Relationship Between Stock Prices and Common Financial Metrics*

Analysis of the Relationship Between Stock Price and EPS & Dividends of the Largest 50 North American Companies over 10 Years

Diana Liu

April 6, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

2 Data

Public companies whose stock are traded on exchanges, must disclose certain financial metrics in their quarterly and yearly financial statements. Investors then use these metrics to help inform their decision when selecting companies to invest in. Some of the most important metrics investors look at are Earnings Per Share, Net Income, and Dividends (Shakespeare 2020). Investments take the form of buying shares or stocks which allows the investor to own a fraction of the company.

The data used for analysis is obtained through Walton Research Data Services (WRDS) by the University of Pennsylvania. WRDS provides access to Compustat, a database of financial, statistical, and market information on global companies since 1962. Our data comes from Compustat's North America Fundamentals Annual database, which contains both financial data from all public North American companies, collected by Compustat from each company's annual financial statements or from stock exchanges (Figure 1). From this data set, we extract the price, EPS, net income, dividends, TIC, and Year of the top 50 largest companies by market capitalization, a measure of the size of the company in terms of total value, for the past 10 years (2013-2023). This Data is cleaned and analysed in R Core Team (2022) with assistance from Wickham et al. (2019), Richardson et al. (2024), Arel-Bundock (2022), Gabry and Mahr

*Code and data are available at: https://github.com/Diana-Guanzhi-Liu/Relationship-Between-Stock-Prices-and-Common-Financial-Metrics

(2024), David Robinson (2023), Auguie (2015), Goodrich et al. (2024), Xie (2023), Iannone Richard (2024), and Team and worldwide (2024).

Similar data sets exist, but WRDS was selected because it hosts the most robust collection in terms of variables and its infrastructure allows users to easily query specific items.

Year	Tic	EPS	Dividends	Net_Income	Price	Paid_Dividend
2014	ABT	1.49	1363	2284	45.02	Yes
2015	ABT	2.92	1464	4423	44.91	Yes
2016	ABT	0.94	1547	1400	38.41	Yes
2017	ABT	0.27	1947	477	57.07	Yes
2018	ABT	1.33	2047	2368	72.33	Yes
2019	ABT	2.06	2343	3687	86.86	Yes

Figure 1: Data from Compustat North America Fundamentals Annual. EPS, Dividends, and Net Income are financial statement data while Price are stock exchange data, and Year and Tic can be found on both.

2.1 Financial Statement Data

Net income, EPS, and dividends are items that are found on financial statements. Specifically, EPS and net income can be found on the income statement while dividends can be found on the statement of retained earnings. This data is likely unbiased and free from error as financial statements are required by law to be audited (verified by an independent third party).

Net income is used to measure profitability or how much income the company keeps after expenses are paid. Higher net income means that the company is profitable by either earning more revenues or reducing expenses. Net income is expected to be positively correlated with price as investors want to invest in companies that are more profitable. Net income can be negative when the company has lost money. In our data set, it is measured in millions USD (x10^6). Net income is normally distributed with median of 7120M USD and right skew, mean of 11752 (Figure 3). Apple (AAPL) had the highest net income in the data set of 99803 achieved in 2022, and Berkshire Hathaway (BRK.B) had the lowest of -22819 in the same year (Table 1).

EPS is a commonly used measure of a given company's value in USD; it is calculated net income divided by the number of shares. It is normally distributed and right skewed, with median 4.56 USD and mean 6.2 (Figure 3). Google (GOOGL) had the highest EPS of 112.2 in 2021, and Tesla (TSLA) had the lowest of -11.83 in 2017, meaning for each share Tesla lost 11.83 USD (Table 1).

Dividends are distributions of a company's income to the owners of its stock in USD. Some companies choose not to pay dividends if they can use their income to reinvest into valuable

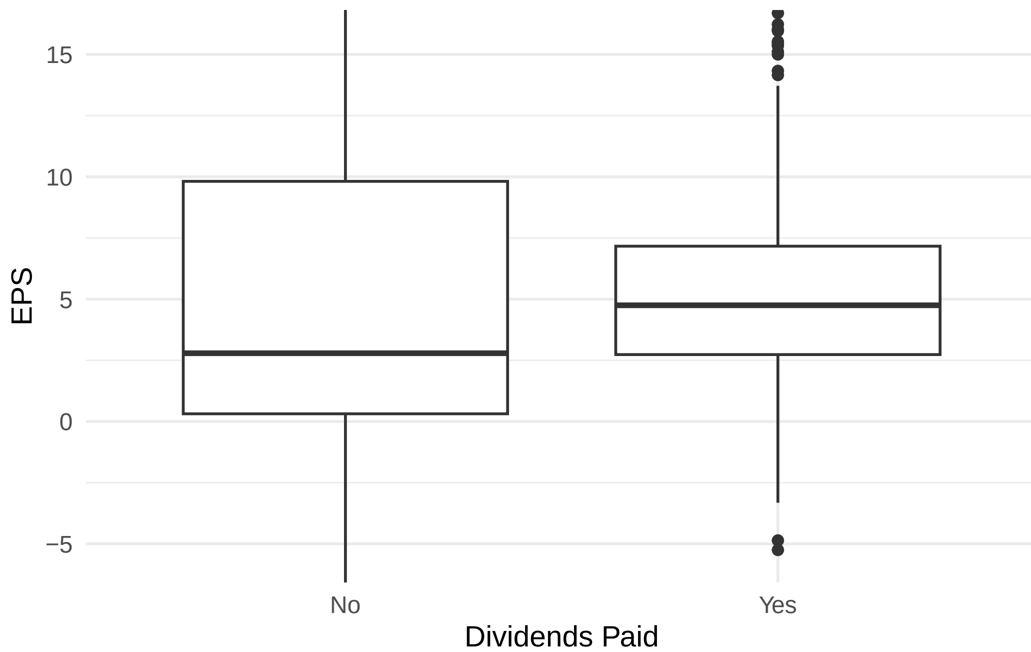


Figure 2: Boxplots of EPS grouped by dividends

projects. For our analysis we created a binary variable for whether or not a company paid dividends, with any amount greater than zero dollars being Yes. Approximately 78% of companies in our data set paid dividends, the largest of which was 25999 USD paid by Caterpillar (CAT) in 2023.

EPS of companies that did not pay dividends has a larger interquartile range than EPS of companies that paid dividends (Figure 2). This means that the observations for companies that paid dividends are more clustered together, likely resulting in smaller standard deviation. The mean and median EPS of companies that paid dividends is also higher, suggesting that paying dividends is correlated with having higher EPS (Figure 2), this makes sense as companies that are profitable enough to have high EPS also have more income to pay out dividends.

2.2 Stock Exchange Data

The price is the amount that each share costs to buy, it is determined by supply and demand for a company. For example, if investors believe a company is profitable, they will buy shares with the goal of eventually sharing in said profit, increasing demand and driving up share price. Share price is normally distributed with median of 113.91 USD and mean of 6.2. The most expensive stock was Amazon (AMZN) which was worth 3334.34 USD in 2021, and the cheapest stock was Advanced Micro Devices, Inc. (AMD) which was worth 2.67 in 2014.

Tic refers to the ticker symbol of each stock. It is a unique string of one to five characters used to denote a particular company's stock so that investors know whose stock are being traded on international exchanges across different languages. Some tick symbols are similar to the name of the company like AAPL for Apple and GOOGL for Google, but this is not always the case, for example American Express is AXP and Coca-cola is KO.

Table 1: Summary statistics of Price, EPS, Net Income, and Dividends of the 50 largest North American Companies by market cap

summary_stats	earnings_per_share	net_income	dividends	price
Min	-11.80	-22819.00	0.00	2.67
1st Quartile	2.30	3003.10	353.50	62.60
Median	4.56	7120.00	2972.00	113.91
Mean	6.20	11752.12	4029.56	192.22
3rd Quartile	7.40	14728.00	6249.50	190.92
Max	112.00	99803.00	25999.00	3334.30

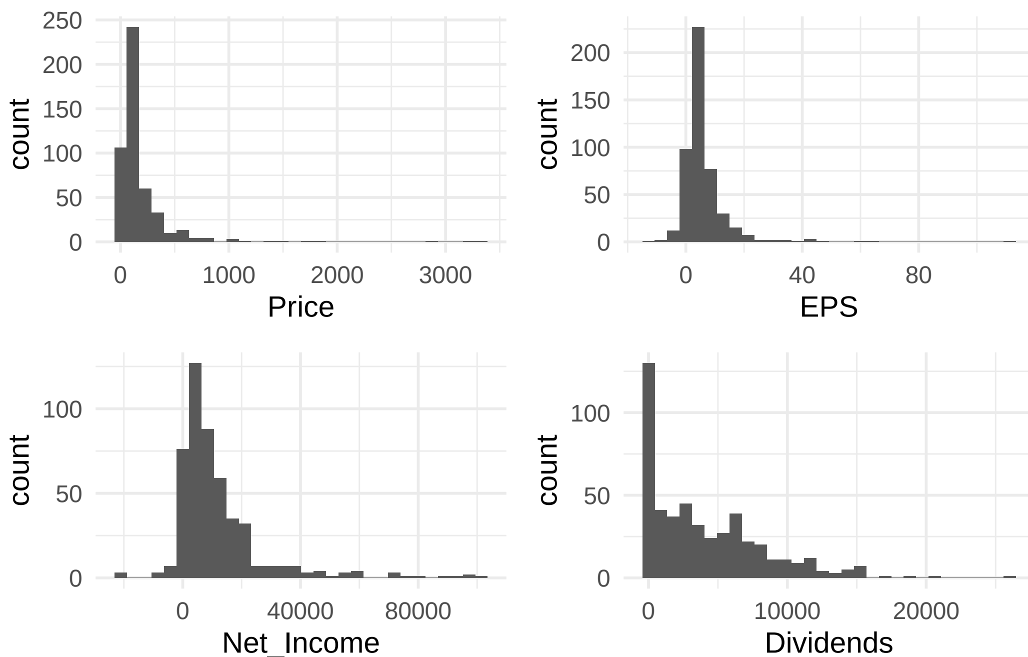


Figure 3: Price, EPS, Net_Income, and Dividends are all normally distributed with right skew

3 Model

The goal of our model is to determine the relationship between EPS and stock price first with a simple linear model. Then we add a binary variable of whether the company paid dividends or not to determine the effect dividends have on this relationship. Background details and diagnostics are included in Appendix B.

3.1 Model set-up

In the first model, we define y_i as the stock price, X_1 as earnings per share, a continuous variable. β_0 is the intercept term, and β_1 is the coefficient for change in the stock price per dollar increase in EPS. ϵ denotes the noise or deviations.

$$\begin{aligned} &\textit{Simple Model} && (1)\\ &Y_i = \beta_0 + \beta_1 X_1 + \epsilon && (2)\\ & && (3) \end{aligned}$$

In the second model, we add X_2 as a binary variable where yes signifies that a company paid dividends, and β_2 is the coefficient for change in stock price if dividends were paid.

$$\begin{aligned} &\textit{Multivariable Model} && (4)\\ &Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon && (5)\\ & && (6) \end{aligned}$$

3.1.1 Model justification

We expect a positive linear relationship between EPS and stock price. This is because a higher EPS indicates greater value and investors should be willing to pay more for a company's shares if they think the company has higher profits relative to its share price. If a company has high EPS, they have greater amount of profit after expenses have been paid. Investors want to own profitable companies, so demand for this company's stock increases, driving up price as the number of stocks that are available to be bought remains the same. The opposite is true for unprofitable companies, if EPS is low or negative, investors may choose to sell off their shares even at a loss in order to prevent further losses should the stock price keep falling (Figure 4).

One might expect dividends to be positively correlated with stock price because investors desire an additional payment, but the opposite is true. This is because dividends paid out to shareholders is value that is taken out of the company, causing price, a reflection of the firm's value, to decrease (Figure 4).

We expect the effects of EPS and dividends to interact with each other, creating two potential outcomes for change in price. Companies that did not pay dividends are expected to have a larger increase in price for every additional dollar of EPS than companies that paid dividends.

4 Results

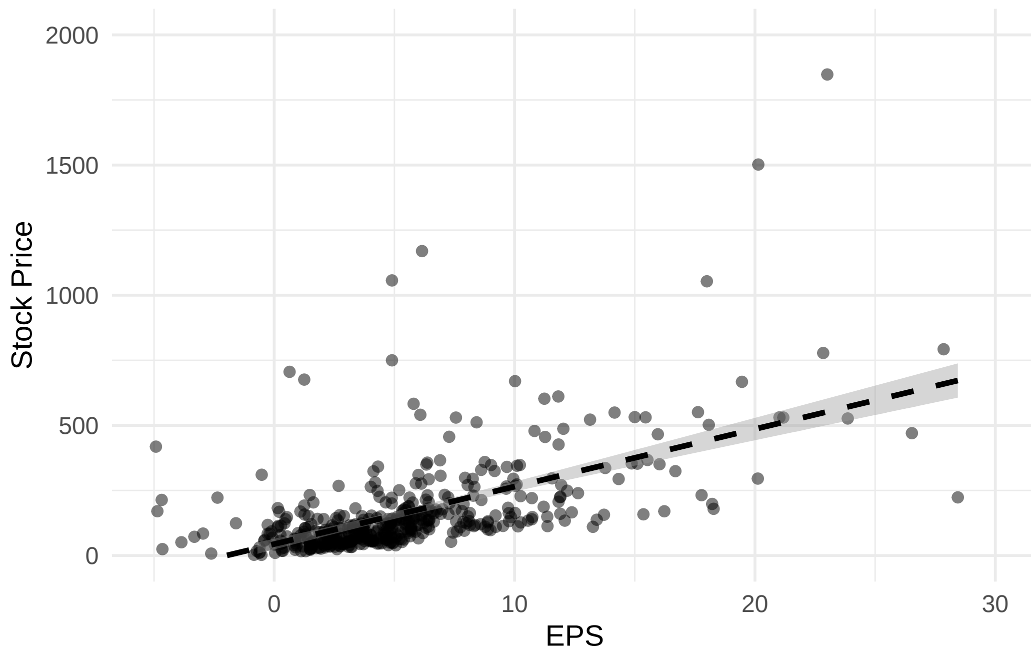


Figure 4: Relationship between variables EPS and stock price

In model 1, only the relationship between the dependent variable price and predictor variable EPS is analyzed. In the first column of Table 2, we have a intercept of 27.25 and with standard error 11.47. The coefficient term β_1 is 26.59 with standard error 1.05 which shows the increase in price for every dollar increase in EPS that can also be observed from Figure 4 as the dashed line with standard error as the shaded area around the dashed line. Positive slope and coefficient show a positive relationship between EPS and stock price. While most observations follow the linear relationship, there is some clustering between 0 to 10.

In model 2, we introduce dividends as an additional predictor variable. Paying dividends has a negative relationship with price as show in Figure 5 where stocks that paid dividends have median price around 100, significantly lower than stocks that paid no dividends with

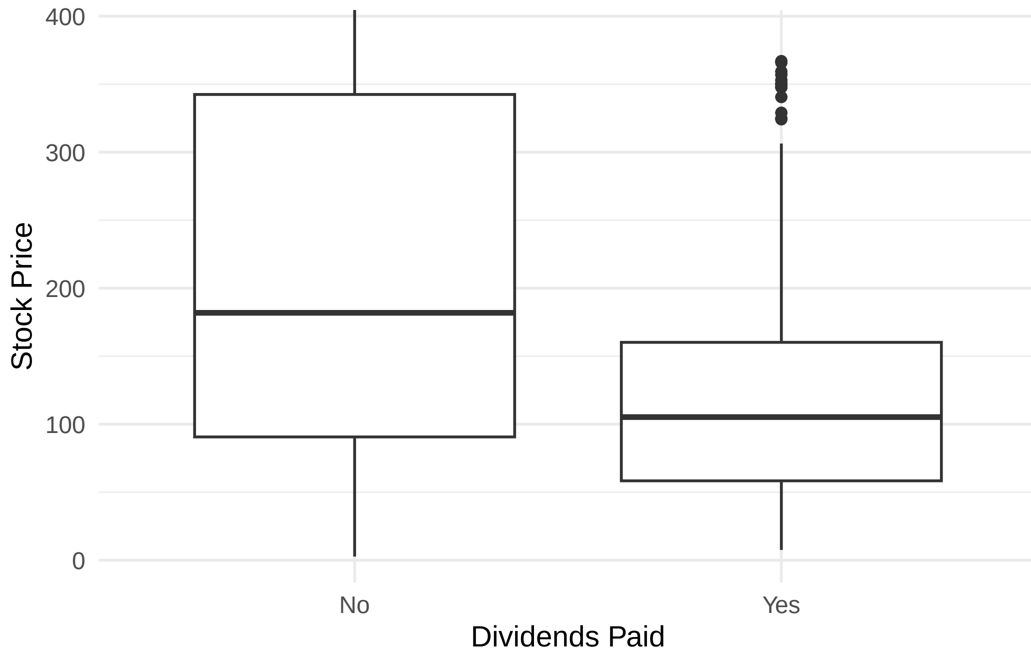


Figure 5: Boxplots of stock price grouped by whether or not dividends were paid.

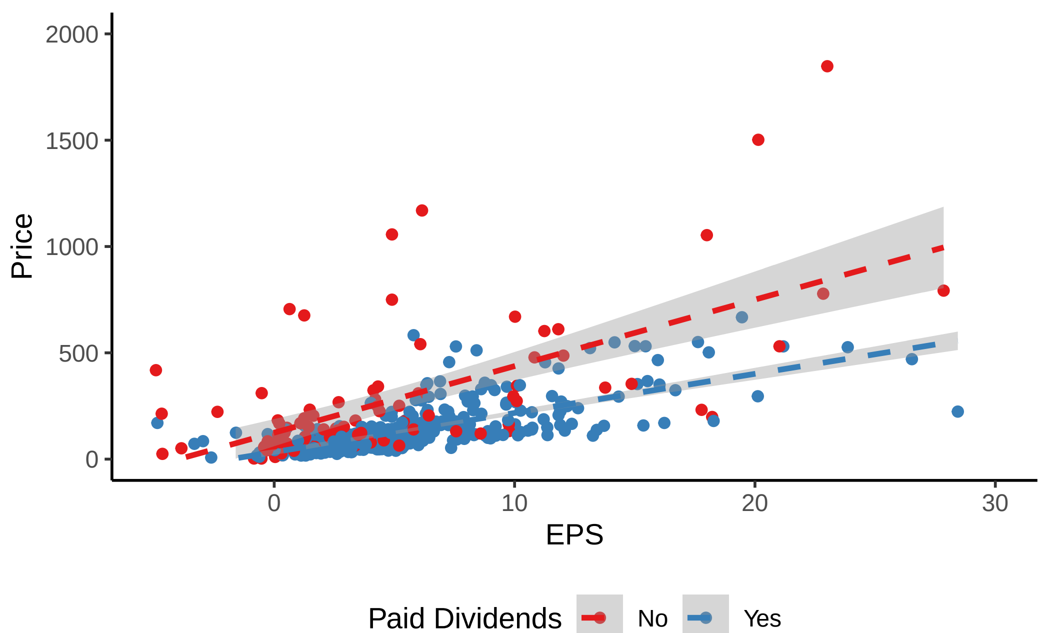


Figure 6: Relationship between EPS, dividends, and stock price.

Table 2: Effect of earnings per share and dividends on stock price.

	EPS Only	With Dividends
(Intercept)	27.25	172.05
	(11.47)	
EPS	26.59	25.59
	(1.05)	
Paid_DividendYes		−177.85
Num.Obs.	483	483
R2	0.569	0.622
R2 Adj.	0.568	0.588
AIC	6526.5	
BIC	6539.0	
Log.Lik.	−3260.248	−3228.568
F	635.550	
ELPD		−3248.2
ELPD s.e.		77.6
LOOIC		6496.5
LOOIC s.e.		155.2
WAIC		6499.6
RMSE	206.66	193.14

median price around 175. Stocks that did not pay dividends also have a significantly larger interquartile range for price of around 250 USD than stocks that did of 100.

The second column of Table 2 shows that the new intercept is 172.05 USD, and that the coefficient of paying dividends β_2 is -177.85. With the addition of the dividends variable, EPS of companies that paid dividends has changed to a coefficient of 25.59 (Table 2). The effects on price from EPS and dividends can be further observed in Figure 6 where the observations and line of best fit are colored blue to indicate that dividends were paid and red if not. EPS without dividends (red line) has very similar slope and coefficient as Figure 4. EPS with dividends (blue line) has a less steep slope and is located below the EPS without dividends.

R^2 and adjusted R^2 are terms that explains how well the model is able to explain data, with R^2 for single variable models and the more conservative adjusted R^2 for multiple variables. Model 1 has a R^2 of 0.57, meaning that 57% of the variability observed in stock price is explained by EPS in the regression model. Model 2 can explain 59% of stock price with EPS and dividends. Both indicate that the model can explain stock price reasonably well, but there is around 40% of variability is attributed to other causes.

5 Discussion

5.1 Companies with Higher Earnings per Share are More Attractive to Investors

From model 1, we obtain a coefficient of 26.59 and standard error of 1.05 for EPS, indicating that for every additional dollar increase in EPS, stock price increases by 26.599 USD (Table 2). The standard error term means that on average the relationship between EPS and stock price differs from the actual relationship by 1.05 USD. The standard error is relatively small, just over a dollar, so model 1 is reasonably accurate. This confirms that there is a positive linear relationship between EPS and stock price. When a company is profitable relative to the amount of shares they have, each shareholder is entitle to a larger portion of their profit, making the company a more attractive investment than competitors that have lower EPS. This demand for shares of the company then drives up stock prices. Alternatively, a lower EPS indicating lower profitability or losses can cause existing shareholders to sell their shares, creating downward pressure on price.

This conclusion has several implications for investors and management of companies alike. For investors, it may be useful to use the causal relationship between EPS and stock price to predict future stock prices. An investor can use available net income information and number of shares from any company to calculate EPS before the information is released to the public. If the estimated EPS is high, the investor can buy up shares at a lower price before the increase in demand and sell those shares once demand increases to make a profit.

Managers of a company who have an incentive for their company to be profitable in the long term would want their share prices to exhibit stable growth over time instead of the short term fluctuations caused by changing demand. If managers anticipates that one year's EPS will be low, the company can use excess income to buy up shares once the price decreases. This alleviates some of the downward pressure on price and creates more stability. The company can then either hold onto the purchased share until a subsequent year when the price rebounds to sell them, or eliminate those shares from the market. Fewer shares means that the denominator of the EPS equation is smaller, causing subsequent EPS to be higher in the future.

5.2 Paying Dividends Causes Stock Prices to Fall

From Figure 5 there is a negative relationship between paying dividends and stock price. When a company chooses to pay dividends, on average their stock price is expected to decrease by about 177.85 USD (Table 2). This is consistent with contemporary literature on dividend policy which suggests that paying dividends causes the stock price to drop by the amount of the dividend per share plus the opportunity cost of lost potential profits the dividend could have been used to generate (Narinderp and Tandon (2019)). Dividends paid out represents value that is extracted from the company and not used to invest into profitable projects. All things equal, this indicates that the company is expected to grow less than competitors who did not pay dividends and reinvested, making the stock less attractive to potential investors who care about growth.

EPS coefficient of companies that paid dividends is 25.59 which is one dollar less than that of model one where dividends were not considered (Table 2). This means that the amount price is expected to increase per dollar EPS increases is smaller, likely due to !! why lol?«<

Managers are faced with the decision of what to do with extra profit. If paid out in dividends, the stock price will fall but existing shareholders, especially large ones with 10% or more equity in the company, would want the dividend income as compensation for their investment. At the same time, paying out dividends makes the stock less attractive to future investors. By knowing the coefficient of paying dividends and coefficient of EPS increases, managers are better able to balance these competing priorities and make better decisions in regards to long-term profitability and growth of their firms.

5.3 Weaknesses and next steps

A weakness in our model is the distribution of the data where they are clustered around zero, this may indicate that a linear model is not appropriate for our observations and the conclusions between EPS, dividends and price cannot be drawn. While the residuals follow normal distribution, they are clustered together instead of randomly spaced apart, indicating that they contain some additional structure that was not captured by our analysis (Figure 7). The posterior prediction check also supports this as the actual data is far larger kurtosis than what the model predicted, meaning that the predictive power of the model is low (Figure 8a). The credibility intervals for the coefficient of paying dividends is also quite large, with probability normally distributed between about -275 and -125 USD, suggesting a relatively high level of uncertainty (Figure 9). Practically, the model worked as intended with no signs of problems from the trace plot or r-hat values (Figure 10). Further work is needed to confirm the positive linear relationship between stock price and EPS.

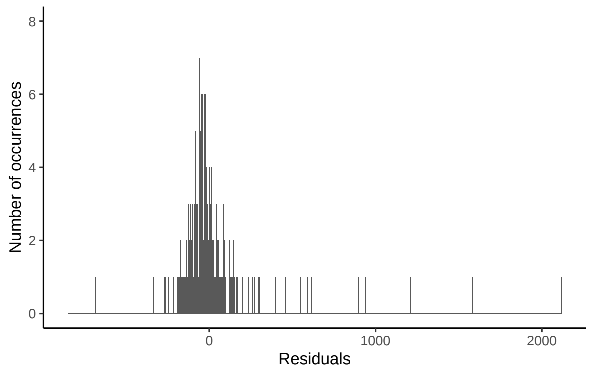
The conclusions drawn from our analysis only applies to our data set of large, North American companies from 2013-2023. More analysis with diverse data sets in terms of need to be done to determine if the same conclusions are applicable to companies of different sizes or different geographic locations. The data is also historic, meaning that the relationships described may not occur in the future. Training the model to have predictive power may provide additional useful information.

Lastly, while financial metrics are important to management and investor decision making, there are many quantitative factors like the strategic direction of a firm and its competitive advantage within the industry should also be considered. Future analysis should aim to quantify these categorical factors and determine their relationship with stock price as well as how they interact with financial metrics.

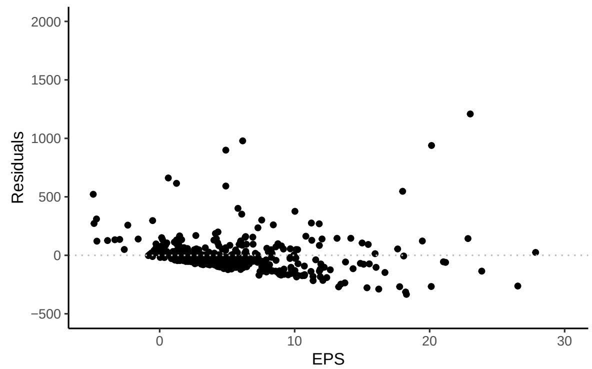
A Appendix

B Model details

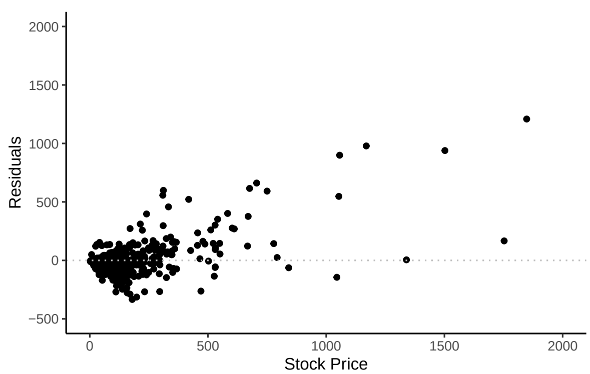
B.1 Residuals



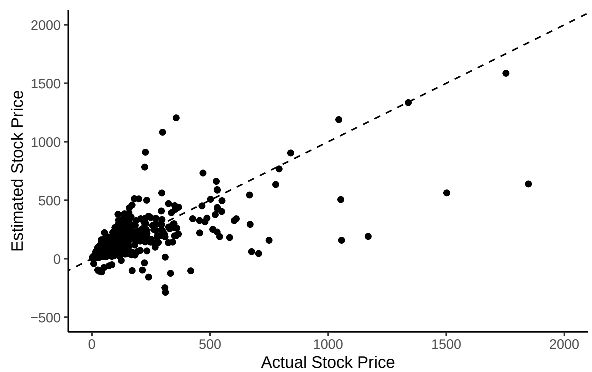
(a) Distribution of residuals



(b) Residuals by EPS



(c) Residuals by Price

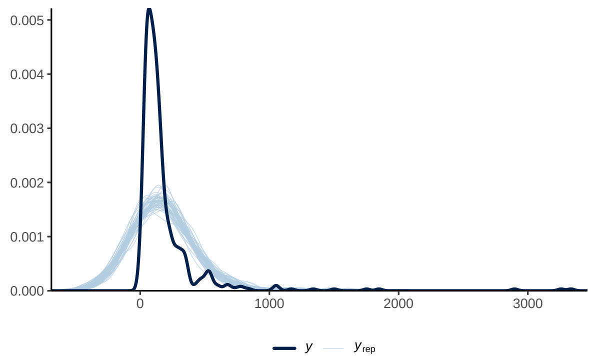


(d) Comparing the estimated price with the actual price

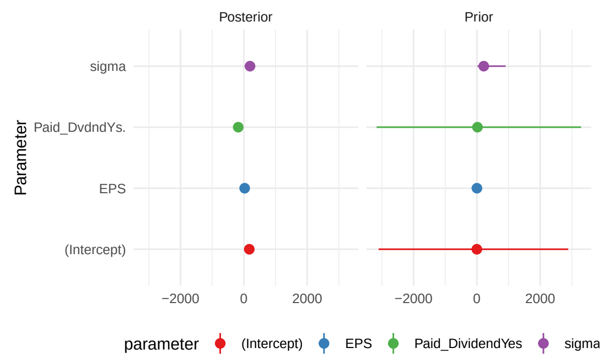
Figure 7: Residuals of model 1, simple linear regression

Figure 7a shows that residuals are normally distributed around zero which is a good sign that our model is able to capture the main patterns in stock price data with our predictor variables. This is supported by residuals scatter plot Figure 7c show that the residual values are randomly scattered around zero. However, in Figure 7b, the residuals show a negative linear pattern which indicates that our linear model may not be appropriate for the data. Lastly, in Figure 7d, the relationship between actual and estimated stock prices appears somewhat linear but with a large cluster around 0. A perfectly predictive model will have a linear slope of one, meaning that our model may lack predictive power.

B.2 Posterior predictive check



(a) Posterior prediction check



(b) Comparing the posterior with the prior

Figure 8: Examining how model 2 fits the data, and how it is affected by the data

Figure 8a compares actual dependent variable outcomes with distributions from the posterior simulation. We can see that the general shape and distribution are similar but actual data has far higher kurtosis than the simulations.

Figure 8b compares priors with posteriors to see how much the estimates change once data are taken into account. Our priors and posteriors are quite similar,

B.3 Credible Intervals

Figure 9 shows the distributions of the 95% credibility intervals of coefficients of the predictor variables of model 2. The Bayesian estimation provides distributions of each coefficient. The actual value has a 95% probability of being within the interval with the distribution showing the likelihood within the interval. Ones that have a wider credibility interval are more spread out like dividends and intercept means that the true coefficient is within a larger range.

B.4 Diagnostics

Markov chain Monte Carlo (MCMC) to was used in the model to obtain samples from the posterior distributions of interest. Figure 10a shows bouncing horizontal lines with overlap, and Figure 10b shows R-hat close to one. Both indicate that there are no issues with the algorithm.

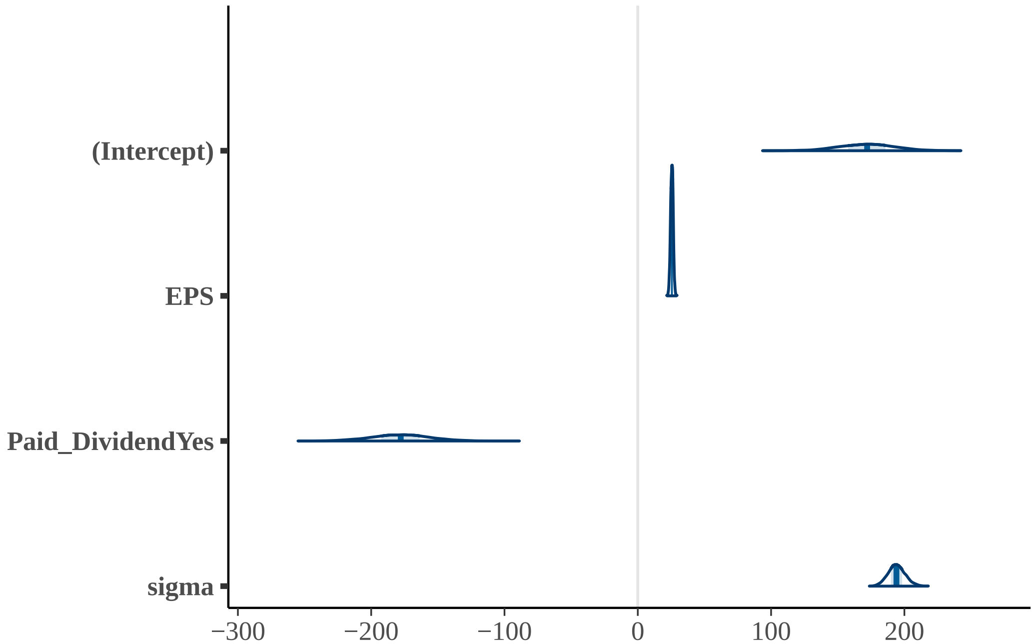
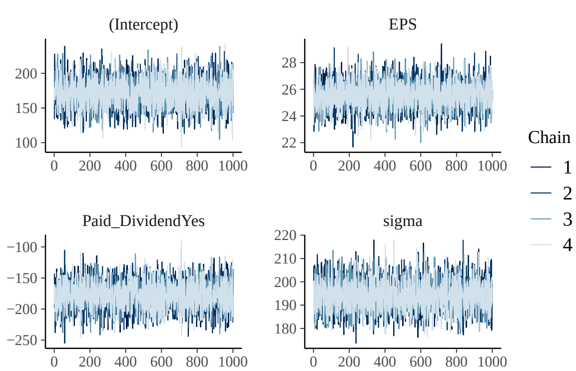
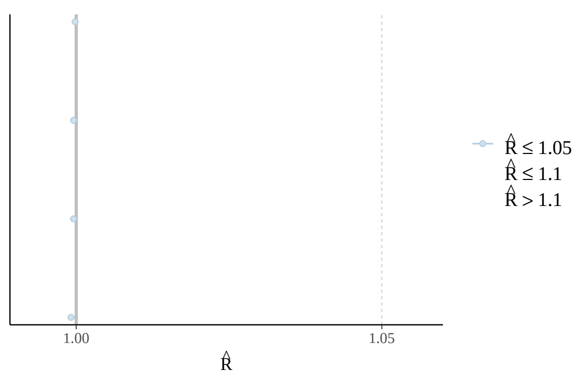


Figure 9: 95% credibility intervals of model 2 variable coefficients



(a) Trace plot



(b) Rhat plot

Figure 10: Checking the convergence of the MCMC algorithm of model 2

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Auguie, Baptiste. 2015. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. http://CRAN.R-project.org/package=gridExtra.

David Robinson, Simon Couch, Alex Hayes. 2023. "Broom: Convert Statistical Objects into Tidy Tibbles."https://broom.tidymodels.org/.

Gabry, Jonah, and Tristan Mahr. 2024. "Bayesplot: Plotting for Bayesian Models." https://mc-stan.org/bayesplot/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2024. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Iannone Richard, Schloerke Barret, Cheng Joe. 2024. "Gt: Easily Create Presentation-Ready Display Tables." https://github.com/rstudio/gt.

Narinderp, Narinder Pal Singh, and Aakarsh Tandon. 2019. "The Effect of Dividend Policy on Stock Price: Evidence from the Indian Market." *Asia-Pacific Journal of Management Research and Innovation* 15 (1). https://doi.org/10.1177/2319510X19825729.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://github.com/apache/arrow/.

Shakespeare, Catherine. 2020. "Reporting Matters: The Real Effects of Financial Reporting on Investing and Financing Decisions." *Accounting and Business Research* 20 (5). https://doi.org/10.1080/00014788.2020.1770928.

Team, R Core, and contributors worldwide. 2024. *The r Graphics Devices and Support for Colours and Fonts*. https://r-universe.dev/manuals/grDevices.html.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.