Cyclical		3	3	1	1	1	1	2		1
Consumer Defensive		1	4					1	1	3

Energy	1		1		2	1		1		
Healthcare	4		3	2		4	3	1	2	1
Industrials		2	3	2	3	1		3		
Technology	6	1		1	4	2	5	1		

Similar results were provided by the model without data reduction. Most notable differences were observed for the communication services model (which is split into 3 groups instead of 5) and the consumer defensive sector (which is split into 6 groups instead of 5).

4. Discussion

It has been proved by the panel regression model that companies operating in different industries are not equally valuated on NYSE and NASDAQ. Financial indicators do affect the valuation of the company; however, their effect should be considered separately for each sector on NYSE and NASDAQ. A consumer defensive, healthcare and technology sectors company would be valuated higher on the market than a company from the communication services and energy sector with the same financial indicator values. A potential investor should keep in mind that investor sentiment may change quickly as the world faces unprecedented challenges. A subject of further research could examine whether this relationship would change when an energy crisis may be a result of the Ukrainian war. Another possibility could be to explore a longer time series to explore whether the investor sentiment against various sectors changed over time.

The cluster model showed the sectors are heterogenous from the financial perspective and groups of companies with similar financial indicators values consist of representants of various sectors. A subject of another analysis could be to determine a relationship among companies in each group, using both financial and non-financial perspectives. More financial indicators may be used in the cluster model, as the TSNE method provides good results in reducing multicollinearity (van der Maaten & Hinton, 2008).

References

Abdallah, A., Afifa, M. A., Saleh, I. H., & Alsufy, F. (2022). Determinants of Market Stock Price: New Evidence from an Emerging Market. *Information Sciences Letters*, *11*(2), 549–558. https://doi.org/10.18576/isl/110223

Aliu, F., Nadirov, O., & Nuhiu, A. (2021). Elements indicating stock price movements: The case of the companies listed on the v4 stock exchanges. *Journal of Business Economics and Management*, *22*(2), 503–517. https://doi.org/10.3846/jbem.2021.14181

Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A Systematic review. *Expert Systems with Applications*, *156*. https://doi.org/10.1016/j.eswa.2020.113464

Chen, L., Yu, J., & Zheng, Y. (2021). Analysis and Research on Internal Influencing Factors of Stock Volatility of Chinese Listed Companies Based on CVM-RBF Neural Network Algorithm. *2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT)*, 15–20. https://doi.org/10.1109/ISCIPT53667.2021.00011

Cipra, T. (2008). *Finanční ekonometrie* (Vol. 30). Ekopress Praha, Czech Republic.

Collins, J. (1957). How to Study the Behavior of Bank Stocks. *Financial Analysts Journal*, *13*(2), 109–113. https://doi.org/10.2469/faj.v13.n2.109

D'Agostino, R. B. (1971). An Omnibus Test of Normality for Moderate and Large Size Samples. *Biometrika*, *58*(2), 341–348. http://www.jstor.org/stable/2334522

Dimitropoulos, P. E., & Asteriou, D. (2009). The value relevance of financial statements and their impact on stock prices: Evidence from Greece. *Managerial Auditing Journal*, *24*(3), 248–265. https://doi.org/10.1108/02686900910941131

Gao, Y., Yao, Y., & Li, Y. (2022). *Analysis on the influence mechanism of corporate stock price based on Lasso-CNN neural network*. 1048–1052. https://doi.org/10.1109/cisai54367.2021.00210

Ligocká, M. (2019). *The Effect of Financial Ratios on the Stock Prices: Evidence from the Polish Stock Exchange Vliv fi nančních ukazatelů na ceny akcií: aplikace na polskou burzu cenných papírů*. *13*, 44–60. www.vsfs.cz/acta

Roldugin, V., & Roldugin, A. (2018). Multiple linear regression of stock quotes of the Lithuanian enterprises. *Economic Annals-XXI*, *173*(9–10), 43–48. https://doi.org/10.21003/ea.V173-07

Sampathkumar, S., Suresh, C. K., & Umamaheswari, S. (2021). An empirical study on effect of financial accounting indicators towards stock market price volatility. *World Review of Science, Technology and Sustainable Development*, *1*(1), 1. https://doi.org/10.1504/wrstsd.2021.10036120

Sholichah, F., Asfiah, N., Ambarwati, T., Widagdo, B., Ulfa, M., & Jihadi, M. (2021). The Effects of Profitability and Solvability on Stock Prices: Empirical Evidence from Indonesia. *Journal of Asian Finance, Economics and Business*, *8*(3), 885–894. https://doi.org/10.13106/jafeb.2021.vol8.no3.0885

Stock Analysis on Net. (2022). *Stock Analysis on Net*. https://www.stock-analysis-on.net/

Takamatsu, R. T., & Lopes Fávero, L. P. (2019). Financial indicators, informational environment of emerging markets and stock returns. *RAUSP Management Journal*, *54*(3), 253–268. https://doi.org/10.1108/RAUSP-10-2018-0102

van der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*(86), 2579–2605. http://jmlr.org/papers/v9/vandermaaten08a.html

Xu, M. (2021). A Study on the Correlation between Financial Status of Listed Companies and Chinese Stock Market: Base on Multiple linear regression analysis. *Proceedings - 2021 2nd International Conference on Big Data Economy and Information Management, BDEIM 2021*, 340–343. https://doi.org/10.1109/BDEIM55082.2021.00075

Yahoo. (2022). *Yahoo Finance*. http://finance.yahoo.com/

Yeo, I.-K., & Johnson, R. A. (2000). A New Family of Power Transformations to Improve Normality or Symmetry Author (s): In-Kwon Yeo and Richard A . Johnson Published by : Oxford University Press on behalf of Biometrika Trust Stable URL : http://www.jstor.org/stable/2673623. *Biometrika*, *87*(4), 954–959.