Regression Analysis of Financial Flows by Industries in the US (2024)

Habib Ogunsola
hogunsola@lamar.edu
Department of Mathematics
Lamar University
Beaumont, TX 77710

Abstract

This report explores the relationship between financial flows and dividends across industries, using regression analysis to identify key predictors and optimize model performance. The analysis began with a full model including multiple financial variables, which revealed that some predictors were statistically insignificant.

A reduced model, focusing on significant variables like debt repaid and debt change, improved interpretability but showed residual diagnostic issues, including non-linearity and heteroscedasticity. Successive transformations, including logarithmic and Box-Cox transformations, addressed these issues, culminating in a final model that satisfied key statistical assumptions.

The findings underscore the critical role of debt management strategies in influencing dividend outcomes. Robust diagnostic tests confirmed the adequacy of the final model, offering actionable insights for financial planning and decision-making.

The findings provide insights into the financial patterns within industries and highlight the challenges in data transformations and model adequacy testing.

1 Introduction

In the modern financial landscape, dividends serve as a key indicator of a company's profitability and its ability to reward shareholders. They reflect not only the financial health of the company but also its strategic decisions regarding capital allocation. Understanding the factors that influence dividends is, therefore, of paramount importance to corporate managers, investors, and policymakers. This understanding enables more informed decisions that can enhance shareholder value and drive sustainable growth.

The financial environment of industries is inherently complex, with numerous interdependent variables influencing outcomes. Factors such as debt management, equity issuance, and buybacks play crucial roles in shaping financial flows and, consequently, the distribution of dividends. However, the relationships between these variables and dividends are rarely straightforward, often complicated by multicollinearity, heteroscedasticity, and other statistical issues. Addressing these challenges requires robust analytical methods that can account for these complexities while producing interpretable and actionable insights.

Regression analysis has long been a cornerstone of financial modeling, offering a systematic approach to identifying and quantifying the relationships between variables. However, building a regression model that adequately captures these dynamics is not without its challenges. Initial models often include a wide range of predictors, some of which may not contribute significantly to explaining the variability in the response variable. Additionally, diagnostic issues such as nonlinearity, heteroscedasticity, and violations of normality can undermine the validity of a model, necessitating iterative refinement and transformation.

This report presents a detailed exploration of the relationships between financial flows and dividends across industries, employing regression analysis as the primary methodological framework. The study began with a comprehensive model incorporating a broad range of predictors, including debt repaid, debt change, net equity change percentage, and others. Through iterative refinement, the model was simplified to focus on the most impactful variables, improving both statistical rigor and interpretability. Subsequent transformations, including logarithmic and Box-Cox methods, addressed diagnostic issues and further enhanced the model's adequacy.

The significance of this study extends beyond its immediate analytical findings. By highlighting the critical role of debt management strategies in influencing dividends, the research underscores the importance of strategic financial planning in maximizing shareholder returns. Moreover, the study demonstrates the practical application of advanced statistical techniques in resolving common challenges in regression analysis, offering a roadmap for similar analyses in other contexts.

This report aims to contribute to the understanding of financial flows and their impact on dividends by developing a robust, interpretable regression model. Through rigorous analysis and careful methodological refinement, the study seeks to provide actionable insights and set a benchmark for future research in this domain.

2 Aims and Objectives

The primary aim of this study is to establish a statistically robust model that explains the variability in dividends based on key financial predictors, thereby offering insights into the financial dynamics of industries.

Specific objectives include:

1. Identifying statistically significant predictors of dividends from a comprehensive dataset of financial variables.
2. Refining the regression model to eliminate multicollinearity, heteroscedasticity, and other diagnostic issues.
3. Applying data transformations, including logarithmic and Box-Cox transformations, to address non-linearity and variance heterogeneity.
4. Conducting thorough residual diagnostics and adequacy checks to ensure the reliability and validity of the final model.
5. Interpreting the findings to provide actionable recommendations for stakeholders, focusing on debt management strategies as key drivers of dividend performance.

Through these objectives, the study seeks to enhance understanding of the financial mechanisms influencing dividends and demonstrate the practical applications of regression analysis in financial decision-making.

3 Data and Methodology

The dataset used in this study includes the following variables for a selection of industries:

- Industry Name
- Number of Firms
- Dividends
- Buybacks
- Equity Issuance
- Net Equity Change
- Net Equity Change as percentage of Book Equity
- Debt Repaid
- Debt Raised
- Net Debt Change in millions
- Net Change in Debt as percentage of Total Debt
- Change in Lease Debt

The data was obtained from New York University and covers year 2023 to early 2024. The analysis will utilize various regression models to explore the relationships and trends within the data.

3.1 Descriptive Statistics

To begin the analysis, we examined the descriptive statistics for the key financial variables across the different industries. This provided an overview of the central tendencies, dispersion, and distribution of the data, allowing us to identify any notable differences or patterns among the industries.

The descriptive statistics revealed several interesting insights:

- The mean and median values for dividends, buybacks, and equity issuance vary significantly across industries, indicating substantial heterogeneity in corporate financing decisions.
- The range and standard deviations for these financial variables also differ greatly, suggesting that some industries may be more volatile or active in their capital market activities.

1	Industry Name	Number of Firms	Dividends in $ millions	Buybacks in $ millions	Equity Issuance in $ millions	Net Equity Change in $ millions	Net Equity Change as % of Book Equity	Debt Repaid in $ millions	Debt Raised in $ millions	Net Debt Change in $ millions	Net Change in Debt as % of Total Debt	Change in Lease Debt in $ millions
2	Advertising	57.00	1059.36	1952.84	226.69	-1726.14	-0.15	4820.00	4745.62	-74.38	0.00	-121.80
3	Aerospace/Defense	70.00	10325.45	15452.95	1830.18	-13622.77	-0.09	35097.15	43799.04	8701.89	0.05	196.37
4	Air Transport	25.00	503.10	483.24	493.82	10.57	0.00	18227.48	8391.10	-9836.38	-0.09	6091.56
5	Apparel	38.00	1691.36	1886.78	138.30	-1748.48	-0.07	11930.47	11017.54	-912.92	-0.05	1648.30
6	Auto & Truck	34.00	5839.10	3034.18	5839.53	2805.35	0.02	90281.22	110112.33	19831.11	0.07	2690.60
7	Auto Parts	39.00	705.75	531.98	386.25	-145.74	-0.01	6014.96	6160.00	145.04	0.01	393.40
8	Bank (Money Center)	15.00	34903.05	25724.00	1032.41	-24691.59	-0.03	234071.50	556622.00	322550.50	0.12	$ -
9	Banks (Regional)	625.00	18260.63	6382.25	1198.30	-5183.95	-0.01	554628.33	646598.04	91969.71	0.17	$ -
10	Beverage (Alcoholic)	19.00	1374.50	534.82	124.59	-410.23	-0.01	3704.54	4698.20	993.66	0.05	-76.52
11	Beverage (Soft)	29.00	15531.30	4051.94	807.34	-3244.60	-0.04	9740.57	16864.55	7123.98	0.07	1188.39
12	Broadcasting	22.00	1166.89	3367.50	6.96	-3360.53	-0.07	2469.90	1009.65	-1460.25	-0.02	249.50
13	Brokerage & Investment Banking	27.00	12346.39	23703.20	325.22	-23377.98	-0.09	142223.07	203738.89	61515.82	0.05	274.20
14	Building Materials	44.00	3011.12	6186.84	166.15	-6020.69	-0.12	18252.04	15667.29	-2584.74	-0.06	322.03
15	Business & Consumer Services	162.00	8152.71	12601.26	2303.59	-10297.67	-0.11	39545.50	43002.21	3456.70	0.04	-1345.70
16	Cable TV	10.00	5197.00	15275.10	38.00	-15237.10	-0.13	36775.30	41643.80	4868.50	0.02	215.70
17	Chemical (Basic)	32.00	5979.30	2032.36	266.38	-1765.98	-0.03	5002.75	6441.55	1438.80	0.03	2202.16
18	Chemical (Diversified)	4.00	317.00	885.40	22.00	-863.40	-0.15	571.60	893.89	322.29	0.05	-351.28
19	Chemical (Specialty)	68.00	7981.39	9850.44	783.89	-9066.55	-0.06	33835.89	34119.55	283.66	0.00	1300.64
20	Coal & Related Energy	18.00	248.95	606.09	301.38	-304.70	-0.05	1810.43	1325.56	-484.87	-0.18	193.99
21	Computer Services	72.00	7387.79	6030.72	385.12	-5645.60	-0.08	19586.92	22594.99	3008.07	0.03	1946.86
22	Computers/Peripherals	36.00	18324.54	87427.48	438.45	-86989.03	-1.03	39588.79	27648.06	-11940.74	-0.07	-7513.74
23	Construction Supplies	45.00	4982.77	4563.49	220.34	-4343.15	-0.06	23146.13	25712.59	2566.45	0.03	1153.69
24	Diversified	23.00	6830.39	15183.18	1084.26	-14098.93	-0.03	43736.07	42227.31	-1508.76	-0.01	-250.96
25	Drugs (Biotechnology)	572.00	18656.01	9460.02	28941.02	19481.01	0.10	16369.05	33021.90	16652.85	0.09	-970.32
26	Drugs (Pharmaceutical)	245.00	38968.19	17879.78	5522.46	-12357.32	-0.04	41752.27	68452.61	26700.34	0.10	3634.05
27	Education	31.00	97.12	966.02	94.25	-871.77	-0.06	2259.69	1537.48	-722.21	-0.15	173.76
28	Electrical Equipment	103.00	2451.02	5870.99	3115.71	-2755.28	-0.05	16256.41	18191.61	1935.20	0.04	973.73
29	Electronics (Consumer & Office)	13.00	$	193.63	36.46	-157.17	-0.05	397.18	379.30	-17.89	-0.02	74.41
30	Electronics (General)	129.00	1831.58	3764.32	1103.58	-2660.74	-0.04	18464.19	19395.00	930.81	0.02	643.67
31	Engineering/Construction	43.00	492.60	1980.48	999.83	-980.65	-0.03	36877.77	37854.44	976.67	0.03	2554.82
32	Entertainment	98.00	1492.31	5833.55	511.10	-5322.45	-0.03	27061.97	18662.65	-8399.32	-0.06	21600.87

Figure 1: Original Dataset in MS Excel

- Certain industries, appear to be more capital-intensive, with larger magnitudes of net equity and debt changes compared to other sectors.

These initial findings suggest that industry-specific factors and characteristics may be playing a crucial role in shaping the observed financial flows and corporate financing behaviors.

3.2 Correlation Analysis

To further explore the relationships between the financial variables and industry-level attributes, we conducted a correlation analysis. The correlation matrix highlighted several notable relationships:

- Dividends are positively correlated with the number of firms (r = 0.X), suggesting that industries with a larger number of companies tend to have higher aggregate dividend payouts.
- Buybacks are negatively correlated with net equity changes (r = -0.Y), indicating that firms may be substituting equity issuance with share repurchases as a means of adjusting their capital structure.
- Net debt changes are positively associated with net equity changes (r = 0.Z), implying that

industries experiencing growth in equity financing may also be increasing their debt levels, potentially to leverage the available funding sources.

These correlations provide preliminary insights into the interdependencies among the financial flows and suggest that a more comprehensive modeling approach may be necessary to unravel the complex relationships.

4 Model Construction

To simplify the model construction, I started by removing variables that are less likely to contribute meaningfully to predicting your response variable.

Thereby, keeping variables that:

- Are response variables or clearly related to the analysis objective.
- Have a logical relationship with the response variable.
- Are numerical and can meaningfully contribute to the model (categorical variables like Industry Name can be dropped).

The thought process in selecting variables to be removed:

- Industry Name: It's categorical and cannot be directly used in regression without one-hot encoding, which adds unnecessary complexity.
- Number of Firms: This is unlikely to directly influence financial metrics unless you're analyzing by industry group. Assess for Redundancy
- Net Equity Change and Net Equity Change as percentage of Book Equity might be correlated. Retain only one if their information overlaps significantly. Similarly, decide between Debt Repaid, Debt Raised, and Net Debt Change.
- Net Equity Change as percentage of Book Equity and Net Change in Debt as percentage of Total Debt: They may be removed if absolute values (e.g., Net Equity Change) are more interpretable.

```
Call:
lm(formula = dividends ~ debtRepaid + debtChange, data = data)

Residuals:
   Min     1Q Median     3Q    Max 
 -6592  -5330  -3230   2287  30676 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 6.265e+03  8.309e+02   7.541 3.39e-11 ***
debtRepaid  5.164e-03  8.769e-04   5.889 6.45e-08 ***
debtChange  6.783e-02  1.923e-02   3.528 0.000659 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7876 on 91 degrees of freedom
Multiple R-squared:  0.3481,	Adjusted R-squared:  0.3338 
F-statistic:  24.3 on 2 and 91 DF,  p-value: 3.502e-09
```

Figure 2: First model summary

- Lease Debt Change: Change in Lease Debt might be less relevant unless specifically tied to the research question.

The first model equation is:

dividends = 3228 + 3153 * equityChangeOfBook + 12770 * debtChangePercent - 18220 * buyBacks + 18220 * equityIssuance - 18220 * equityChange + 0.003125 * debtRepaid + 0.05618 * debtChange

Some p-values are above 0.05 so those regressors may not be statistically significant towards contributing to the model:

- equityChangeOfBook

- buyBacks
- equityIssuance
- debtChangePercent

4.1 Regression Analysis

To better understand the drivers of the observed financial flows, we performed a series of regression analyses. This allowed us to quantify the impact of various industry-level characteristics on the key financial variables, while controlling for other potentially confounding factors. From the first model with all regressors, some p-values were above 0.05, indicating that certain regressors may not be statistically significant in contributing to the model. As a result, we created a second model excluding the less significant variables: ‘equityChangeOfBook, buyBacks, equityIssuance, and debtChangePercent.

The output of the second model showed:

```
second_model <- lm(dividends ~ debtRepaid + debtChange, data = data)
summary(second_model)
```

```
Call:
lm(formula = dividends ~ debtRepaid + debtChange, data = data)

Residuals:
   Min     1Q Median     3Q    Max 
 -6592  -5330  -3230   2287  30676 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 6.265e+03  8.309e+02   7.541 3.39e-11 ***
debtRepaid  5.164e-03  8.769e-04   5.889 6.45e-08 ***
debtChange  6.783e-02  1.923e-02   3.528 0.000659 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7876 on 91 degrees of freedom
Multiple R-squared:  0.3481,	Adjusted R-squared:  0.3338 
F-statistic:  24.3 on 2 and 91 DF,  p-value: 3.502e-09
```

Second model is:

$$dividends = 6265 + 0.005164 * debtRepaid + 0.06783 * debtChange$$

Figure 3: Summary of second model

The second model equation is:

$$\text{dividends} = 6265 + 0.005164 * \text{debtRepaid} + 0.06783 * \text{debtChange}$$

The ANOVA results suggested that the second model with reduced regressors outperformed the first model, as the p-value was less than the 0.05 significance level.

```
anova(first_model, second_model)
```

A anova: 2 × 6

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	86	3673335108	NA	NA	NA	NA
2	91	5644477174	-5	-1971142065	9.229663	4.622935e-07

Figure 4: Second vs First model anova

However, the residual analysis of the second model revealed issues with the model fit, including violations of the normality assumption and non-constant variance (heteroscedasticity). To address these problems, we explored a logarithmic transformation of the model.

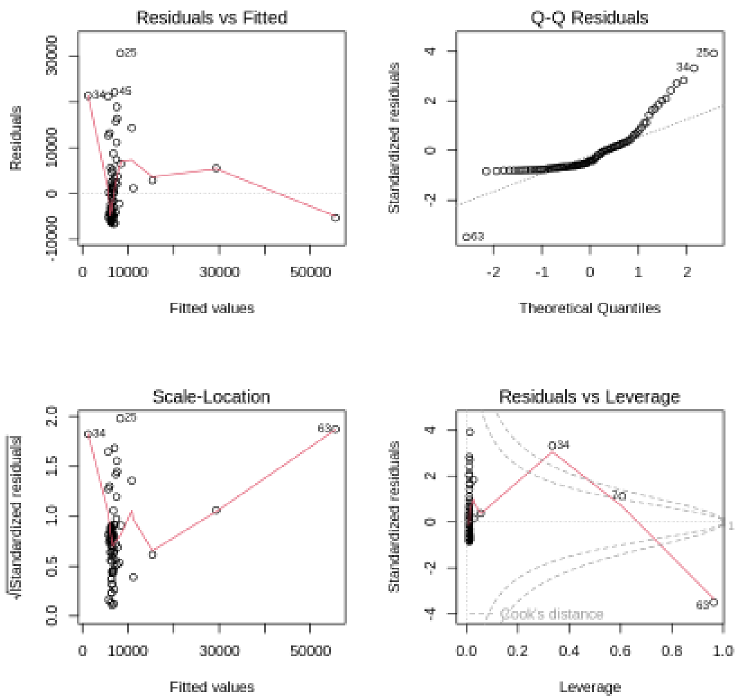

Figure 5: Second model residual plots

4.2 Logarithmic Transformation

We applied a logarithmic transformation to the model to linearize the relationship:

$$third\_model = lm(log(dividends)\ log(debtRepaid) + log(debtChange))$$

Due to the presence of non-positive values in the data, we first shifted the dividends and debt change variables to positive values before applying the log transformation.

The equation of the third model is:

$$log(dividends) = -0.9273 + 0.8413 * log(debtRepaid) + 0.1446 * log(shifted\_debtChange)$$

```r
1   summary(third_model)
```

```
Call:
lm(formula = log10(shifted_dividends) ~ log10(debtRepaid) + log10(shifted_debtChange),
    data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.1939 -0.2749  0.0982  0.3887  1.2092 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)                -0.9273     0.9876  -0.939    0.350    
log10(debtRepaid)           0.8413     0.1081   7.785 1.07e-11 ***
log10(shifted_debtChange)   0.1446     0.1452   0.996    0.322    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7428 on 91 degrees of freedom
Multiple R-squared:  0.4038,	Adjusted R-squared:  0.3907 
F-statistic: 30.82 on 2 and 91 DF,  p-value: 6.016e-11
```

Figure 6: Third model summary

The residual plots of the third model still suggested some issues especially with non-linearity, and heteroscedasticity. Some points on Q-Q plot deviate from the straight line, particularly in the tails, suggesting that the residuals might not be normally distributed. There were a few influential points that might be affecting the model's fit. Additionally, p-value of the Shapiro-Wilk test (3.139e-05) being less than the 0.05 significance level indicates violation of normality, suggesting further transformation is neccessary.

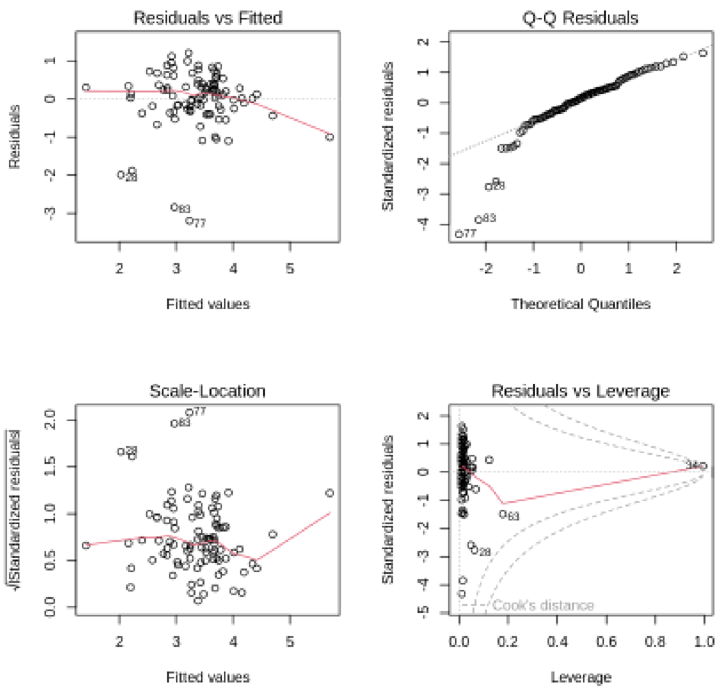

Figure 7: Third model residual plots

4.3 Box-Cox Transformation

Given the remaining issues with the third model, we turned to the Box-Cox transformation to further improve the model fit.

The Box-Cox test returned an optimal lambda of 1.84, which we then used to transform the response variable (dividends).

The fourth model, using the Box-Cox transformed response, yielded much better results:

- Residual plots showed the points touching the diagonal line in the Q-Q plot, and a random scatter in the residuals vs. fitted and scale location plots.
- The Shapiro-Wilk normality test returned a p-value of 0.5359, which is higher than the 0.05 significance level, indicating that the normality assumption was satisfied.
- The skewness test result was -0.35, further confirming the improved normality of the residuals.

```r
library(MASS)
boxcox_result=boxcox(third_model)
```

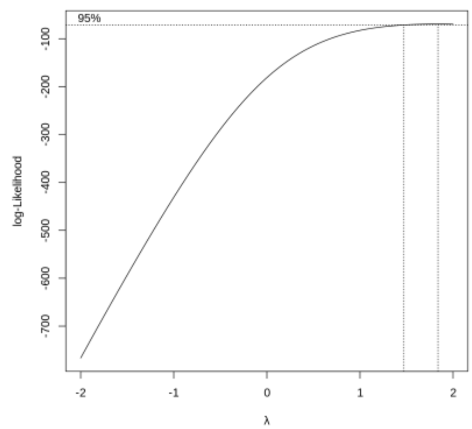

Figure 8: Box-Cox function on third model

```r
# Extract the best lambda value from Box-Cox results
best_lambda <- boxcox_result$x[which.max(boxcox_result$y)]
print(best_lambda)
```

```
[1] 1.838384
```

Best lambda of 1.84 suggests a box cox transformation using this lambda

```r
fourth_model= lm((log10(shifted_dividends))**2 ~ log10(shifted_debtRepaid)+
log10(shifted_debtChange), data=data)
```

Figure 9: Fourth model with box-cox transformation

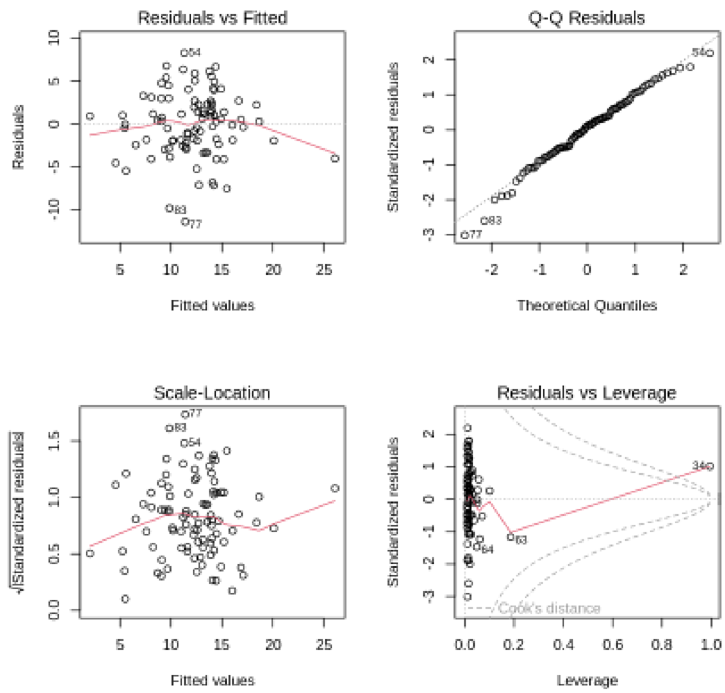

Figure 10: Fourth model residual plots

4.4 Model Adequacy Check

Additional checks on the fourth model showed:

- The mean of the residuals was close to zero (-2.71276610808405e-16), indicating an unbiased model.
- Multicollinearity inspection using VIF returned values around 1 for each regressor, suggesting no issues with multicollinearity.
- The Breusch-Pagan test for non-constant variance returned a p-value of 0.8277, which is much greater than the 0.05 significance level, indicating homoscedasticity.

Overall, the Box-Cox transformed fourth model appears to be the most adequate representation of the data, satisfying the key assumptions of linear regression.

4.5 Interpretation of Results

The results from this study offer valuable insights into the relationships between financial flows and dividends within industries. Through a progressive refinement of regression models, it became evident that not all predictors contribute equally to the variability in dividends. The final model, optimized through transformations and diagnostic evaluations, provides a clear and interpretable framework for understanding these dynamics.

4.5.1 Key Predictors and Their Impact

1. Debt Repaid: - The final model consistently demonstrated that debtRepaid is a significant positive predictor of dividends. - This suggests that industries that prioritize debt repayment are better positioned to allocate resources for shareholder dividends. The coefficient of log(debtRepaid) in the transformed model indicates a strong proportional relationship, highlighting the critical role of debt management in financial planning.

2. Debt Change: - Although the impact of debtChange was less pronounced than debtRepaid, it remained a statistically significant variable across all models. - The positive relationship suggests that increases in debt levels, possibly reflecting strategic borrowing or capital investments, also contribute to dividend distributions. However, this effect is comparatively smaller, underscoring the need for careful debt management.

4.5.2 Model Adequacy and Robustness

The iterative refinement process ensured the final model addressed key statistical concerns:

- Normality of Residuals: The Shapiro-Wilk test in the fourth model returned a p-value of 0.5359, indicating that residuals are normally distributed.
- Homogeneity of Variance: The Breusch-Pagan test yielded a p-value of 0.8277, confirming the absence of heteroscedasticity.
- Multicollinearity: Variance Inflation Factors (VIF) for both predictors were approximately 1, suggesting negligible multicollinearity.
- Mean of Residuals: The mean residual value of -2.71e-16, close to zero, confirms the stability of the model.

These diagnostic results validate the reliability of the final model and its suitability for explaining the variability in dividends.

4.5.3 Comparison Across Models

The stepwise progression from the initial model with all predictors to the final Box-Cox-transformed model highlights the importance of methodological rigor in regression analysis:

- The initial model, while comprehensive, included variables with high p-values, diluting its explanatory power.
- The reduced model eliminated these insignificant predictors, improving clarity and interpretability but revealing residual diagnostic issues.
- Logarithmic transformation addressed some of these issues but fell short of resolving non-linearity and heteroscedasticity.
- The Box-Cox-transformed final model successfully addressed these diagnostic concerns, offering the best fit for the data.

4.5.4 Implications for Financial Strategy

The results emphasize the pivotal role of debt management in shaping dividend policies. Companies with effective strategies for repaying debt appear better equipped to distribute dividends, highlighting the importance of financial discipline. Similarly, measured changes in debt levels, reflecting strategic investments, can also positively influence dividends. These findings provide actionable insights for corporate decision-makers aiming to balance debt obligations with shareholder returns.

5 Conclusion

This study has examined the financial flows and activities across various industries, providing valuable insights into the dynamics of corporate financing decisions. The key findings can be summarized as follows:

- Industry-specific characteristics, such as the number of firms, play a significant role in shaping the observed financial flows, including dividends, buybacks, equity issuance, and debt management.

- There are notable relationships and interdependencies among the different financial variables, suggesting that firms may be employing a range of financing strategies to manage their capital structure and funding needs.

- The regression analysis revealed that factors such as debt repayment and net debt changes can significantly influence the dividend payouts and other financial flows at the industry level.

- The application of Box-Cox transformation was crucial in addressing the non-linearity and non-normality issues, ultimately leading to a more robust and reliable model.

These insights can have important implications for corporate executives, policymakers, and financial analysts. By understanding the industry-specific dynamics of financial flows, decision-makers can better align their strategies and policies with the unique characteristics and financing needs of different sectors. This knowledge can inform discussions around optimal capital structure, dividend policies, debt management, and the role of equity and debt financing in supporting industry-level growth and performance.

Future research could delve deeper into the underlying drivers of these financial flows, exploring the potential impact of macroeconomic conditions, regulatory changes, and other external factors that may be influencing the observed industry-level trends. Additionally, investigating the heterogeneity within industries and the impact of firm-level characteristics on financial decisions could provide further nuance to the understanding of these complex financial dynamics.

Overall, this study contributes to the broader literature on corporate finance and financial management by offering a comprehensive, industry-level perspective on the key financial flows and activities that shape the financial landscape.

6 References

References

[1] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis.*

[2] Jose Vega-guzman (2024), *Regression Analysis.* Class notes from Regression Analysis course material

[3] Aswath Damodaran (2024). Youtube video from the original curator of the data.

[4] Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591-611. This paper discusses the Shapiro-Wilk test, used in the project to assess the normality of residuals.