My title*

My subtitle if needed

First author Another author

March 30, 2024

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

Table of contents

*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).

Year	Tic	EPS	Dividends	Net_Income	Price	Dividend_Yield	Paid_Dividend
2014	ABT	1.49	1363	2284	45.02	30.28	Yes
2015	ABT	2.92	1464	4423	44.91	32.60	Yes
2016	ABT	0.94	1547	1400	38.41	40.28	Yes
2017	ABT	0.27	1947	477	57.07	34.12	Yes
2018	ABT	1.33	2047	2368	72.33	28.30	Yes
2019	ABT	2.06	2343	3687	86.86	26.97	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.

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 2). 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 as net income divided by the number of shares of stock. A higher EPS indicates greater value as investors are willing to pay more for a company's shares if they think the company has higher profits relative to its share price (Shakespeare 2020). We expect earnings to be positively correlated with stock price. It is normally distributed and right skewed, with median 4.56 and mean 6.2 (Figure 2). 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. It can be measured as dividends per share, but in the data set it is total dividends. Some companies choose not to pay dividends if they can use their income to reinvest into valuable projects, approximately 78% of companies in our data set paid dividends. We might expect dividends to be positively correlated with stock price because investors desire an additional payment, but the opposite. This is because dividends paid out to shareholders is value that is taken out of the company, causing price which should be a reflection of the firm's value to decrease. The largest dividend in the data set was 25999 USD paid by Caterpillar (CAT) in 2023.

Year is the year in which the financial statement information is released.

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.

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

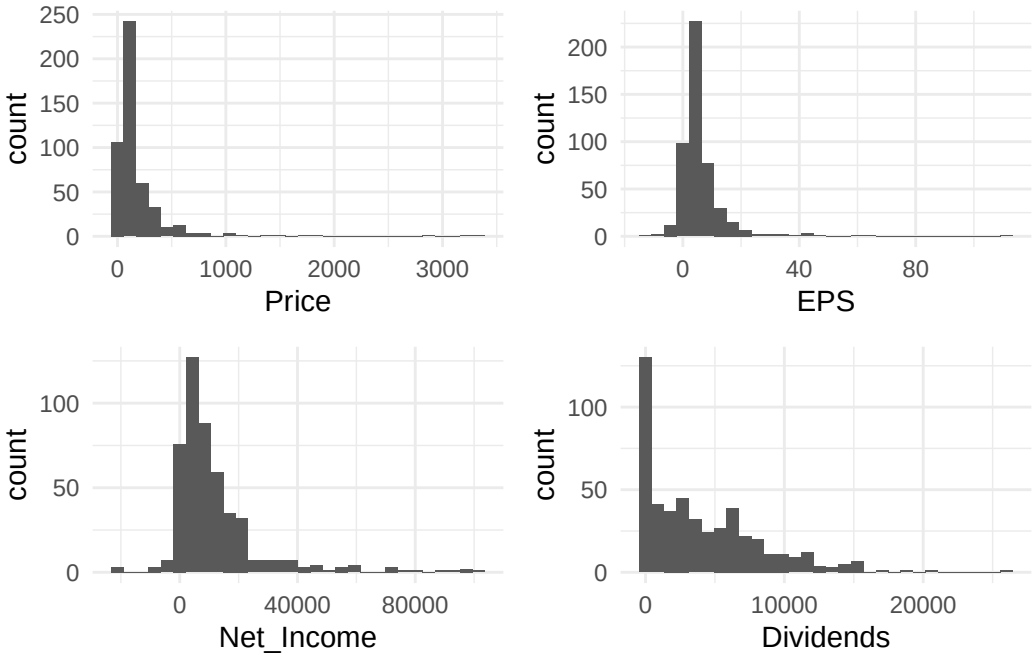


Figure 2: Price, EPS, Net_Income, and Dividends are all normally distributed with right skew

Talk more about it.

Table 2

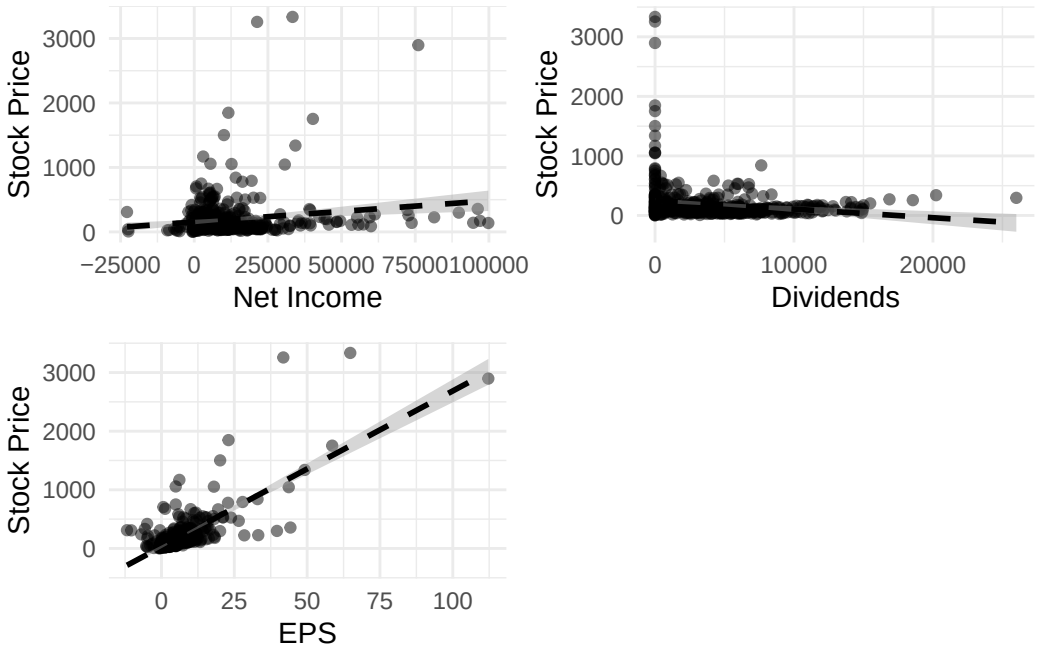


Figure 3Relationship between me and ur mom

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

3 Model

The goal of our modelling strategy is twofold. Firstly,…

Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

Simple Model

$$y = \beta_0 + \beta_1 X + \epsilon \tag{1}$$
$$Y \sim \text{Normal}(\beta, \sigma^2) \tag{2}$$

Where Y is the stock price and X is earnings per share.

We run the model in R (**citeR?**) using the `rstanarm` package of (**rstanarm?**). We use the default priors from `rstanarm`.

Multivariable Model

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 - \beta_3 X_3 + \epsilon \tag{3}$$
$$Y \sim \text{Normal}(\beta, \sigma^2) \tag{4}$$

Where X1 is EPS, X2 is net income, and X3 is dividends.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance θ.

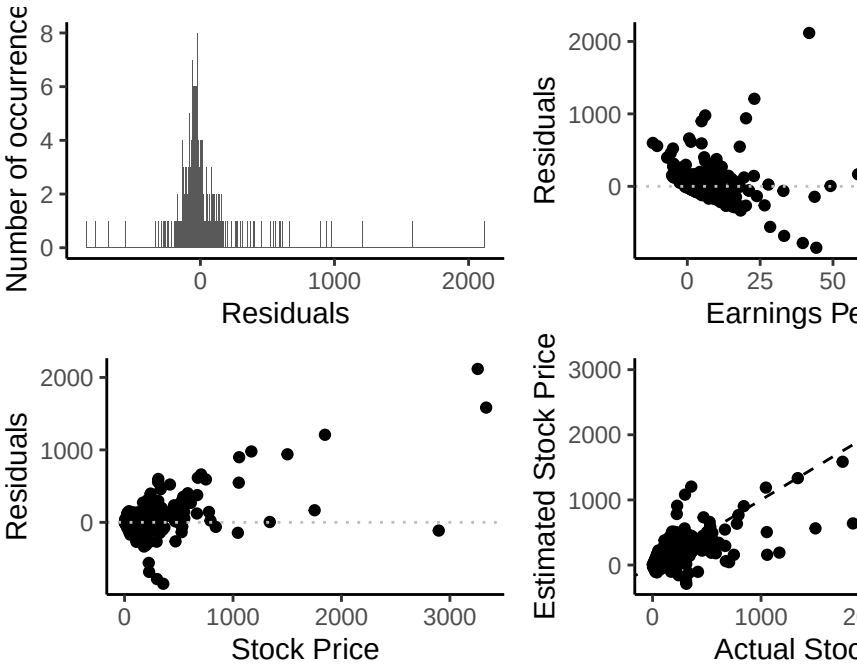
4 Results

Our results are summarized in Table 3.

::: {#tbl-poopy but hole .cell tbl-cap='Explanatory models of flight time based on wing width

Table 3: Explanatory models of flight time based on wing width and wing length

	Net_Income
(Intercept)	27.249
	(11.472)
EPS	26.591
	(1.055)
Num.Obs.	483
R2	0.569
R2 Adj.	0.568
AIC	6526.5
BIC	6539.0
Log.Lik.	−3260.248
F	635.550
RMSE	206.66



and wing length'} ::: {.cell-output-display}
::: :::

Table 4: Explanatory models of flight time based on wing width and wing length

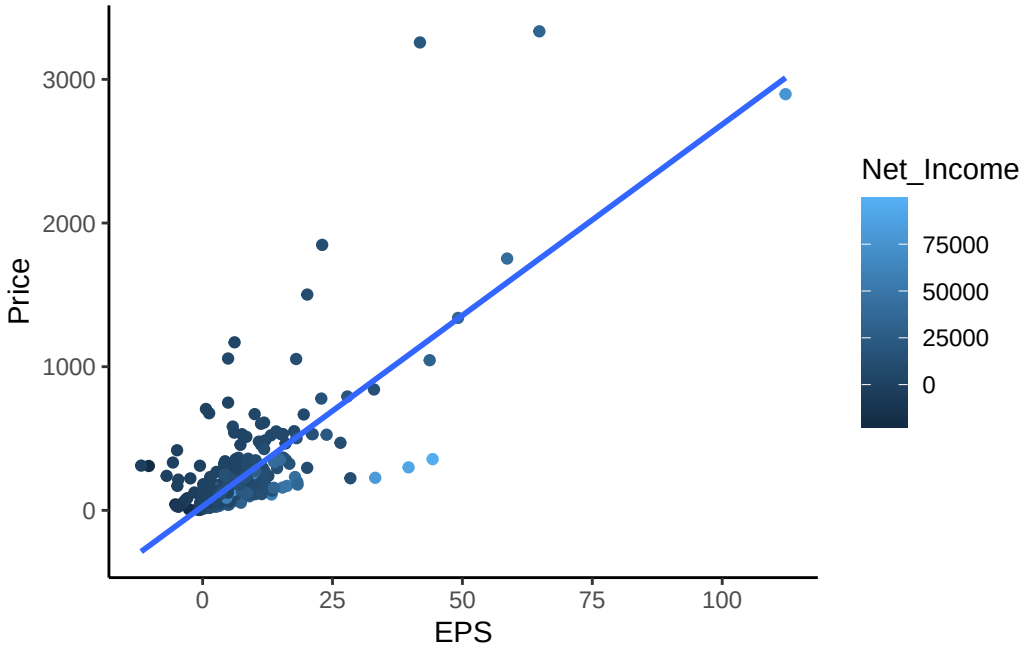


Table 5: Explanatory models of flight time based on wing width and wing length

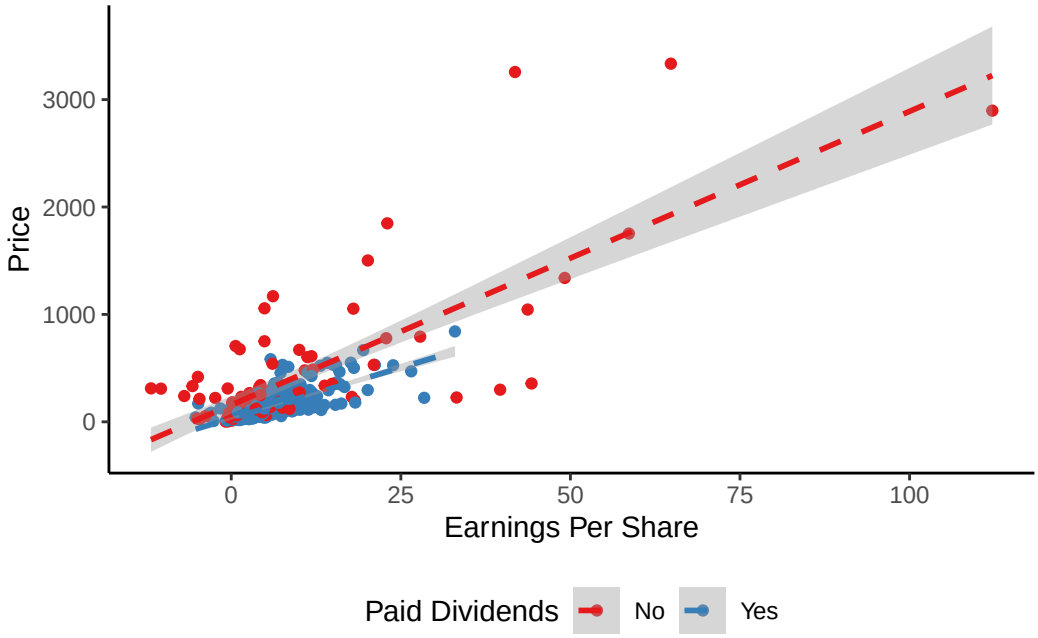


Table 6: Explanatory models of flight time based on wing width and wing length

	EPS Only	With Dividends
(Intercept)	27.25	171.67
	(11.47)	(20.38)
EPS	26.59	25.56
	(1.05)	(0.99)
Paid_DividendYes		−177.80
		(21.32)
Num.Obs.	483	483
R2	0.569	0.624
R2 Adj.	0.568	0.622
AIC	6526.5	6463.2
BIC	6539.0	6479.9
Log.Lik.	−3260.248	−3227.579
F	635.550	397.813
RMSE	206.66	193.14

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be 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.

Table 7: priors

	Non-scaled priors	Auto-scaling priors
(Intercept)	−100.48	60.85
EPS	24.32	30.86
Net_Income	0.00	−0.01
Num.Obs.	483	483
R2	0.438	0.619
R2 Adj.	0.208	0.586
Log.Lik.	−3410.327	−3230.747
ELPD	−3424.6	−3248.6
ELPD s.e.	82.7	74.4
LOOIC	6849.2	6497.1
LOOIC s.e.	165.5	148.8
WAIC	6850.8	6501.0
RMSE	270.70	194.10

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In Figure 4a we implement a posterior predictive check. This shows...

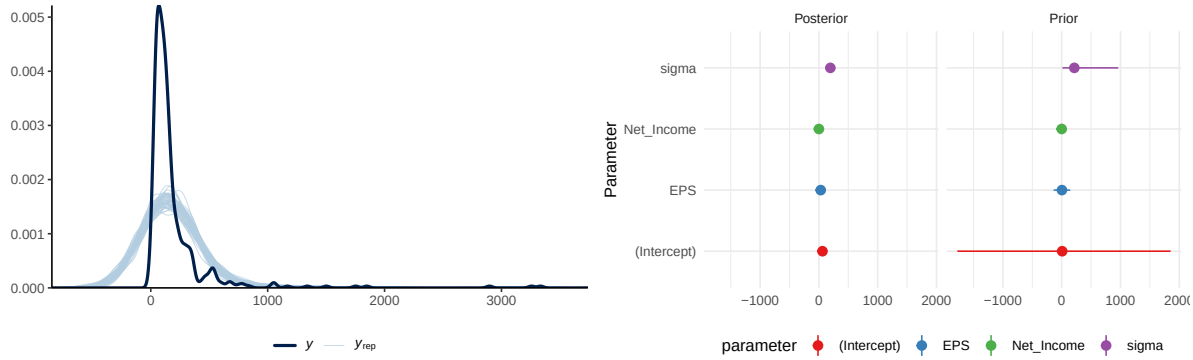
In Figure 4b we compare the posterior with the prior. This shows...

```
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, 0.05])

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



(a) Posterior prediction check

(b) Comparing the posterior with the prior

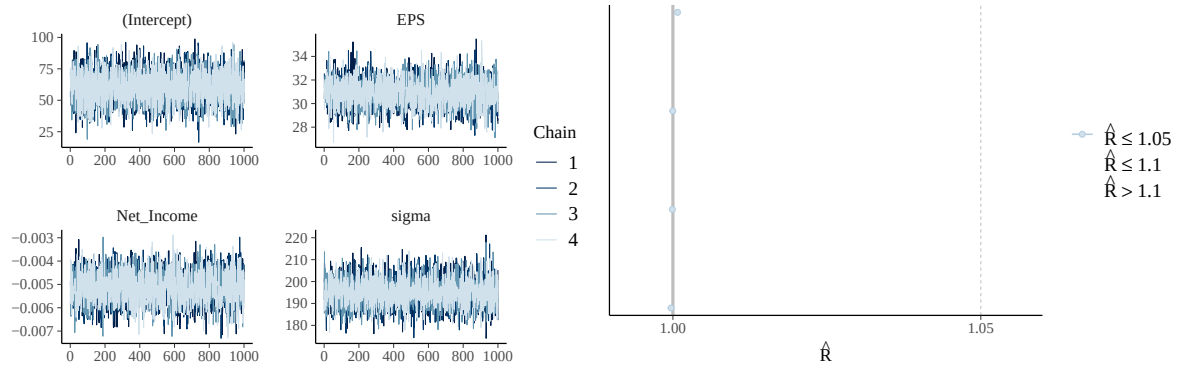
Figure 4: Examining how the model fits, and is affected by, the data

B.2 Diagnostics

Figure 5a is a trace plot. It shows... This suggests...

Figure 5b is a Rhat plot. It shows... This suggests...

‘‘



(a) Trace plot (b) Rhat plot

Figure 5: Checking the convergence of the MCMC algorithm

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/.

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.

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/.