

# My title\*

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

Diana Liu

March 31, 2024

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

### Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data</b>	<b>2</b>
2.1	Financial Statement Data . . . . .	2
2.2	Stock Exchange Data . . . . .	4
<b>3</b>	<b>Model</b>	<b>5</b>
3.1	Model set-up . . . . .	5
3.1.1	Model justification . . . . .	6
<b>4</b>	<b>Results</b>	<b>6</b>
<b>5</b>	<b>Discussion</b>	<b>9</b>
5.1	First discussion point . . . . .	9
5.2	Second discussion point . . . . .	9
5.3	Third discussion point . . . . .	9
5.4	Weaknesses and next steps . . . . .	9
	<b>Appendix</b>	<b>10</b>
<b>A</b>	<b>Model details</b>	<b>10</b>
A.1	Posterior predictive check . . . . .	10
A.2	Diagnostics . . . . .	10

---

\*Code and data are available at: [LINK](#).

## 1 Introduction

You can and should cross-reference sections and sub-sections.

The remainder of this paper is structured as follows. ?@sec-data....

## 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. 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 (2024), David Robinson (2023), Auguie (2015), Goodrich et al. (2024), Xie (2023)

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.

### 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

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.

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 ( $\times 10^6$ ). 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 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.

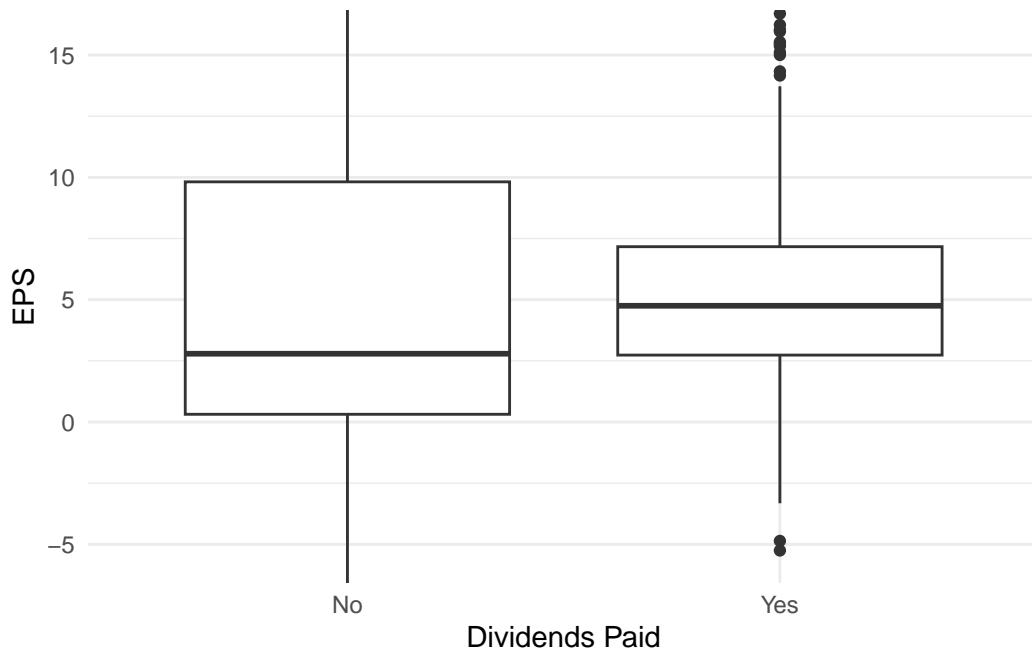


Figure 2: Boxplots of EPS grouped by dividends

## 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

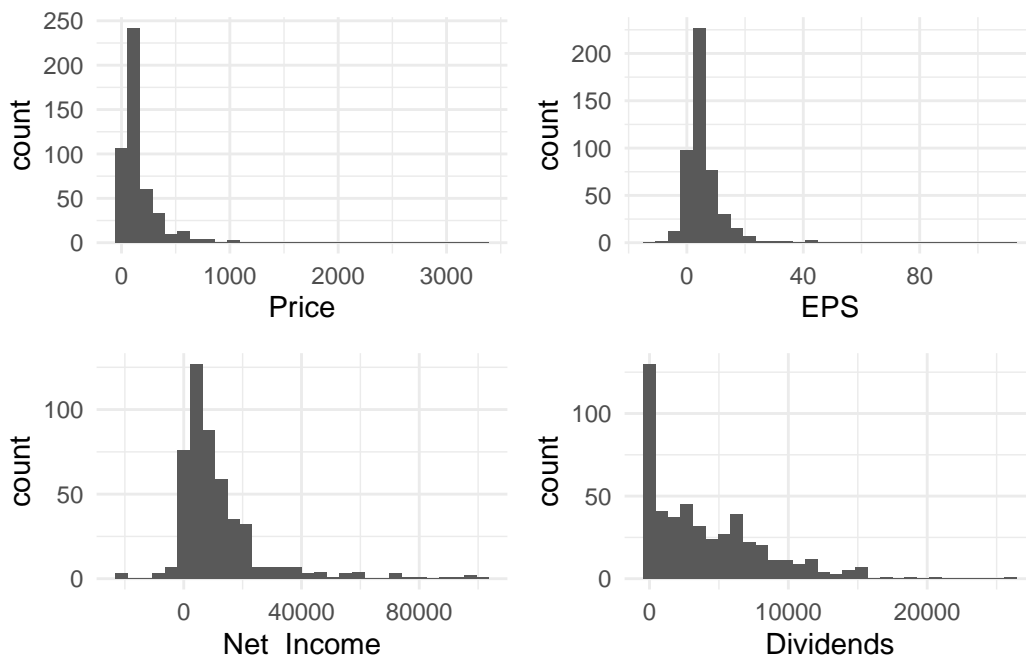


Figure 3: Price, EPS, Net\_Income, and Dividends are all normally distributed with right skew

### 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 A](#).

#### 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.  $\beta_0$  is the intercept term, and  $\beta_1$  is the coefficient for change in the stock price per dollar increase in EPS.  $\epsilon$  denotes the noise or deviations.

$$\text{Simple Model} \tag{1}$$

$$Y_i = \beta_0 + \beta_1 X_1 + \epsilon \tag{2}$$

$$\tag{3}$$

In the second model, we add  $X_2$  as a binary variable where yes signifies that a company paid dividends, and  $\beta_2$  is the coefficient for change in stock price if dividends were paid.

$$\text{Multivariable Model} \tag{4}$$

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \tag{5}$$

$$\tag{6}$$

### 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

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 -124.69 with standard deviation of 7.1 and coefficient of 21.79 with S.D. 1.1. The coefficient term  $\beta_1$  which shows the increase in price for every dollar increase in EPS can be observed from Figure 4 as the

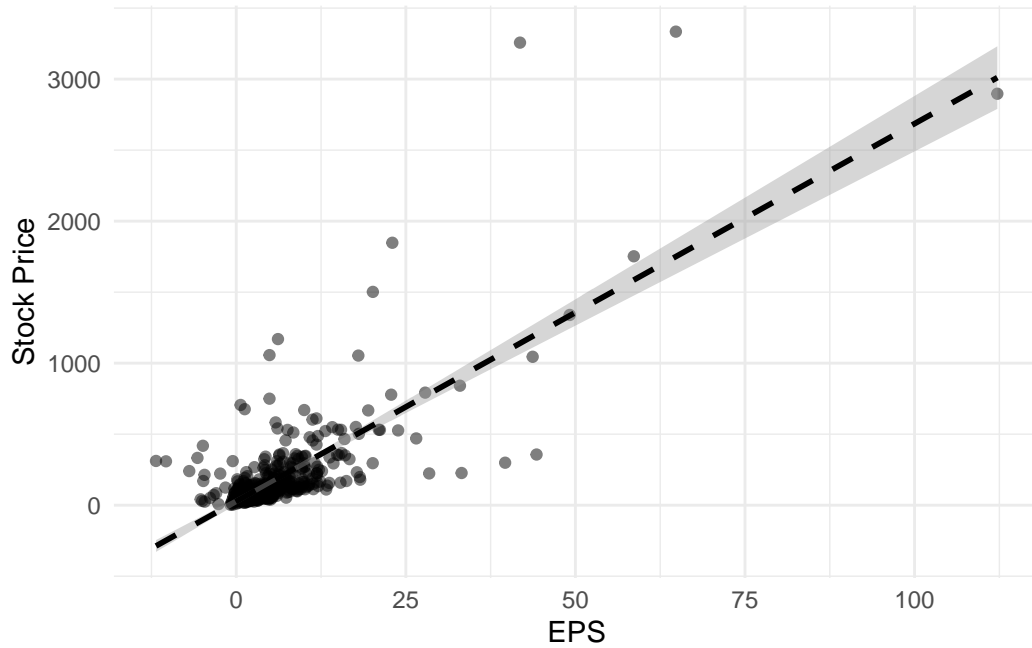


Figure 4: Relationship between variables EPS and stock price

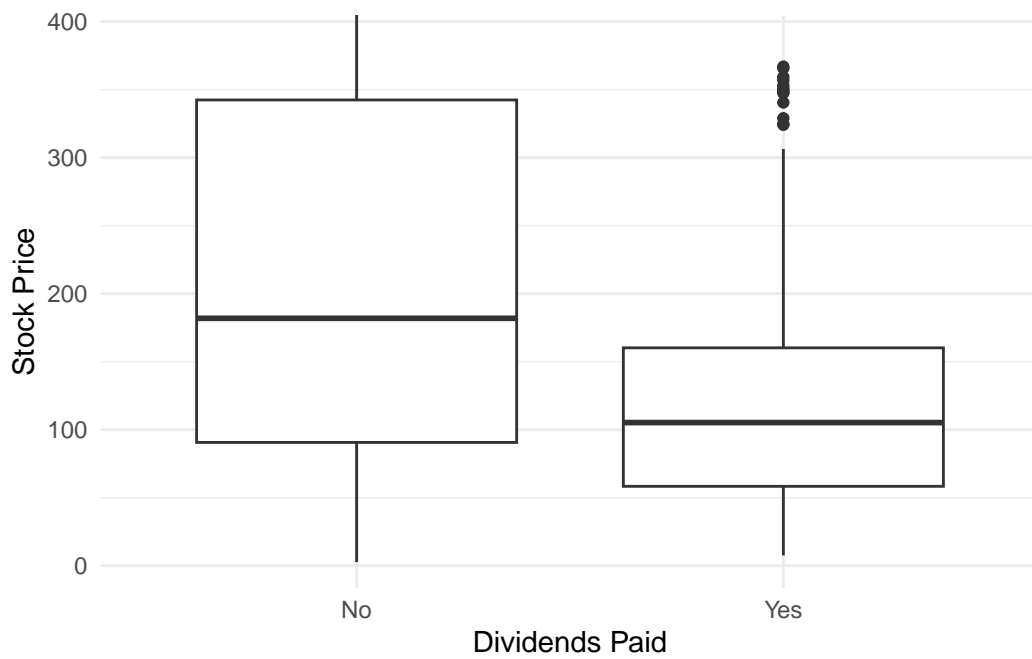


Figure 5: Boxplots of stock price grouped by whether or not dividends were paid.

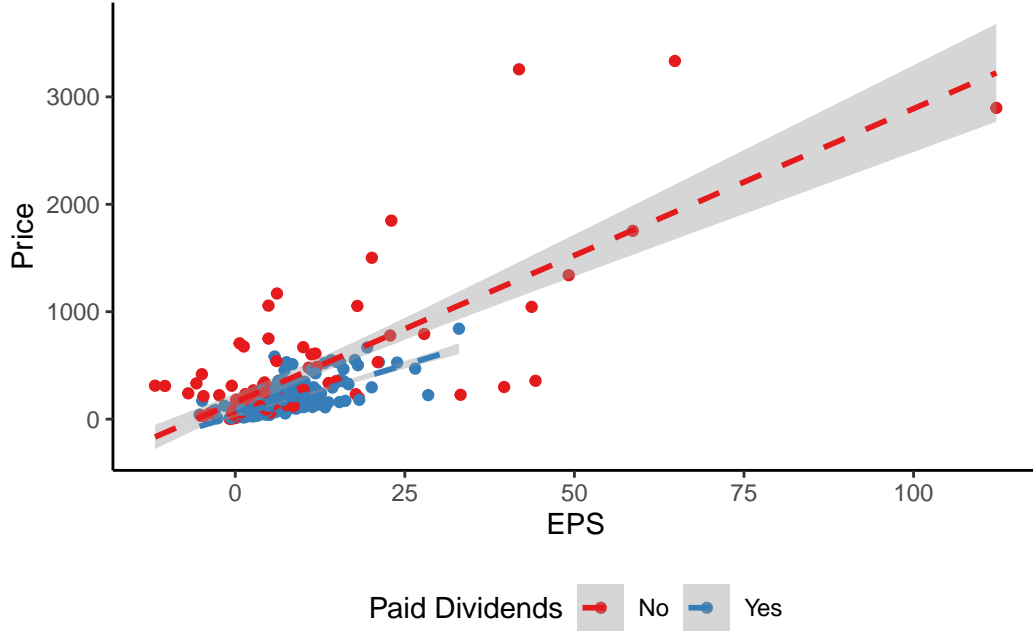


Figure 6: Linear regression

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.

In model 2, we introduce dividends as an additional predictor variable. Stocks that paid dividends have median price around 100, significantly lower than stocks that paid no dividends with median price around 175 (Figure 5), demonstrating a negative relationship between paying dividends and stock price. The second column of Table 2 shows that the new intercept is 172.05 USD, and that the coefficient of paying dividends  $\beta_2$  is -177.85. EPS of companies that paid dividends has intercept of 172.05 and coefficient of 25.59 (Table 2). The new standard deviation for EPS is 1.0 and 21.6 for dividends.

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, and EPS with dividends (blue line) is below it and significantly shorter.



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

	EPS Only	With Dividends
(Intercept)	−124.69	172.05
EPS	21.79	25.59
Paid_DividendYes		−177.85
Num.Obs.	483	483
R2	0.419	0.622
R2 Adj.	0.171	0.588
Log.Lik.	−3425.062	−3228.568
ELPD	−3437.5	−3248.2
ELPD s.e.	82.7	77.6
LOOIC	6875.1	6496.5
LOOIC s.e.	165.4	155.2
WAIC	6877.3	6499.6
RMSE	278.49	193.14

## 5 Discussion

### 5.1 First discussion point

If my paper were 10 pages, then should be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

### 5.2 Second discussion point

### 5.3 Third discussion point

### 5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

## Appendix

### A Model details

#### A.1 Posterior predictive check

In Figure 7a we implement a posterior predictive check. This shows...

In Figure 7b we compare the posterior with the prior. This shows...

```
prior_summary(model_rstanarm_2)
```

```
Priors for model 'model_rstanarm_2'
-----
Intercept (after predictors centered)
  Specified prior:
    ~ normal(location = 0, scale = 2.5)
  Adjusted prior:
    ~ normal(location = 0, scale = 788)

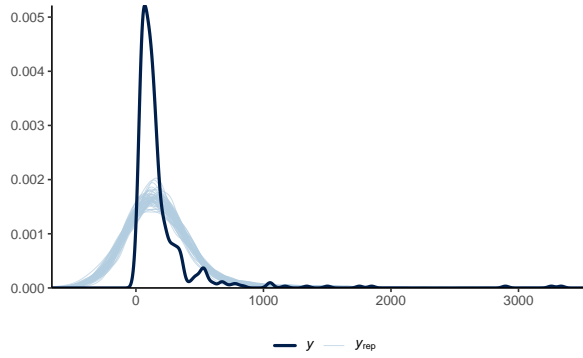
Coefficients
  Specified prior:
    ~ normal(location = [0,0], scale = [2.5,2.5])
  Adjusted prior:
    ~ normal(location = [0,0], scale = [ 88.11,1889.19])

Auxiliary (sigma)
  Specified prior:
    ~ exponential(rate = 1)
  Adjusted prior:
    ~ exponential(rate = 0.0032)
-----
See help('prior_summary.stanreg') for more details
```

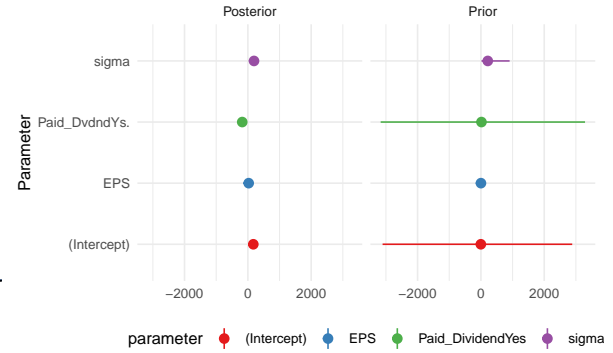
#### A.2 Diagnostics

?@fig-fig-trace-1 is a trace plot

?@fig-fig-trace-2 is a Rhat plot. It shows... This suggests...

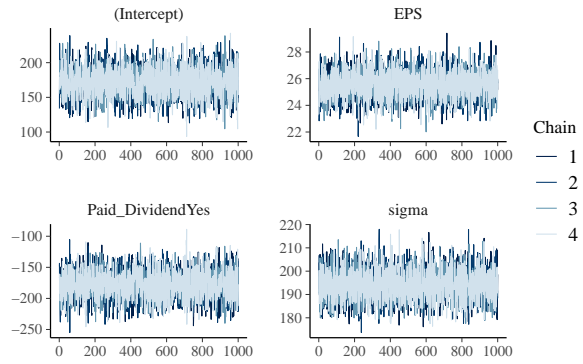


(a) Posterior prediction check

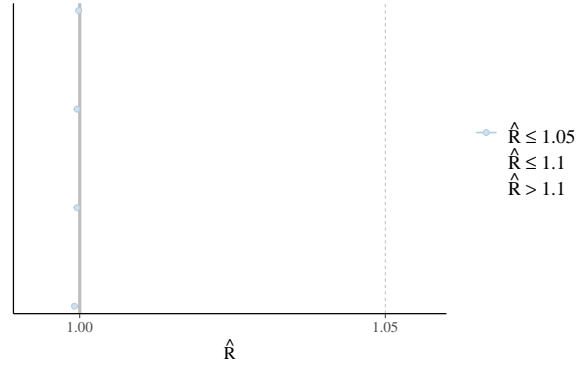


(b) Comparing the posterior with the prior

Figure 7: Examining how the model fits, and is affected by, the data



(a) Trace plot



(b) Rhat plot

Figure 8: Checking the convergence of the MCMC algorithm

## References

- Arel-Bundock, Vincent. 2022. “modelssummary: 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/>.
- 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, Dragoş 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>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, 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/>.