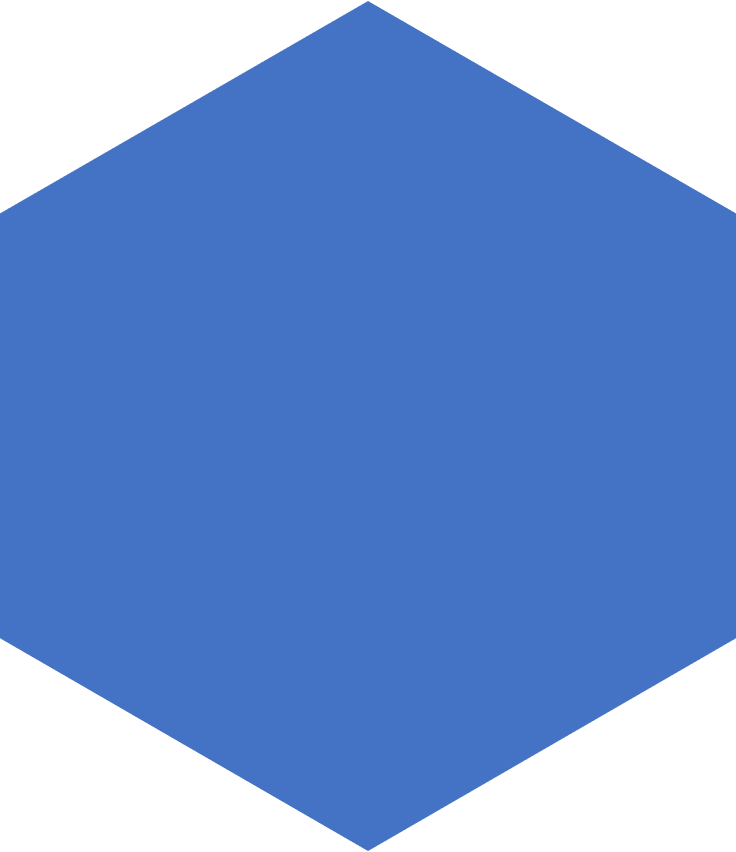


|  |
| --- |
| **Analysis of Google Stock Prices** |
| **MM916 Project** |
| The project is about creating a linear regression model that analyses and predicts Google’s stock price from daily closing stock prices of top 25 companies listed on NYSE. The project involved analyzing 5 year NYSE data, finding the best attributes with the help of correlation, leaps and stepwise selection, interpretation plots, satisfying assumptions with a series of transformations and improving the prediction accuracy of the regression model. |
|  |



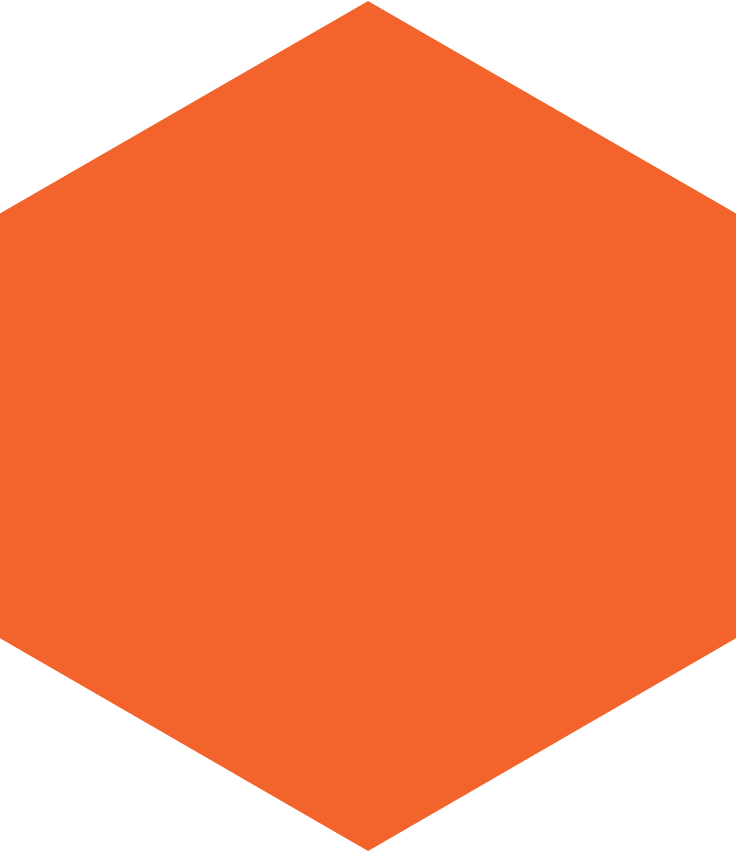
Group K

Cameron Welsh

Dimitur Stefanov

Heng Wu

Sonal Jain



Contents

[Part 1 2](#_Toc27612271)

[Assessing correlations 2](#_Toc27612272)

[Model with most correlated variables 2](#_Toc27612273)

[Variable selection 3](#_Toc27612274)

[Checking assumptions 3](#_Toc27612275)

[Part2 6](#_Toc27612276)

[Prediction competition 6](#_Toc27612277)

[Cross-validation 6](#_Toc27612278)

[References 7](#_Toc27612279)

[Appendix 7](#_Toc27612280)

[Dataset 7](#_Toc27612281)

[Main models 7](#_Toc27612282)

[Correlations bar charts (non-absolute values) 8](#_Toc27612283)

[Variable Selection Output 9](#_Toc27612284)

[Leaps and bounds 9](#_Toc27612285)

[Forward selection 11](#_Toc27612286)

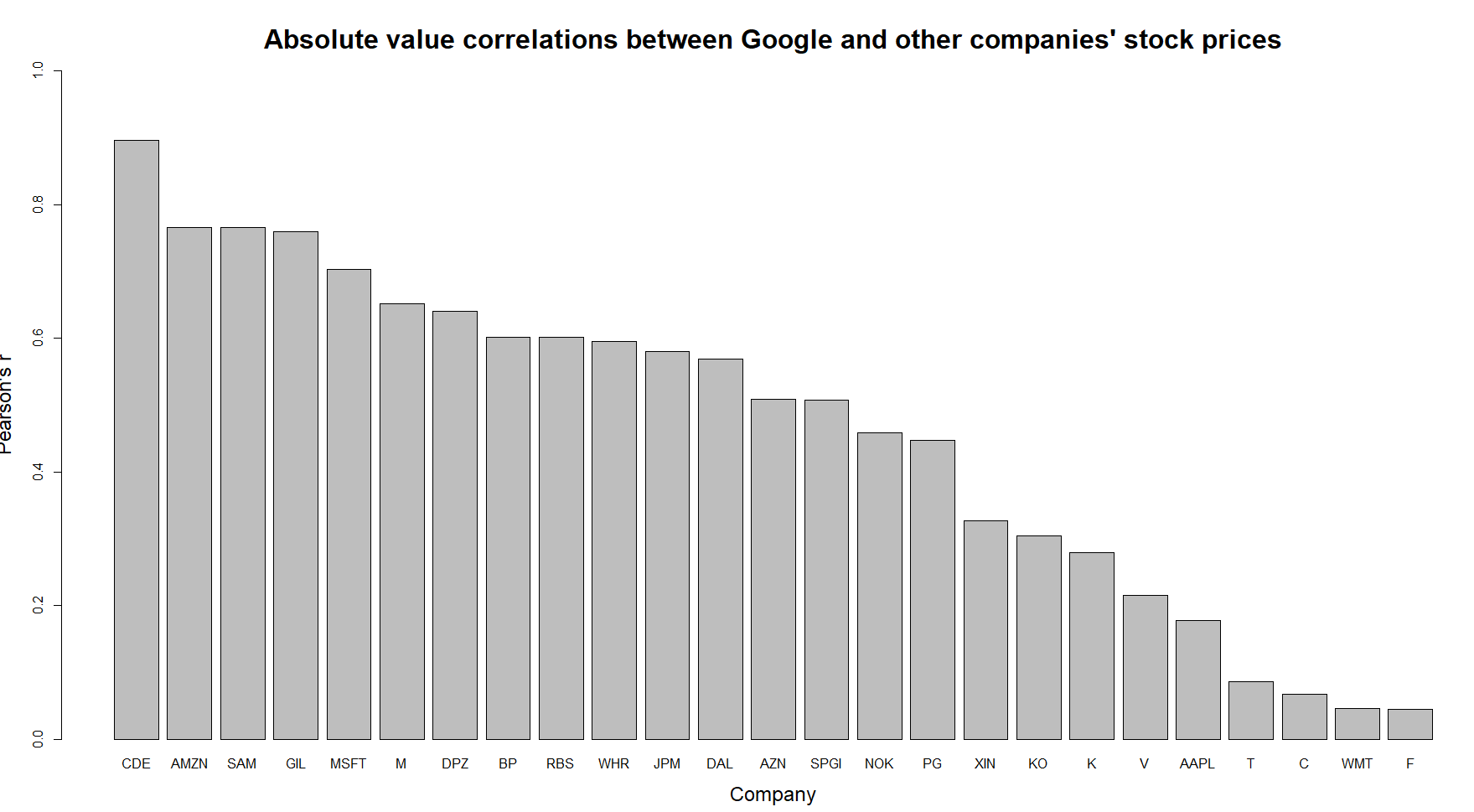
[Stepwise selection 12](#_Toc27612287)

# Part 1

## Assessing correlations

A subset of the data frame was created, excluding the date, year, month and weekday variables. Then a vector was created to store the correlation coefficients between Google and other companies’ closing prices. The coefficients were converted to absolute values and sorted in descending order to better assess strength of relationships (Fig.1).

CDE had the highest correlation (-0.9), followed by AMZN (-0.77), SAM (0.76), GIL (0.76) and MSFT (-0.7), M (0.65). Later on, M and DPZ were also considered when looking for a better model due to their high correlations: 0.65 and -0.64, respectively (see variable selection).

Fig1

## Model with most correlated variables

A model was created with Google as the dependent variable and AMZN, SAM, GIL, MSFT and M as independent variables (M1). This model gave R2=0.8423, meaning that it accounts for 84.2% of the variability with respect to GOOGL. The respective Adjusted R2 was 0.8417. The model was not put through to the online competition; instead, a cross validation was performed which gave a logged mean square error (LMSE) of 10.277.

The model can be summarized with the following equation:

*GOOGL = 448.322 - 63.137 x CDE + 77.778 x AMZN + 40.058 x SAM + 52.864 x GIL - 86.971 x MSFT*

A second model was fitted (M2), containing only the most correlated variable CDE. It managed to provide better predictions (Adjusted R2 = 0.8037, LMSE=9.749), suggesting that some variables in M1 may be redundant. Details of this and other models can be found in the appendix.

## Variable selection

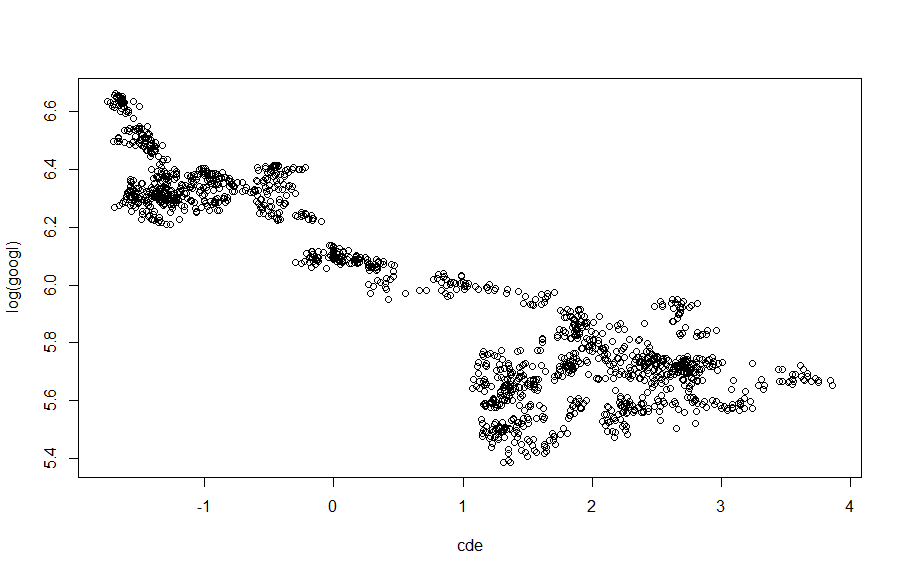
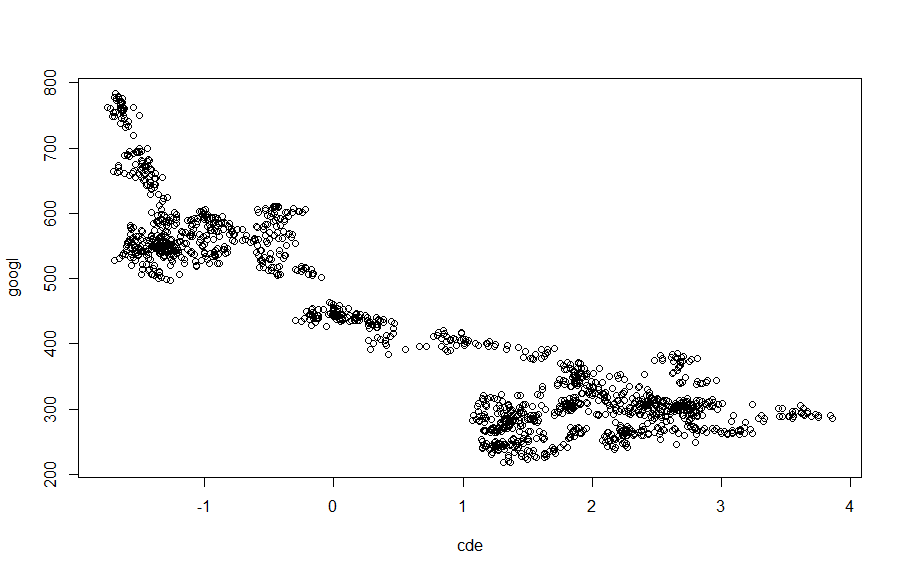
Using the seven variables most strongly correlated with Google, three different techniques were employed to select appropriate predictors. First, R’s leaps() function looks for the best combination of independent variables (rdocumentation.org). For the purpose, the best five subsets for each size were reported and Adjusted R2 and Mallow’s Cp were used to measure performance (see Appendix). Looking at Adjusted R2, models that performed worse than M2 were not considered – this eliminated all models with one independent variable. Including GIL to M2 increased the Adjusted R2 to 0.823. Models with three and four predictive variables improved this by around 0.01 at best, and models with five or more did not improve enough to justify the increasing complexity.

The next variable selection technique was forward selection using R’s add1() function. Here, a scope of independent variables is specified; the algorithm starts with an intercept-only model and computes changes of fit for adding each variable individually (rdocumentation.org). At the first step, CDE provided the best improvement as its F-value by far the highest (see Appendix). After adding CDE, the algorithm showed that GIL would be the next best addition. The next proposed addition was SAM, however, it only increased Adjusted R2 to 0.834. Afterwards, MSFT was proposed, but this did not improve the model enough, so no further additions were considered.

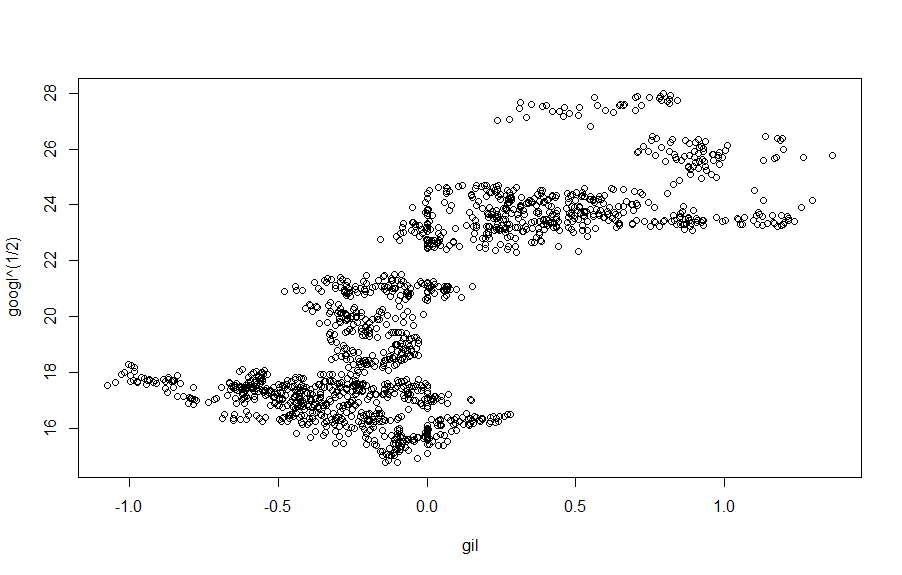
Finally, stepwise variable selection was performed with step() which iterates through the specified scope using both add1() and drop1()(rdocumentation.org). Stepwise selection uses Akaike’s Information Criterion (AIC) to determine adding and/or removing variables. The intercept-only model had an AIC of 14878.23; the algorithm added CDE which reduced AIC to 12436.91. At the next step, GIL was added, lowering AIC by 152.24. The following step suggested adding SAM which reduced AIC by 100.3, while the following added MSFT but the AIC decrease was only 17.19.

From the above it was decided to use CDE and GIL as independent variables (M3). Cross-validation showed improvement in predictions (LMSE=9.563).

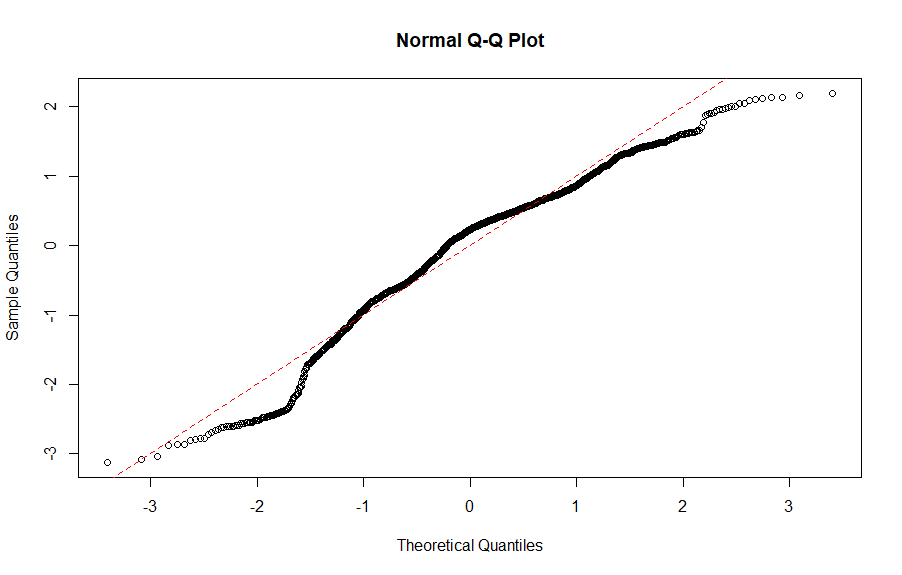
## Checking assumptions



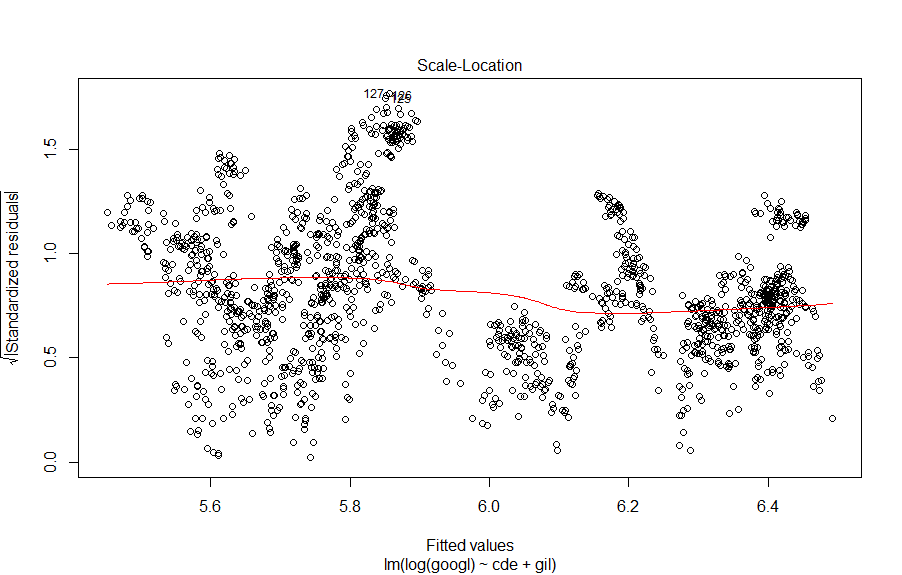
There appears to be a non-linear relationship between GOOGL and CDE (as shown in the left panel above). The data can be transformed (according to Tukey’s ladder) by considering the relationship between CDE and the natural logarithm of GOOGL. The result of plotting the latter against the former reduces the curvature in the relationship, as shown in the right panel. The natural logarithm of GOOGL is also plotted against the other independent variable, GIL, in the figure below.



The linear model with response log GOOGL and explanatory variables CDE and GIL is fitted under the assumptions that the errors are independent and identically normally distributed with mean zero and constant, positive variance σ2. The validity of these assumptions is assessed using the plots shown below.

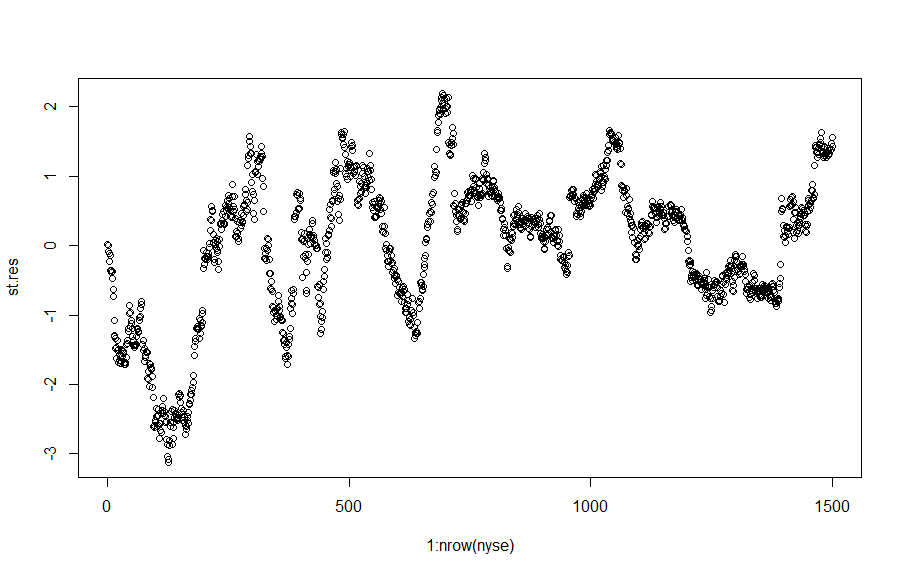


The normal Q-Q plot graphs the standardised residuals on the vertical axis against the quantiles of the standard normal distribution. If the residuals are normally distributed, then the points representing the standardised residuals would lie on the red dashed line with intercept zero and slope one. This is not the case, so the normality assumption is violated.



The scale-location plot should exhibit a random scatter of points if the errors have constant, positive variance. This assumption appears largely to have been met; however, the residuals appear to be “bunched” together in places, which may imply that the model is not explaining the entire relationship between the response and explanatory variables.

The biggest violation of the model assumptions is the serial correlation in the residuals, as shown by the versus-order plot below (which should show a random scatter with no obvious relationship between the order of observation and the residuals). This is to be expected given that the data are time series observations from stock price listings.



Given these adjustments, M3 scored 8.679 on LMSE.

# Part2

While looking for a good set of independent variables, different models were considered and evaluated via the online prediction competition which used root mean squared prediction error (RMSPE), and cross-validation (LMSE).

## Prediction competition

The first submission had CDE and SAM as independent variables and both were squared; GOOGL’s cube root was used as a dependent variable. This gave an RMSPE of 12.14.

The following submission included the logged GOOGL, and squared CDE and GIL as predictors. The result of that was an RMSPE of over 13. Another variation of this model was submitted but it did not show significant improvements.

These results helped identify that the transformations undertaken were not appropriate. Particularly with CDE, a square root would have been better. A constant was added to avoid squaring negative numbers, however, the competition had closed by then.

## Cross-validation

Using an algorithm from MyPlace, cross-validation was performed where 1000 data points were used for training the model, and the rest of the dataset was used for generating predictions.

A model with CDE and SAM as predictors gave an LMSE of 9.714; squaring these did not bring any changes. Including either the log or the cube root of GOOGL increased it to roughly 12.8. Again, this helped identify issues with the transformations applied to the data. After taking the square roots of SAM and CDE, the LMSE dropped to 9.01. This came really close to M3 which could explain why these sets of variables had very similar Adjusted R2 when fitted in the leaps and bounds function.

For exploratory purposes, a model with CDE, GIL, SAM, M and DPZ was considered, however, it scored 10.313. Afterwards, the independent variables were squared which increased the LMSE to 12.752.

With the above in mind it is clear that M3 performed the best for predicting closing stock prices.

# References

rdocumentation.org (no year) *RDocumentation*. Available on: <https://www.rdocumentation.org/>. [accessed 18 December 2019]

# Appendix

## Dataset

The dataset’s columns are as follows:

All of the stock prices apart for Google’s have been centred, scaled and detrended by the provider, meaning that they no longer have meaningful units and strong collinearity has been accounted for. Data for Google’s stocks is measured in US Cents (¢) and has not been modified.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column number** | **Column name** | **Description** | **Column number** | **Column name** | **Description** |
| 1 | Date | Date of closing prices (YYYY-mm-dd) | 16 | JPM | JP Morgan |
| 2 | Year | Year (YYYY) ∈ {2010, . . . , 2016} | 17 | K | Kellogs |
| 3 | Month | Month (mm) ∈ {1, . . . , 12} | 18 | KO | Coca-Cola Company |
| 4 | Weekday | Day of the week ∈ {Monday, . . . , Friday} | 19 | M | Macy’s Inc. |
| 5 | GOOGL | Google | 20 | MSFT | Microsoft |
| 6 | AAPL | Apple | 21 | NOK | Nokia |
| 7 | AMZN | Amazon | 22 | PG | Procter & Gamble |
| 8 | AZN | Astrazeneca | 23 | RBS | Royal Bank of Scotland |
| 9 | BP | British Petroleum | 24 | SAM | Boston Beer Company |
| 10 | C | Citigroup | 25 | SPGI | S&P Global Inc. |
| 11 | CDE | Couer Mining | 26 | T | AT&T |
| 12 | DAL | Delta Airlines | 27 | V | Visa Inc. |
| 13 | DPZ | Domino’s Pizza | 28 | WMT | Walmart |
| 14 | F | Ford | 29 | WHR | Whirlpool Inc. |
| 15 | GIL | Gidlan Activewear | 30 | XIN | Xinyuan Real Estate Co. Ltd. |

## Main models

Model 1 (M1)

*GOOGL = 448.322 - 63.137 x CDE + 77.778 x AMZN + 40.058 x SAM + 52.864 x GIL - 86.971 x MSFT*

R2 = 0.8423. Adjusted R2 = 0.8037. All variables were significant at the 0.001 level.

Model 2 (M2)

*GOOGL = 472.534 - 80.05 x CDE*

R2 = 0.8039. Adjusted R2 = 0.8417. All variables were significant at the 0.001 level.

Model 3 (M3)

*GOOGL = 705.719 - 199.632 x sqrt(CDE) + 32.337 x GIL*

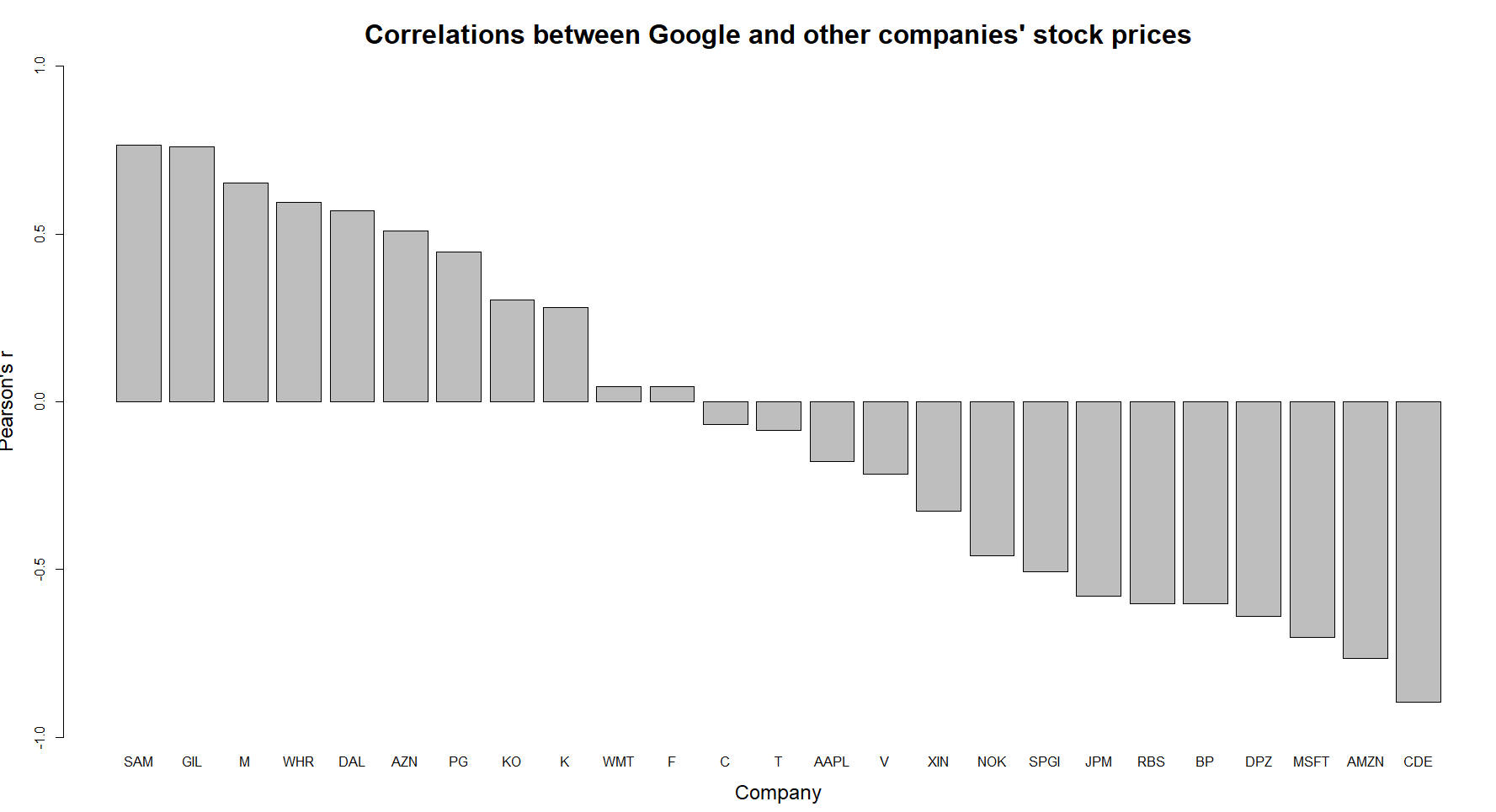
R2 = 0.8467. Adjusted R2 = 0.8465. All variables were significant at the 0.001 level.

Model 4 (M4)

This model was included in the Appendix since it came very close to the best performing one – M3. *GOOGL = 450.544 - 67.054 x CDE + 51.567 x SAM*

R2 = 0.8187. Adjusted R2 = 0.8185. All variables were significant at the 0.001 level.

## Correlations bar charts (non-absolute values)



## Variable Selection Output

### Leaps and bounds

#### Adjusted R2

Size AdjR2 CDE AMZN SAM GIL MSFT M DPZ

1 2 0.804 TRUE FALSE FALSE FALSE FALSE FALSE FALSE **(Model2)**

2 2 0.587 FALSE TRUE FALSE FALSE FALSE FALSE FALSE

3 2 0.586 FALSE FALSE TRUE FALSE FALSE FALSE FALSE

4 2 0.576 FALSE FALSE FALSE TRUE FALSE FALSE FALSE

5 2 0.494 FALSE FALSE FALSE FALSE TRUE FALSE FALSE

6 3 0.823 TRUE FALSE FALSE TRUE FALSE FALSE FALSE **(Model3)**

7 3 0.818 TRUE FALSE TRUE FALSE FALSE FALSE FALSE

8 3 0.814 TRUE FALSE FALSE FALSE FALSE TRUE FALSE

9 3 0.811 TRUE FALSE FALSE FALSE TRUE FALSE FALSE

10 3 0.804 TRUE FALSE FALSE FALSE FALSE FALSE TRUE

11 4 0.834 TRUE FALSE TRUE TRUE FALSE FALSE FALSE

12 4 0.832 TRUE FALSE FALSE TRUE TRUE FALSE FALSE

13 4 0.830 TRUE FALSE FALSE TRUE FALSE TRUE FALSE

14 4 0.826 TRUE FALSE FALSE TRUE FALSE FALSE TRUE

15 4 0.824 TRUE TRUE FALSE TRUE FALSE FALSE FALSE

16 5 0.836 TRUE FALSE TRUE TRUE TRUE FALSE FALSE

17 5 0.836 TRUE FALSE TRUE TRUE FALSE FALSE TRUE

18 5 0.836 TRUE FALSE TRUE TRUE FALSE TRUE FALSE

19 5 0.835 TRUE TRUE FALSE TRUE TRUE FALSE FALSE

20 5 0.835 TRUE TRUE TRUE TRUE FALSE FALSE FALSE

21 6 0.851 TRUE TRUE FALSE TRUE FALSE TRUE TRUE

22 6 0.851 TRUE TRUE TRUE FALSE FALSE TRUE TRUE

23 6 0.846 TRUE TRUE TRUE TRUE FALSE FALSE TRUE

24 6 0.842 TRUE TRUE TRUE TRUE TRUE FALSE FALSE **(Model1)**

25 6 0.841 TRUE TRUE FALSE TRUE TRUE TRUE FALSE

26 7 0.865 TRUE TRUE TRUE TRUE FALSE TRUE TRUE

27 7 0.855 TRUE TRUE FALSE TRUE TRUE TRUE TRUE

28 7 0.852 TRUE TRUE TRUE FALSE TRUE TRUE TRUE

29 7 0.848 TRUE TRUE TRUE TRUE TRUE FALSE TRUE

30 7 0.844 TRUE TRUE TRUE TRUE TRUE TRUE FALSE

31 8 0.866 TRUE TRUE TRUE TRUE TRUE TRUE TRUE

#### Mallow’s CP

Size Cp CDE AMZN SAM GIL MSFT M DPZ

1 2 691.361 TRUE FALSE FALSE FALSE FALSE FALSE FALSE **(Model2)**

2 2 3105.874 FALSE TRUE FALSE FALSE FALSE FALSE FALSE

3 2 3116.671 FALSE FALSE TRUE FALSE FALSE FALSE FALSE

4 2 3233.387 FALSE FALSE FALSE TRUE FALSE FALSE FALSE

5 2 4143.472 FALSE FALSE FALSE FALSE TRUE FALSE FALSE

6 3 479.623 TRUE FALSE FALSE TRUE FALSE FALSE FALSE **(Model3)**

7 3 527.845 TRUE FALSE TRUE FALSE FALSE FALSE FALSE

8 3 578.307 TRUE FALSE FALSE FALSE FALSE TRUE FALSE

9 3 606.944 TRUE FALSE FALSE FALSE TRUE FALSE FALSE

10 3 687.410 TRUE FALSE FALSE FALSE FALSE FALSE TRUE

11 4 351.509 TRUE FALSE TRUE TRUE FALSE FALSE FALSE

12 4 382.715 TRUE FALSE FALSE TRUE TRUE FALSE FALSE

13 4 400.380 TRUE FALSE FALSE TRUE FALSE TRUE FALSE

14 4 439.077 TRUE FALSE FALSE TRUE FALSE FALSE TRUE

15 4 470.105 TRUE TRUE FALSE TRUE FALSE FALSE FALSE

16 5 330.071 TRUE FALSE TRUE TRUE TRUE FALSE FALSE

17 5 332.711 TRUE FALSE TRUE TRUE FALSE FALSE TRUE

18 5 336.780 TRUE FALSE TRUE TRUE FALSE TRUE FALSE

19 5 340.980 TRUE TRUE FALSE TRUE TRUE FALSE FALSE

20 5 348.479 TRUE TRUE TRUE TRUE FALSE FALSE FALSE

21 6 168.333 TRUE TRUE FALSE TRUE FALSE TRUE TRUE

22 6 171.756 TRUE TRUE TRUE FALSE FALSE TRUE TRUE

23 6 221.563 TRUE TRUE TRUE TRUE FALSE FALSE TRUE

24 6 271.099 TRUE TRUE TRUE TRUE TRUE FALSE FALSE **(Model1)**

25 6 276.501 TRUE TRUE FALSE TRUE TRUE TRUE FALSE

26 7 10.103 TRUE TRUE TRUE TRUE FALSE TRUE TRUE

27 7 127.503 TRUE TRUE FALSE TRUE TRUE TRUE TRUE

28 7 159.948 TRUE TRUE TRUE FALSE TRUE TRUE TRUE

29 7 198.475 TRUE TRUE TRUE TRUE TRUE FALSE TRUE

30 7 241.264 TRUE TRUE TRUE TRUE TRUE TRUE FALSE

31 8 8.000 TRUE TRUE TRUE TRUE TRUE TRUE TRUE

### Forward selection

> mod.fwd <- lm(GOOGL ~ 1, data = nyse7)

> add1(mod.fwd, test = "F",

+ scope = ~ CDE + GIL + AMZN + M + DPZ + SAM + MSFT)

Single term additions

Model:

GOOGL ~ 1

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 30422866 14878

CDE 1 24455454 5967412 12437 6139.1 < 2.2e-16 \*\*\*

GIL 1 17520465 12902401 13594 2034.2 < 2.2e-16 \*\*\*

AMZN 1 17868339 12554527 13553 2132.0 < 2.2e-16 \*\*\*

M 1 12923161 17499704 14051 1106.2 < 2.2e-16 \*\*\*

DPZ 1 12467339 17955527 14089 1040.1 < 2.2e-16 \*\*\*

SAM 1 17838883 12583983 13556 2123.5 < 2.2e-16 \*\*\*

MSFT 1 15037633 15385232 13858 1464.2 < 2.2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**CDE added. This gives Model 2.**

> mod.fwd2 <- lm(GOOGL ~ CDE, data = nyse7)

> add1(mod.fwd2, test = "F",

+ scope = ~ CDE + GIL + AMZN + M + DPZ + SAM + MSFT)

Single term additions

Model:

GOOGL ~ CDE

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 5967412 12437

GIL 1 583104 5384308 12285 162.1206 < 2.2e-16 \*\*\*

AMZN 1 2241 5965171 12438 0.5625 0.45337

M 1 313881 5653531 12358 83.1126 < 2.2e-16 \*\*\*

DPZ 1 16234 5951178 12435 4.0837 0.04348 \*

SAM 1 451548 5515864 12321 122.5498 < 2.2e-16 \*\*\*

MSFT 1 235755 5731657 12378 61.5748 8.054e-15 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**GIL added. This gives Model 3.**

> mod.fwd3 <- lm(GOOGL ~ CDE+GIL, data = nyse7)

> add1(mod.fwd3, test = "F",

+ scope = ~ CDE + GIL + AMZN + M + DPZ + SAM + MSFT)

Single term additions

Model:

GOOGL ~ CDE + GIL

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 5384308 12285

AMZN 1 31423 5352885 12278 8.7819 0.003091 \*\*

M 1 221643 5162665 12224 64.2261 2.214e-15 \*\*\*

DPZ 1 116072 5268235 12254 32.9606 1.137e-08 \*\*\*

SAM 1 354968 5029340 12184 105.5867 < 2.2e-16 \*\*\*

MSFT 1 269836 5114472 12210 78.9278 < 2.2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**SAM added. Adjusted R2 = 0.8344**

> mod.fwd4 <- lm(GOOGL ~ CDE+GIL+SAM, data = nyse7)

> add1(mod.fwd4, test = "F",

+ scope = ~ CDE + GIL + AMZN + M + DPZ + SAM + MSFT)

Single term additions

Model:

GOOGL ~ CDE + GIL + SAM

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 5029340 12184

AMZN 1 13724 5015616 12182 4.0907 0.0432966 \*

M 1 45641 4983699 12173 13.6914 0.0002232 \*\*\*

DPZ 1 56741 4972599 12169 17.0591 3.824e-05 \*\*\*

MSFT 1 63943 4965397 12167 19.2522 1.226e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**MSFT added. That’s interesting since the leap models didn’t have it. Adjusted R2 = 0.8364**

> mod.fwd5 <- lm(GOOGL ~ CDE+GIL+SAM+MSFT, data = nyse7)

> add1(mod.fwd5, test = "F",

+ scope = ~ CDE + GIL + AMZN + M + DPZ + SAM + MSFT)

Single term additions

Model:

GOOGL ~ CDE + GIL + SAM + MSFT

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 4965397 12167

AMZN 1 166340 4799057 12118 51.7836 9.772e-13 \*\*\*

M 1 30311 4935086 12160 9.1761 0.002494 \*\*

DPZ 1 6287 4959110 12167 1.8940 0.168954

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**AMZN added . Adjusted R2 would be 0.8417. This model with 5 IVs is worse than the models discovered with stepwise so it is not considered.   
No new models were discovered.**

### Stepwise selection

Start: AIC=14878.23

GOOGL ~ 1

Df Sum of Sq RSS AIC

+ CDE 1 24455454 5967412 12437

+ AMZN 1 17868339 12554527 13553

+ SAM 1 17838883 12583983 13556

+ GIL 1 17520465 12902401 13594

+ MSFT 1 15037633 15385232 13858

+ M 1 12923161 17499704 14051

+ DPZ 1 12467339 17955527 14089

<none> 30422866 14878

Step: AIC=12436.91

GOOGL ~ CDE

Df Sum of Sq RSS AIC

+ GIL 1 583104 5384308 12285

+ SAM 1 451548 5515864 12321

+ M 1 313881 5653531 12358

+ MSFT 1 235755 5731657 12378

+ DPZ 1 16234 5951178 12435

<none> 5967412 12437

+ AMZN 1 2241 5965171 12438

- CDE 1 24455454 30422866 14878

Step: AIC=12284.67

GOOGL ~ CDE + GIL

Df Sum of Sq RSS AIC

+ SAM 1 354968 5029340 12184

+ MSFT 1 269836 5114472 12210

+ M 1 221643 5162665 12224

+ DPZ 1 116072 5268235 12254

+ AMZN 1 31423 5352885 12278

<none> 5384308 12285

- GIL 1 583104 5967412 12437

- CDE 1 7518093 12902401 13594

Step: AIC=12184.37

GOOGL ~ CDE + GIL + SAM

Df Sum of Sq RSS AIC

+ MSFT 1 63943 4965397 12167

+ DPZ 1 56741 4972599 12169

+ M 1 45641 4983699 12173

+ AMZN 1 13724 5015616 12182

<none> 5029340 12184

- SAM 1 354968 5384308 12285

- GIL 1 486524 5515864 12321

- CDE 1 3491147 8520487 12973

Step: AIC=12167.18

GOOGL ~ CDE + GIL + SAM + MSFT

Df Sum of Sq RSS AIC

+ AMZN 1 166340 4799057 12118

+ M 1 30311 4935086 12160

<none> 4965397 12167

+ DPZ 1 6287 4959110 12167

- MSFT 1 63943 5029340 12184

- SAM 1 149075 5114472 12210

- GIL 1 520174 5485571 12315

- CDE 1 2874355 7839752 12850

Step: AIC=12118.06

GOOGL ~ CDE + GIL + SAM + MSFT + AMZN

Df Sum of Sq RSS AIC

+ DPZ 1 203583 4595474 12055

+ M 1 86849 4712208 12093

<none> 4799057 12118

- AMZN 1 166340 4965397 12167

- SAM 1 196102 4995159 12176

- MSFT 1 216559 5015616 12182

- GIL 1 371590 5170647 12228

- CDE 1 2547608 7346664 12755

Step: AIC=12055.04

GOOGL ~ CDE + GIL + SAM + MSFT + AMZN + DPZ

Df Sum of Sq RSS AIC

+ M 1 525098 4070375 11875

<none> 4595474 12055

- MSFT 1 68442 4663915 12075

- DPZ 1 203583 4799057 12118

- SAM 1 362390 4957863 12167

- AMZN 1 363637 4959110 12167

- GIL 1 471890 5067364 12200

- CDE 1 2512614 7108088 12707

Step: AIC=11875.04

GOOGL ~ CDE + GIL + SAM + MSFT + AMZN + DPZ + M

Df Sum of Sq RSS AIC

<none> 4070375 11875

- MSFT 1 11192 4081568 11877

- SAM 1 331476 4401851 11990

- GIL 1 419992 4490367 12020

- M 1 525098 4595474 12055

- DPZ 1 641833 4712208 12093

- AMZN 1 847149 4917525 12157

- CDE 1 2781372 6851747 12654

Call:

lm(formula = GOOGL ~ CDE + GIL + SAM + MSFT + AMZN + DPZ + M,

data = nyse)

Coefficients:

(Intercept) CDE GIL SAM MSFT AMZN DPZ

395.70 -66.59 57.42 58.89 -21.81 291.46 -305.64

M

63.98

**Again, no significant improvements after Model 3**