Statistical Programming 2023/2024

Term One

SINGAPORE MANAGEMENT UNIVERSITY

Group Project

Module Code: ACCT 337

Prepared by:

Full Name	Student ID
Aayisha Basheer Ahamed	01447068
Kohchet-Chua Dean Cedric Lopez	01404709
Jocelyn Anastasia	01405762
Ng Wee Kim	01414947
Perumal Muthukrishnan	01438255

Section: G1

Professor: Sterling Huang

Page Count (excluding cover page, contents page, references and appendix): 20 pages

Contents Page

1 Introduction

Executives are an integral part of a company. They contribute to company performance and activities by providing strategic direction, effective decision-making and efficient resource allocation. With the intention to encourage executives to make decisions and actions that enhance organisation's performance and shareholder value, compensation packages are tailored to drive these executives. Executive compensation packages usually consist of a combination of basic pay, variable pay, bonuses, stock options and other forms of benefits that are subjected to complex performance-based metrics and contractual agreements.

In today's highly competitive labour market, organizations strive to construct an equitable pay structure to attract, retain and motivate talented individuals. However, the big question of how executives should be rewarded and whether the executive pay structure truly aligns with and consequently enhances company performance remains unclear. This report investigates whether there is a relationship between executive pay and company performance.

2 Literature Review

Through an examination of academic articles and studies, the group believes that there is a positive relationship between CEO compensation and the performance of a firm.

In developed cities such as Australia, Japan and Singapore, both the CEO's compensation and the total salaries were higher when a company did better. This can be attributed to factors such as an effective corporate governance framework, which played a significant role in this outcome. (Kayani, Umar & Christopher, 2022)

Research conducted on US publicly traded companies, using the S&P ExecuComp database, suggests a significant and positive correlation between CEO compensation and firm performance (Long Kweh, Qian, et al., 2022). Similarly, a separate study focusing on Nigeria firms has revealed a mutually cause-and-effect relationship between CEO pay and firm performance. This suggests that higher CEO pay serves as both an incentive and a reward for improved firm performance, with higher-performing firms reciprocating by providing better CEO compensation. (Olayini, 2018)

The role of pay is also seen as multifaceted, encompassing aspects that impact employee motivation, performance, and job satisfaction. Higher CEO pay is seen to be a great motivator to CEOs to perform better. At the same time, however, better performing firms often reciprocate the effort of their CEO with high pay ranges, provided they do not face financial limitations. (Kweh et al., 2022)

Despite the known effects of a higher CEO compensation on a firm's performance, research also highlights the negative effects of higher pay on factors such as employee motivation, employee morale, and shareholder's perspectives (Hendriks et al., 2023). Other studies also seem to suggest that a CEO's pay has little to no influence on a firm's performance, and such a factor is insignificant, especially when other aspects of a firm, such as firm size are factored in. (Wijeweera et al., 2021).

However, with increasing discussion on CEO's compensation in recent years and the correlative better performance in firms compensate their CEO's generously, we are more inclined to believe there is a high positive correlation between executive pay and company financial performance.

3 Data Sources

All data collected for the purposes of this project were derived from the Wharton Research Data Services (WRDS). The group believes that datasets from this database will best serve the needs of determining the hypothesis. Using the Compustat – Capital IQ function in WRDS to generate two main reports. The first report the project will be using is the Fundamental Annual Report for North America. A list of all important variables extracted from the database are listed in the Appendix A. The other report that will be used is the Annual Compensation report generated by the database's Execucomp platform. A list of all important variables extracted from the database is listed in the Appendix B.

4 Merging, cleaning and preparing of data

<u>Cleaning company financial data:</u>

1. Analysing and replacing missing NA values in *fyear* : Firstly, *fyear* represents the financial year of the company and analysing the NAs, we identified that the reason is because of the may month cutoff in CompStat database and therefore if the month in *datadate* variable is greater than 5, then the *fyear* will remain the same otherwise, NA *fyear* will be replaced by the previous year. This process removes the NAs in *fyear* and allows for creation of index.

2. Filtering based on industries: As the original dataset consist of companies that are both industrial and financial in nature. As the financial companies follow a different set of financial regulation and representation, we filtered the *indfmt* variable to consist of only industrial companies.
3. Creating index and removing duplicated index: GV key does not uniquely identify a particular observation and therefore we created an *index* variable combining *gvkey* and *fyear* separated by an underscore. Next, the index contained few duplicates as a company may have 2 or more observation in a same *fyear* due to change in financial year end. Therefore, we identified the duplicates and removed them the dataset and now, each observation can be uniquely identified by the *index* variable.

<u>Cleaning executive compensation data:</u>

1. Filtering only to include CEO compensation CEOANN variable indicates the executive that served as CEO for all or most of fiscal year. We filtered the raw dataset by value "CEO" to shortlist the CEO compensation.
2. Converting the GVKEY to character: Secondly, as GVKEY is numerical in this dataset, we converted it to character for ease of combining this dataset to company financial. After converting to character, If the GVKEY is 4 digits, then we added two zeros in front and if its 5 digits then we added one in front of the characters. For example, if the GVKEY 1001, after conversion will be 001001, and If the GVKEY is 50010, after conversion will be 050010.
3. Creating index and removing duplicated index: We followed same steps in company financial dataset to create unique index for each observation and removed the duplicated indices.

<u>Combining the datasets</u>: To combine the company financial and executive compensation data frames, we used a function of the dplyr package called inner_join(). It allows us to to merge the 2 data frames based on common columns, and it retains only the rows where there is a match in both data frames. As we are working with 2 datasets with related information, rows with non-matching values in the specified column should be excluded from the output. With this, we combined the cleaned data frames *comp_data_clean* and *exec_data_clean* with inner_join() by *index* variable, resulting in an output of 44,535 observations.

Calculating ratios: To investigate the relationship between CEO compensation and firm performances, we have chosen various ratios from 4 different categories (profitability, efficiency, leverage and efficiency ratio) as the proxy for firm financial performance.

As a prerequisite to compute ratios like return on assets, return on equity and inventory turnover, which requires an average value of total asset *at*, inventory *invt* and total equity *seq*, it is crucial to incorporate lagged values. Before creating lagged values, it is imperative to arrange the data in ascending order based on *index* variable, as it is a combination of *gvkey* and *fyear*. The mutate function is then employed to generate lagged variable (*'at_lag', 'invt_lag' and 'seq_lag'*) within our data frame. Moreover, an ifelse() function is added to create a lagged values only for rows where the *fyear* is one year ahead of the previous year.

After generating lagged variables, a mutate function from the dplyr package has been used to create new columns representing the calculated ratios. Table 1 provides a brief description of the newly created variables.

Variable Name	**Formula**	**Variable description**
net_profit_margin	Net income/Revenue (*ni/revt*)	A key profitability measure that represents the percentage of profit a company retains from its total revenue after deducting all expenses.
current_ratio	Current asset/ Current liabilities (*act/lct*)	A liquidity measure that assesses a company's ability to meet short-term obligations with its short-term assets.
debt_to_equity_ratio	Total liabilities/Total equity (*lt/seq*)	A leverage measure that evaluates the proportion of a company's financing that comes from debt compared to equity.
asset_turnover	Revenue/Net assets (*revt/(at-lt)*)	An efficiency measure that gauges a company's ability to generate revenue from its assets.

roa	Net income/Average total asset (*ni/((at+at_lag)/2*)	A profitability ratio that measures a company's ability to generate earnings from its total asset.
inventory_turnover	COGS/Average inventory *cogs*/(*(invt+invt_lag)*/2)	An efficiency ratio that assesses how quickly a company sells and replaces its inventory within a specific period.

Table 1: Formula and descriptions of the dependent variables

5 Justification of sample period & sample selection

A total selection of 6 compensation components has been chosen and analysed. A sample period of 22 years from 2000-2022 has also been chosen with the following objective:

Long-term Trends and Events: The 22-year sample period enables the analysis of long-term trends and patterns in CEO compensation. It is also helpful in understanding how trends have evolved and developed over time, especially in response to changes in the economy, societal or regulatory changes. The two-decade time frame also enables the analysis of patterns in CEO compensation due to significant global events, such as financial crises or pandemics.

Selection of 6 Compensation Components: The following 6 main compensation components have been chosen, with the following objectives:

1: Salary: As the fundamental element of compensation, analysis of the salary allows the identification of changes in base pay overtime and provides insights into the wage growth.

2: Bonus: Bonus figures tend to vary significantly over time. Usually with better firm performance, the bonuses increase in value. Analysis of this figure over time uncovers trends, especially in performance-based pay models.

3: Other Compensation (OTHCOMP): The OTHCOMP component includes figures such as other benefits, perks and non-standard compensation items. Analysisng this figure allows the examination of supplementary compensation components over the 20 years.

4: Restricted Stock Grants (RSTKGRNT): As a form on equity-based compensation scheme, RSTKGRNT helps to align CEO's interests with the performance of a company by offering a stake in the company. Analysing this value allows the understanding of how these grants have played a role in attracting, retaining and motivating the employees and CEO to enhance the performance of a company.

5: Option Awards Block Value (OPTION_AWARDS_BLK_VALUE): Evaluating option awards over the time span of two decades allows the examination of how companies use options as a tool to incentivise and motivate the CEOs, and the employees. These stock options are often a significant component of executive compensation packages and are essential tools in gaining wealth. As a result, these options are expected to play a significant role in motivating the financial performance of a company.

6: Long-Term Incentive Plan (LTIP): LTIPs are often designed with the objective of achieving a company's long-term objectives and retaining key management, such as the CEO. Analysing the LTIP figures will highlight whether a company is able to adapt to changing business environments and maintain financial performance. It will also highlight how LTIPs have changed, especially as the industry and market shifts.

6 Justification of any restrictions imposed on the sample

Firstly, we filtered our raw data set according to industry. Our raw data set included data that was both industrial and financial in nature. As such, we filtered the *indfmt* variable to account for only the industrial companies. This is important as companies of different industries follow a different set of financial regulations, accounting standards and reporting requirements. For example, the financial reporting standards for healthcare, a public service, will vary from those of a corporate firm. As such, imposing such sample restrictions based on industries is important as it ensures that the compensation practices are within the context of a specific industry.

To remove duplications of indexes, we have combined *gvkey* with *fyear*. We have also ensured that every index can be uniquely identifiable with the index variable. Moreover, to clean up the executive compensation data, we have filtered the raw dataset by the value of "CEO". To combine our dataset, we have converted our *gvkey* to characters. This makes it easier for the dataset to be combined with the financial data of the company. Character data offers more flexibility when working with data, as they allow for string manipulations and comparisons.

Due to vast differences in company size (total assets), we have restricted regression analysis to use "OTHCOMP/at" as the independent variable. Only the OTHCOMP/at may help alleviate the effects of company size on the performance of a firm. In comparing CEO compensation, even if it appears that two companies may be paying their CEOs the same amount, one company may be paying a more significant portion of its assets. The difference in company size will highlight variances that could influence the profit margin and the CEO compensation.

7 Design of Statistic Model

With the central hypothesis that higher CEO compensation will lead to better firm performance, we aim to analyse the relationship between CEO compensation and company's financial performance and determine which compensation components affect financial performance significantly.

To test the hypothesis, we will employ multi-regression analysis, to assess the relationship between six different compensation components and several financial ratios, including profitability, liquidity, leverage, and operational efficiency. Therefore, we construct the following executive pay-performance sensitivity model with the control variables that may affect financial performance, based on prior studies: where the dependent variable is the financial ratios, and the independent variable is compensation components. However, the compensation components may not be directly related to company performance due to variance in company size. For example, a company with total assets 10,000,000 paying its CEO other compensation of 1,000,000 compared to a company with total assets 5,000,000 paying its CEO other compensation of 1,000,000. Below are the control variables that will affect the respective financial ratios.

<table>
<tr><th>Financial Ratio</th><th>Control Variable</th><th>Purpose</th></tr>
<tr><td rowspan="3">Profitability
- Profit Margin

Efficiency
- Return on Assets</td><td>at</td><td>Setting total assets as a control variable to account for the size of the firm.</td></tr>
<tr><td>debt_to_equity_ratio</td><td>Setting debt to equity ratio as a control variable to account for the variations in firm's capital structure.</td></tr>
<tr><td>ceo_years</td><td>Setting the the number of years the executive has served as the CEO as a control variable to account for the variations in leadership experience.</td></tr>
</table>

Liquidity - Current Ratio	*at*	Setting total assets as a control variable to account for the size of the firm.
	debt_to_equity_ratio	Setting debt to equity ratio as a control variable to account for the variations in firm's capital structure.
	asset_turnover	Setting assets turnover ratio as a control variable to account for the variations in operational efficiency of the firm.
Leverage - Debt to Equity Ratio	*at*	Setting total assets as a control variable to account for the size of the firm.
	net_profit_margin	Setting net profit margin as a control variable to account for the variations in profitability and financial health of the firm.
	asset_turnover	Setting assets turnover ratio as a control variable to account for the variations in operational efficiency of the firm.

Moreover, we have constructed a separate model to include firm and year fixed effects, as it may help to control for time-invariant characteristics of firms that might be correlated with the independent variables and to control for common shocks or trends (e.g., economic recession) affecting all firms each year. This helps to account for the unobserved heterogeneity at both firm and time levels, thereby it may provide a more robust estimates for the coefficient of interest.

$$Financial\ Ratio = \beta 0 + \beta 1 \times Compensation\ Component + Control\ Variables$$

β0 is the intercept, representing the baseline financial ratio when the CEO compensation component and all control variables are zero. On the other hand, β1 is the coefficient that measures the sensitivity of the financial ratio to the changes in CEO compensation component. A positive β1 will suggest that higher CEO compensation component is associated with an increase in the financial ratio, while a negative value implies the opposite.

We removed the observations whereby total assets were NA based on the assumption that the total assets are unlikely to be zero for a company. In addition, over 9000 values of inventory turnover are infinite as the actual inventory values that were NA converted to zero during the data cleaning process. Therefore, we removed the infinite values in inventory turnover as it may affect the regression model later.

Next, we analyzed the outliers by plotting the *fyear* in x-axis and various financial ratios including profit margin, debt to equity ratio, inventory turnover, asset turnover, current ratio and return on asset (roa) in y-axis to identify the distribution and any outliers. After analyzing the outliers using ggplot, we used quantile to remove the top and bottom 0.01 quantile to remove the potential outliers affecting the regression model. In addition, for CEO years, we limited it to 20 years as the density plot shows a significant decrease in CEO years past 20.

8 Regressing Salary with the Dependent variables

All regression reports are attached in appendix C for reference.

8.1 IV: Salary / Total asset, DV: Net profit margin

Based on the regression result, model 1 has a slope of *salary_at* (-0.01187) indicates a negative relationship, where every one-unit increase in salary/total asset, the model predicts a decrease of approximately 0.01187 units in profit margin. The associated p-value (<0.1) suggest that this slope is statistically significant, meaning that there is evidence to reject the null hypothesis that the true slope is zero.

However, the adjusted r-squared of 0.0348 suggests that the model might not effectively explains the variation in the dependent variable (profit margin). Upon adding firm and year fixed effects, there is a substantial improvement in the adjusted R-squared to 0.329, and residual standard error reduces to 0.116. This suggests that the model with fixed effects explains a much larger proportion of the variability in the DV and indicates that there is better precision in predicting the dependent variable. Nevertheless, the p-value rises to >0.1, suggesting a reduction in degrees of freedom and a potential overfitting issue, which can lead to less stable estimates.

8.2 IV: Salary / Total asset, DV: Return on assets

The p-value across 3 models for the salary/total asset variable remain less than 0.1, indicating statistical significance. When comparing the 3 models, we can observe that after adding firm and year fixed effect, the r-squared increases from 0.0379 to 0.329 and the residual standard error decreases to 0.0701. Similar to the previous regression, the negative coefficient for *salary_at* suggests a negative relationship with return on asset, although the effect size is small.

8.3 IV: Salary / Total asset, DV: Current ratio

The model without fixed effects demonstrates a positive and statistically significant relationship between *salary_at* and Current Ratio (slope:0.1109 and $p<0.1$). Despite this, the models yield an extremely low adjusted r-squared at 0.108. With fixed effects, it alters the dynamics, resulting in a negative relationship (slope: -0.04943). The change in sign of the coefficient suggests that there are unobserved factors captured by fixed effects play a crucial role in understanding the relationship between salary/total asset and current ratio. Comparing with the previous regressions with salary and roa, though model 3 yields the highest adjusted r-squared, it returns the highest AIC indicating a weaker relationship fit between bonus and current ratio.

8.4 IV: Salary / Total asset, DV: Debt to equity ratio

Regression model 1 and 2 is statistically significant in explaining the variation in debt-to-equity ratio. The negative slope of salary/total asset (-0.1169) suggests an inverse relationship with debt-to-equity ratio. In other words, as a salary/total asset increase, the debt-to-equity ratio tends to decrease. When comparing to Model 3, the adjusted R squared increases to 0.795 and residual standard error reduces to 1.16. The overall model, as well as the relationship between IV and DV, with fixed effects stays statistically significant at the 0.1 level. However, model 3 shows that adding fixed effects has reduced the size of the effect of salary/total asset on the debt-to-equity ratio with the slope of -0.0393.

In summary, the regression analysis indicates a weak association between salary and financial firm performance, evidenced by a slope coefficient close to 0. When examining the regression of salary/at against profit margin or return on asset, a statistically significant negative correlation is observed, but the model's ability to explain profit margin and return on asset variability is limited. Conversely, when using debt-to-equity ratio as the dependent variables, the p value remains below 0.1 level, and adjusted r squared is high with relatively low residual error. With this, we can deduce that the regression model suggests a reasonable correlation between CEO salary vs debt to equity ratio:

Debt-to-equity-ratio = - 0.039374 * salary_at + 0.00000399 * total asset - 0.63801114 * net profit margin + 0.76662036 * asset turnover

9 Regressing Bonus with the Dependent variables

9.1 IV: Bonus / Total asset, DV: Net profit margin

The negative slope coefficient for bonus/total asset (-0.005952) is statistically significant with a p-value less than 0.1, indicating an inverse relationship between the IV and DV. However, the models exhibit a low adjusted R-squared of 0.0154. The inclusion of firm and year fixed effects enhances the adjusted r-squared to 0.33, though remain to be moderately low. The change of IV coefficient from negative to positive (0.008249) may be attributed to unobserved factor specific to each firm or year that were not accounted for in the initial model. With this, we can conclude that model 3 pertains to be the best model when regressing bonus/total asset against profit margin.

9.2 IV: Bonus / Total asset, DV: Return on assets

Model 1 and 2 exhibits a slope coefficient of -0.002121 with its statistically significant p-value, suggesting an adverse association between CEO salary and return on asset. In alignment with regression outcome in 9.1, it is still necessary to include fixed effects to capture unobserved factors, given the models' exceedingly low adjusted R-squared values of 0.0143. When adding fixed effects, there is also a shift in sign and magnitude of the slope in model 3 to 0.004766. Nevertheless, Model 3 results in an elevated P-value>0.1 for the independent variable, indicative of a reduction in degrees of freedom and a potential overfitting concern, leading to less stable estimates.

9.3 IV: Bonus / Total asset, DV: Current ratio

We can deduce that there is a consistent pattern observed in the analysis where salary was regressed on the current ratio. Adding fixed effects alters the dynamics, resulting in a negative relationship between bonus and current ratio (from 0.09794 to –0.03978). Moreover, there is a substantial improvement in the model's explanatory power and precision in predicting the current ratio, as shown in the improvement in adjusted r squared to 0.717 and residual standard error to 0.795. Comparing to previous regression model with profit margin and roa, this model with fixed effects returns the highest adjusted r-squared while also maintaining the p-value below threshold. This regression outcomes suggest a notable and statistically significant negative association between CEO bonus and current ratio.

9.4 IV: Bonus / Total asset, DV: Debt to equity ratio

When examining the correlation between bonus and debt to equity ratio, it is evident that Model 1 and 2 yield a moderate level of adjusted R-squared (0.464), with a p-value<0.1. In Model 3, though the fixed effect has increased the adjusted r-squared to 0.795, there is a simultaneous increase of the p-value above 0.1, suggesting that the model may not be statistically significant.

Upon analysing the various regression results aimed at assessing the correlation between CEO bonus and financial performance, it becomes apparent that a very weak relationship exists, this is indicated by an exceedingly low adjusted r squared and a negative coefficient close to 0. Despite the incorporation of fixed effects to enhance the model's reliability by accounting for the unobservable factors, the resultant increase in the p-value raises concerns about its robustness. Nevertheless, when the dependent variable is current ratio, the regression findings appear to offer a more substantiated and viable basis for investigating the relationship between CEO bonus and this financial metric.

Current ratio = -0.03978 * bonus_at - 0.000001211 * total asset + 0.005079 * debt to equity ratio - 0.03595 * asset turnover

10 Regressing Other Compensation with the Dependent variables

10.1. IV: Other compensation/ Total asset, DV: Net profit margin

Based on the results under (IV + Control Variables), the slope of the *othcomp_at* in the regression output (-0.0082) indicates a negative relationship, where every one-unit increase in other compensation/total asset, the model predicts a decrease of approximately 0.0082 units in profit margin. The associated p-value (<0.1) suggest that this slope is statistically significant, and that there is evidence to reject the null hypothesis that the true slope is zero. In other words, the other compensation/total asset variable appears to have an impact on profit margin in this model.

However, the adjusted r-squared of 0.0183 suggests that the only 1.83% of variation in the profit margin can be explained by other compensation / total assets. Upon adding firm and year fixed effects, there is a substantial improvement in the adjusted R-squared to 0.329, and residual standard error reduces to 0.116 indicating the model with control variable and fixed effects could better predict the variations in profit margin.

10.2. IV: Other compensation / Total asset, DV: Return on asset

Based on the result, the p-value across 3 models for the other compensation/total asset variable remain less than 0.1, indicating statistical significance. When comparing the 3 models, we can observe that after adding firm and year fixed effect, the r-squared increases substantially from 0.0199 to 0.391 and the residual standard error decreases to 0.0702. This suggests that the model with fixed effects explains a much larger proportion of the variability in the DV and indicates that there is better precision in predicting the dependent variable. Similar to the previous regression, the negative coefficient for other compensation/Total Asset suggests a negative relationship with return on asset, although the effect size is relatively small.

10.3. IV: Other compensation / Total asset, DV: Current ratio

Based on the result of this regression model, we can observe that the adjusted r square of IV + fixed effects model is highest with 71.7% however, the relationship between other compensation/ at and current ratio in this model may not be statistically significant as the P-value of the other compensation/at is >0.1. However, the IV + control variables model is statistically significant with <0.1 P value for other compensation/at and F statistic. Through this regression model, we can observe that other compensation/at is positively correlated to current ratio and over 9.32% of variance in current ratio might be explained by other compensation/at based on adjusted r square. In addition, compared to previous regression model with profit margin and roa, this model returns the highest AIC indicating a weaker relationship fit between other compensation / at and current ratio.

10.4. IV: Other compensation / Total asset, DV: Debt to equity ratio

Regression model 1 and 2 is statistically significant in explaining the variation in debt-to-equity ratio. The negative slope of other compensation/total asset suggests an inverse relationship with debt-to-equity ratio and these models indicates that 46.8% of the variability in the debt-to-equity ratio can be explained by other compensation / at. When fixed effects are added, the adjusted R squared increases to 0.794 and residual standard error reduces to 1.17 and the model stays statistically significant at the 0.1 level. However, model 3 shows that adding fixed effects has reduced the size of the effect of salary/total asset on the debt-to-equity ratio (as indicated by the smaller slope).

Furthermore, for all 4-regression model with other compensation /at and the 4 different financial ratios, the KS static of residuals is higher than the 1.36 kscritvalue indicating sufficient evidence that the residuals did not meet the normality assumption of the simple linear regression.

In summary, the regression model of other compensation/at against profit margin and return on asset leads to statistically significant negative correlation, however the model's ability to explain variability of profit margin and return on asset is limited due to lower adjusted r square. Furthermore, regressing other compensation/at against current ratio exhibits a high adjusted r squared of 0.719, however, the p value exceeds the 0.1 threshold, suggesting statistical insignificance in the corelation. Conversely, when using debt-to-equity ratio as the dependent variable, the p value remains below 0.1 level, and adjusted r squared is high at 0.794 with relatively low residual error. With this, we can deduce that the regression model suggests a reasonable correlation between CEO other compensation and debt to equity ratio, indicating that other compensation can be utilized to predict a company's leverage through a given model:

Debt to equity-ratio = - 0.02106 * othcomp_at + 0.00000394 * total asset - 0.6299 * net profit margin + 0.76648 * asset turnover

11 Regressing Restricted Stock Grant with the Dependent variables

11.1 IV: RSTKGRNT, DV: Net Profit Margin

Based on the regression results, the slope of the *RSTKGRNT* (IV+CV) suggests a meagre relationship between *RSTKGRNT* and the net profit margin, whereby the model predicts an increase of approximately 0. 000002162 units in profit margin. The p-value of <0.1 also highlights the statistically insignificant slope and hence, there is evidence to reject the null hypothesis that the slope is zero. The positive adjusted r-square of 0.0147 highlights that 1.47% of the variation in profit margin can be attributed to the *RSTKGRNT*. With the addition of the firm and fixed year effects however, there is an improvement, indicating that the model with the CV and fixed effects can better predict the variations in the profit margin.

11.2 IV: RSTKGRNT, DV: Return on Asset

Most p-values across the 3 models remain below 0.1, highlighting statistical significance. When the firm and year fixed effects are added, it can be observed that there is a significant increase in the r square from 0.00517 to 0.388, and the residual error also falls from 0.0897 to 0.0703. More significantly, the p value becomes more than 0.1, highlighting that the figures become statistically insignificant, signaling overfitting. As such, this highlights that the latter model is better able to predict the effects of the DV. The positive but meagre correlation between *RSTKGRNT* and the return on assets suggest a weak relationship between these two factors.

11.3 IV: RSTKGRNT, DV: Current Ratio

Most p-values in the 3 models have remained below 0.1, suggesting statistical significance. When the firm and fixed year effects were added, the adjusted r square saw significant improvements from 0.0936 to 0.717., while the residual standard error fell from 1.42 to 0.796. More notably however, the p value increases to above 0.1, indicating statistical insignificance. The slope here, however, is negative, suggesting a negative relationship between the current ratio and the RSTKGRNT. As such, this factor has a weak relationship with the *RSTKGRNT*.

11.4 IV: RSTKGRNT, DV: Debt to Equity Ratio

This model has resulted in the highest adjusted r square of IV + fixed effects with 79.5%, with a p-value of >0.1, suggesting statistical insignificance. The debt-to-equity ratio is the only factor whereby the figures become statistically insignificant. The r square value also experiences a significant increase from 0.462 to 0.795, and the residual error falls from 1.89 to 1.17 suggesting that the values are more precise with the addition of the firm and fixed year effects. This model has returned the highest AIC, indicating a weaker relationship fit between *RSTKGRNT* and the debt-to-equity ratio.

The KS statistic for all 4 regression models with *RSTKGRNT* and the 4 different financial ratios have resulted in values below the 1.36 kscritvalue, indicating that there is sufficient evidence to suggest that the residuals met the normality assumption of the simple linear regression. The generally negative or near 0 value coefficients however prove that there is very little to no relationship between RSTKGRNT and the 4 financial ratios. When the return on asset is used, the p value remains below 0.1 and the adjusted r square is relatively high, with a lower residual error. Hence, its regression model will be able to highlight a reasonable correlation.

Return on Asset = 0.000001065 * RSTKGRNT + 0.000042918 * total asset - 0.000997851 * net profit margin + 0.000872509 * asset turnover

12 Regressing Stock Option Schemes with the Dependent variables

12.1 IV: OPTION_AWARDS_BLK_VALUE, DV: Net Profit Margin

Based on the regression analysis, the IV + CV + Fixed Effects Model has the highest Adjusted R Square at 0.329 and lowest residual standard error at 0.116, compared to the other models at 0.0147. and 0.1409 respectively However, the P-Value is greater than 0.1, which implies that the model is not statistically significant. Using the F-statistic shows that the model is statistically significant, though the effect between the independent variable and dependent variable is very minimal with a slope of -0.0000005164. Therefore, the model implies a very weak negative relationship between net profit margin and stock options.

12.2 IV: OPTION_AWARDS_BLK_VALUE, DV: Return on Asset

Based on the regression analysis, the IV + CV + Fixed Effects Model has the highest Adjusted R Square at 0.39 and lowest residual standard error at 0.0702, compared to the other models at 0.0142 and 0.0893 respectively. This model also has a reasonable P-Value that is under 0.1, indicating a level of statistical significance, though with a very small negative slope, there is a weak negative regression between both variables. This regression has a greater impact than its relationship with net profit margin given the higher R Square and lower residual standard error and P-Value.

12.3 IV: OPTION_AWARDS_BLK_VALUE, DV: Current Ratio

Based on the regression analysis, the IV + CV + Fixed Effects Model has the highest Adjusted R Square at 0.717 and lowest residual standard error at 0.796, compared to the other models at 0.0924 and 1.43 respectively. The model is more fit for the variables than that shown in 12.1 and 12.2, with the P-Value showing that the model is statistically significant, despite the slope being small. Therefore, the regression shows that there is a positive relationship between the variables, although the slope may not show an impactful change. Without fixed effects, the slope would be negative, meaning the fixed effect has importance to the model.

12.4 IV: OPTION_AWARDS_BLK_VALUE, DV: Debt to Equity Ratio

Based on the regression analysis, the IV + CV + Fixed Effects Model has the highest Adjusted R Square out of all the regressions at 0.795, as well as having a significant P-Value that is under 0.1, compared to other models at 0.4618. It also has the lowest residual standard error the three models at 1.17, though this statistic is higher than the standard error in other regressions. Though the slope may be minimal, this regression seems to show the best fit between the variables. Hence, the regression shows a negative relationship between variables despite the small slope.

The KS statistic for all 4 regression models with *OPTION_AWARDS_BLK_VALUE* and the 4 different financial ratios have resulted in values below the 1.36 kscritvalue, indicating that there is sufficient evidence to suggest that the residuals met the normality assumption of the simple linear regression. In summary, the regression model indicates a weak association between *OPTION_AWARDS_BLK_VALUE* and financial firm performance, evidenced by a slope coefficient close to 0. A statistically significant negative correlation is observed when examining the regression of *OPTION_AWARDS_BLK_VALUE* against net profit margin, return on assets and debt-to-equity-ratio. Yet, a statistically significant positive correlation is observed when examining the regression of *OPTION_AWARDS_BLK_VALUE* against current ratio. However, the model's ability to explain profit margin, current ratio and return on assets variability is limited. Conversely, when using debt-to-equity ratio as the dependent variables, the p value remains below 0.1 level, and adjusted r squared is high with relatively low residual error.

Debt-to-equity-ratio = -0.00000256 * OPTION_AWARDS_BLK_VALUE+ 0.00000391 * total asset - 0.62493929 * net profit margin + 0.76633488 * asset turnover

13 Regressing Long Term Incentive Payment with the Dependent variables

13.1 IV: LTIP/Total Assets, DV: Net Profit Margin

The adjusted R square of 0.0145 from Model 1 and 2 indicates that 1.45% of the variation in profit margin is explained by the independent variable. The slope coefficient of 0.005357836 from Model 1 and 2 also indicates that the relationship between *LTIP* and profitability is positive.

With a higher R square of 0.328, Model 3 is the best model when regressing *LTIP* against profitability. Model 3 also has a lower residual standard error of 0.116 than in Model 1 and 2. A lower residual standard error would mean that the model's predictions are on average closer to the actual observed values, which implies a more accurate and precise model. Since the p value is more than 0.1, *LTIP* is unlikely to be a significant predictor of firm's profitability. However, the F statistic associated with the overall model is less than 0.1, it implies that the model as a whole is statistically significant.

13.2 IV: LTIP/Total Assets, DV: Return on Asset

The adjusted R square of 0.0144 from Model 1 and 2 indicates that 1.44% of the variation in return on assets is explained by the independent variable. The slope coefficient of 0.0085003758 from Model 1 and 2 also indicates that the relationship between LTIP and operation efficiency is positive.

With a higher R square of 0.39, Model 3 is the best model when regressing *LTIP* against return on assets. Model 3 also has a lower residual standard error of 0.0702 than in Model 1 and 2, implying that Model 3's predictions are on average closer to the actual observed values. Since the p value is more than 0.1, *LTIP* is unlikely to be a significant predictor of firm's operation efficiency. However, the F statistic associated with the overall model is less than 0.1, it implies that the model is statistically significant.

13.3 IV: LTIP/Total Assets, DV: Current Ratio

The adjusted R square of 0.00932 from Model 1 and 2 indicates that 0.932% of the variation in profit margin is explained by the independent variable. The slope coefficient of -0.230164345 from Model 1 and 2 also indicates that the relationship between *LTIP* and liquidity is negative.

With a higher R square of 0.717 and a lower residual standard error of 0.796 than in Model 1 and 2, Model 3 is the best model when regressing *LTIP* against current ratio. Since the p value is more than 0.1, *LTIP* is unlikely to be a significant predictor of firm's liquidity. However, the F statistic associated with the overall model is less than 0.1, it implies that the model as a whole is statistically significant.

13.4 IV: LTIP/Total Assets, DV: Debt to Equity Ratio

The adjusted R square of 0.0462 from Model 1 and 2 indicates that 4.62% of the variation in profit margin is explained by the independent variable. The slope coefficient of 0.033826292 from Model 1 and 2 also indicates that the relationship between *LTIP* and profitability is positive.

On the other hand, Model 3 has a higher R square of 0.795 and a lower residual standard error of 1.17 than Model 1 and 2. This implies that Model 3 is the best model when regressing *LTIP* against profitability. Since the p value is less than 0.1, *LTIP* is unlikely to be a significant predictor of firm's leverage. However, the F statistic associated with the overall model is less than 0.1, it implies that the model as a whole is statistically significant.

In summary, the regression model indicates a weak association between *LTIP* and financial firm performance, evidenced by a slope coefficient close to 0. A statistically significant negative correlation is observed when examining the regression of *LTIP*/at against profit margin or current ratio. Yet, a statistically significant positive correlation is observed when examining the regression of *LTIP*/at against return on assets. However, the model's ability to explain profit margin, current ratio and return on assets variability is limited. With the p value of *LITP_at* more than 0.1 for all the models, we conclude that *LITP* is unlikely to be a significant predictor of firm's performance.

14 Conclusion

In the analysis of our IV with our 6 compensation components mentioned in section 5, we can predict the ratios of the variables in concluding our research. Based on our regression models, we observe that only a few regressions have a r square of IV of over 75% with P-values below 0.1, suggesting a statistically significant relationship between the independent variable (CEO compensation) and the dependent variable (firm financial performance). One such example is the relationship between CEO other compensation and debt to equity ratio. However, the regression model of other compensation with other financial ratios remains insignificant. Likewise, the regression models of other IVs remain largely insignificant. As such, we only observe weak relationship between CEO's compensation components and the financial performance of the company and in certain cases, it was a negative relationship therefore we reject our null hypothesis that there is a positive association between executive compensation and the financial performance of a firm. In reaching in this conclusion, we faced certain limitations including size of dataset, outliers, missing values and uncontrollable external factors affecting the regression results. The size of the dataset may be a limitation in regression analysis, a small dataset may not capture the full variability of the underlying population, leading to imprecise estimates and biased results. In such cases, the statistical power of the analysis is reduced and limiting the true relationship between variables. Moreover, while addressing outliers is crucial for improving model robustness, the choice of quantiles threshold for outliers' removal can impact the results significantly. There is also a large number of NA values are removed, reducing the effective sample size which can impact the statistical power of the analysis.

15 References

Hendriks, Martijn, et al. (2022). "The Influence of CEO Compensation on Employee Engagement." Review of Managerial Science, vol. 17, no. 2, Feb. 2023, pp. 607–33. DOI.org (Crossref), https://doi.org/10.1007/s11846-022-00538-4.

Kayani, Umar Nawaz, and Christopher Gan.(2022). "Executive Compensation and Firm Performance Relationship." Review of Pacific Basin Financial Markets and Policies, vol. 25, no. 01, Mar. 2022, p. 2250008. DOI.org (Crossref), https://doi.org/10.1142/S0219091522500084.

Kweh, Qian Long, et al. (2022). "CEO Compensation and Firm Performance: Evidence from Financially Constrained Firms." Research in International Business and Finance, vol. 61, Oct. 2022, p. 101671. DOI.org (Crossref), https://doi.org/10.1016/j.ribaf.2022.101671

Olaniyi, C. O., Obembe, O. B., & Oni, E. O. (2017). Analysis of the Nexus between CEO Pay and Performance of Non-Financial Listed Firms in Nigeria. *African Development Review*, *29*(3), 429–445. https://doi.org/10.1111/1467-8268.12279

Wijeweera, Albert, et al. (2022). "Executive Remuneration and Firm Financial Performance: Lessons from Listed Companies in Australia and Implications for Their APEC Counterparts." Asia. Pacific Business Review, vol. 28, no. 2, Mar. 2022, pp. 260–72. DOI.org (Crossref), https://doi.org/10.1080/13602381.2022.2013614.

16 Appendix

Appendix A: Variable description (comp_financial.csv)

Variable name	Definition
gvkey	Unique identifier for each company in the database
datadate	Annual close of fiscal period. dd/mm/yy format
fyear	It refers to the Company's accounting year. If the current fiscal year-end month falls in January- May, fyear will be current calendar year – 1. If the date is 31/5/2000, fyear is 1999. If the date is 31/12/2000, fyear is 2000
cusip	Another firm unique identifier for each company on Compustat. nine-digit code: the first six digits identify the issuer, the seventh and eighth digit identify the issue, and the ninth digit is the check digit. If there's X in the last 6 digit of the of the CUSIP means the issue is inactive
conm	Company name
city	City Location. "City" typically refers to the city where the company is located or has its headquarters.
sic	Standard Industrial Classification (SIC) Code. The SIC code is a classification system used to categorize industries based on their primary economic activities. Each industry is assigned a specific SIC code. We use SIC codes to categorize companies by industry and analyze industry-specific trends, to compare financial performance and ratios within the same industry using SIC codes.
state	State Location.
fyr	Month in the datadate column. If datadate is 31/5/00, fyr is 5
act	**Total Current Asset** A component of total asset in (AT) A sum of Cash and Short-Term Investments (CHE), Receivables - Total (RECT), Inventories - Total (INVT), Current Assets - Other - Total (ACO)
at	**Total Asset**

	Sum of current assets (ACT) plus net property, plant, and equipment (PPENT) plus other noncurrent assets, including intangible assets (INTAN), deferred items and investments and advances (IVAEQ, IVAO).
ch	**Cash** A component of cash and short-term investment (CHE) that includes cash on hand, bank and finance company receivables, bank drafts, etc. Not available for banks.
cogs	**Cost of goods sold** Available for north-american banks in the industrial format Represents all costs directly allocated by the company to production, such as material, labor and overhead.
invt	**Total Inventories** This variable represents the total value of all inventories held by the company. Inventories include raw materials, work-in-progress, and finished goods that the company plans to sell.
lct	**Total Current Liabilities** This variable represents the total value of a company's current liabilities, which are obligations or debts expected to be settled within one year. It includes accounts payable, short-term loans, and other short-term financial obligations.
lt	**Total Liabilities** This variable represents the total value of all the company's liabilities, including both current (short-term) and long-term liabilities. It encompasses all financial obligations the company owes to creditors and others.
ni	**Net Income** Net income is the bottom line of a company's income statement and represents the profit or loss after accounting for all expenses, taxes, and interest. It is a measure of the company's overall financial performance.
revt	**Total Revenue**

	This variable represents the total revenue generated by the company. It typically includes revenue from sales, services, and other operating activities before deducting expenses.
seq	**Total Stockholders' Equity** "SEQ" stands for the total stockholders' equity of the company, which is the difference between the total assets and total liabilities. It represents the ownership interest in the company's assets.

Appendix B: Variable description (exec_comp.csv)

CEOANN	**Annual CEO Flag** It indicates that this executive served as CEO for all or most of the indicated fiscal year.
CO_PER_ROL	**ID Number for each Executive** It represents the unique ID number for each executive
BECAMECEO	**Date Became CEO** It indicates the date the individual became CEO
SALARY	**Salary** It is the dollar value of the base salary earned by the named executive officer during the fiscal year.
BONUS	**Bonus** It is the dollar value of a bonus earned by the named executive officer during the fiscal year.
OTHCOMP	**Other compensation** It is the other compensation received by the director including perquisites and other personal benefits, contributions to defined contribution plans (e.g. 401K plans), life insurance premiums, gross-ups and other tax reimbursements, discounted share purchases, consulting fees, awards under charitable award programs etc.

RSTKGRNT	**Restricted Stock Grant** It is the value of restricted stock granted during the year (determined as of the date of the grant).
OPTION_AWARDS_BLK_VALUE	**Stock Options Awarded** It is the aggregate value of stock options granted to the executive during the year as valued using Standard & Poor's Black-Scholes methodology.
LTIP	**Long Term Incentive Pay** It is the amount paid out to the executive under the company's long-term incentive plan. These plans measure company performance over a period of more than one year (generally three years).

Appendix C: Regression results

8.1 – Regression results of IV: Salary / Total asset, DV: Net Profit Margin

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0348	0.0348	0.329
Residual standard error	0.139	0.139	0.116
Salary_at(slope)	-0.01187	-0.01187	-0.003836
Salary_at (P value)	<0.1	<0.1	>0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		-19084	
RMSE		0.1386	

kscritvalue: 0.01033, KS statistic: 0.19

8.2 – Regression results of IV: Salary / Total asset, DV: Return on Asset

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0379	0.0379	0.329
Residual standard error	0.0882	0.0882	0.0701
Salary_at(slope)	-0.008151	-0.008151	-0.004102

Salary_at (P value)	<0.1	<0.1	<0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		-34960	
RMSE		0.08696	

kscritvalue: 0.01033, KS statistic: 0.1243

8.3 – Regression results of IV: Salary / Total asset, DV: Current Ratio

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.108	0.108	0.719
Residual standard error	1.41	1.41	0.794
Salary_at(slope)	0.1109	0.1109	-0.04943
Salary_at (P value)	<0.1	<0.1	<0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		61129	
RMSE		1.436	

kscritvalue: 0.01033, KS statistic: 0.134

8.4 – Regression results of IV: Salary / Total asset, DV: Debt to Equity Ratio

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.468	0.468	0.795
Residual standard error	1.88	1.88	1.16
Salary_at(slope)	-0.1169	-0.1169	-0.0393
Salary_at (P value)	<0.1	<0.1	<0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		70952	
RMSE		1.983	

kscritvalue: 0.01033, KS statistic: 0.16

9.1– Regression results of IV: Bonus / Total asset, DV: Net Profit Margin

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0154	0.0154	0.33
Residual standard error	0.141	0.141	0.116
bonus_at(slope)	-0.005952	-0.005952	0.008249
bonus_at (P value)	<0.1	<0.1	<0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		-18739	
RMSE		0.1393	

kscritvalue: 0.01033, KS statistic: 0.1974

9.2– Regression results of IV: Bonus / Total asset, DV: Return on Asset

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0143	0.0143	0.389
Residual standard error	0.0893	0.0893	0.0703
bonus_at(slope)	-0.002121	-0.002121	0.004766
bonus_at (P value)	<0.1	<0.1	>0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		-34541	
RMSE		0.08741	

kscritvalue: 0.01033, KS statistic: 0.1296

9.3– Regression results of IV: Bonus / Total asset, DV: Current Ratio

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0945	0.0945	0.717
Residual standard error	1.42	1.42	0.795
bonus_at(slope)	0.09794	0.09794	-0.03978

bonus_at (P value)	<0.1	<0.1	<0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		61396	
RMSE		1.449	

kscritvalue: 0.01033, KS statistic: 0.1357

9.4– Regression results of IV: Bonus / Total asset, DV: Debt to Equity Ratio

	Model 1: IV + CV	Model 2: Stepwise/Backward/Forwar	Model 3: IV + CV + FE
Adjusted R square	0.464	0.464	0.795
Residual standard error	1.88	1.88	1.17
bonus_at(slope)	-0.1621	-0.1621	-0.02845
bonus_at (P value)	<0.1	<0.1	>0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		71081	
RMSE		1.995	

kscritvalue: 0.01033, KS statistic: 0.1547

10.1 - Regression result of IV: Other compensation / total assets, DV: Net Profit Margin

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0183	0.0183	0.329
Residual standard error	0.1406	0.1406	0.116
othcomp_at(slope)	-0.0082	-0.0082	-0.0041734
Othcomp_at (P value)	<0.1	<0.1	<0.1 and >0.01
F statistic (P value)	<0.1	<0.1	<0.1
AIC		-67951	
RMSE		0.1387	

kscritvalue: 0.01033, KS statistic: 0.1967

10.2. IV: Other compensation / Total asset, DV: Return on Asset

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0199	0.0199	0.391
Residual standard error	0.0890	0.0890	0.0702
othcomp_at(slope)	-0.0064	-0.0064	-0.003293
othcomp_at (P value)	<0.1	<0.1	<0.1 and >0.01
F statistic (P value)	<0.1	<0.1	<0.1
AIC		-83798	
RMSE		0.087	

kscritvalue: 0.01033, KS statistic: 0.1304

10.3. IV: Other compensation / Total asset, DV: Current Ratio

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0932	0.0932	0.717
Residual standard error	1.42	1.42	0.796
othcomp_at(slope)	0.0404	0.0404	-0.000506
othcomp_at (P value)	<0.1	<0.1	>0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		12261	
RMSE		1.448	

kscritvalue: 0.01033, KS statistic: 0.1384

10.4. IV: Other compensation / Total asset, DV: Debt to Equity Ratio

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.4622	0.4622	0.794
Residual standard error	1.88	1.88	1.17
othcomp_at(slope)	-0.0514	-0.0514	-0.02091

othcomp_at (P value)	<0.1	<0.1	<0.1 and >0.01
F statistic (P value)	<0.1	<0.1	<0.1
AIC		21,970	
RMSE		1.994	

kscritvalue: 0.01033, KS statistic: 0.2424

11.1– Regression results of IV: RSTKGRNT, DV: Net Profit Margin

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0147	0.0147	0.329
Residual standard error	0.141	0.141	0.116
RSTKGRNT (slope)	0.000002162	0.000002162	0.000001928
RSTKGRNT (P-Value)	<0.1	<0.1	<0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		-18736	
RMSE		0.1391	

kscritvalue: 0.01033, KS Statistic: 0.1975

11.2– Regression results of IV: RSTKGRNT, DV: Return on Asset

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.00517	0.00517	0.388
Residual standard error	0.0897	0.0897	0.0703
RSTKGRNT (slope)	0.000002254	0.000002254	0.000001065
RSTKGRNT (P-Value)	<0.1	<0.1	>0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		-34387	
RMSE		0.08773	

kscritvalue: 0.01033, KS Statistic: 0.1298

11.3– Regression results of IV: RSTKGRNT, DV: Current Ratio

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0936	0.0936	0.717
Residual standard error	1.42	1.42	0.796
RSTKGRNT (slope)	-0.000051612	-0.000051612	0.000001804

RSTKGRNT (P-Value)	<0.1	<0.1	>0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		61406	
RMSE		1.449	

kscritvalue: 0.01033 , KS Statistic: 0.1357

11.4– Regression results of IV: RSTKGRNT, DV: Debt to Equity Ratio

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.462	0.462	0.795
Residual standard error	1.89	1.89	1.17
RSTKGRNT (slope)	-0.000017647	-0.000017647	-0.00000966
RSTKGRNT (P-Value)	>0.1	>0.1	>0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		71135	
RMSE		1.995	

kscritvalue: 0.01033, KS Statistic: 0.1542

12.1– Regression Results of IV: OPTION_AWARDS_BLK_VALUE (OABV), DV: Net Profit Margin

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0147	0.0147	0.329
Residual standard error	0.1409	0.1409	0.116
OABV (slope)	-0.000000499	-0.000000499	-0.0000005164
OABV (P-Value)	<0.1	<0.1	>0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		12277	
RMSE		0.139	

kscritvalue: 0.01033, KS Statistic: 0.1981

12.2– Regression Results of IV: OPTION_AWARDS_BLK_VALUE, DV: Return on Asset

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0142	0.142	0.39
Residual standard error	0.0893	0.0893	0.0702
OABV (slope)	-0.0000002347	-0.0000002347	-0.0000002893
OABV (P value)	>0.1	>0.1	<0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		21985	
RMSE		0.08732	

kscritvalue: 0.01033, KS Statistic: 0.1293

12.3– Regression Results of IV: OPTION_AWARDS_BLK_VALUE, DV: Current Ratio

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.0924	0.0924	0.717
Residual standard error	1.43	1.43	0.796
OABV (slope)	-0.0000004	-0.0000004	0.000003444
OABV (P value)	>0.1	>0.1	<0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		61428	
RMSE		1.449	

kscritvalue: 0.01033, KS Statistic: 0.1352

12.4– Regression Results of IV: OPTION_AWARDS_BLK_VALUE, DV: Debt to Equity Ratio

	Model 1: IV + CV	Model 2: Backward/ Forward/ Stepwise	Model 3: IV + CV + FE
Adjusted R square	0.4618	0.4618	0.795
Residual standard error	1.8860	1.8860	1.17

OABV (slope)	0.000001245	0.000001245	-0.00000256
OABV (P value)	>0.1	>0.1	<0.1
F statistic (P value)	<0.1	<0.1	<0.1
AIC		71145	
RMSE		1.996	

kscritvalue: 0.01033, KS Statistic: 0.1542

13.1– Regression results of IV: LTIP/ Total assets, DV: Net Profit Margin

	Model 1: IV + CV	Model 2: Stepwise/Backward/Forward	Model 3: IV + CV + FE
Adjusted R square	0.0145	0.0145	0.328
Residual standard error	0.141	0.141	0.116
LTIP_at (slope)	0.005357836	0.005357836	-0.0018263303
LTIP_at (P value)	> 0.1	> 0.1	> 0.1
F statistic (P value)	< 0.1	< 0.1	< 0.1
AIC		-67884	
RMSE		0.139	

kscritvalue: 0.01033, KS statistic: 0.1981

13.2 – Regression results of IV: LTIP/Total Assets, DV: Return on Asset

	Model 1: IV + CV	Model 2: Stepwise/Backward/Forward	Model 3: IV + CV + FE
Adjusted R square	0.0144	0.0144	0.39
Residual standard error	0.0893	0.0893	0.0702
LTIP_at (slope)	0.0085003758	0.0085003758	0.0003515685
LTIP_at (P value)	< 0.1	< 0.1	> 0.1
F statistic (P value)	< 0.1	< 0.1	< 0.1
AIC		-83701	
RMSE		0.08725	

kscritvalue: 0.01033, KS statistic: 0.1295

13.3 – Regression results of IV: LTIP/Total Assets, DV: Current Ratio

	Model 1: IV + CV	Model 2: Stepwise/Backward/Forward	Model 3: IV + CV + FE
Adjusted R square	0.0932	0.0932	0.717
Residual standard error	1.42	1.42	0.796
LTIP_at (slope)	-0.230164345	-0.230164345	-0.068259086
LTIP_at (P value)	< 0.1	< 0.1	> 0.1
F statistic (P value)	< 0.1	< 0.1	< 0.1
AIC		12262	
RMSE		1.448	

kscritvalue: 0.01033, KS statistic: 0.1357

13.4 – Regression results of IV: LTIP/Total Assets, DV: Debt to Equity Ratio

	Model 1: IV + CV	Model 2: Stepwise/Backward/Forward	Model 3: IV + CV + FE
Adjusted R square	0.462	0.462	0.795
Residual standard error	1.89	1.89	1.17
LTIP_at (slope)	0.033826292	0.033826292	0.00602919
LTIP_at (P value)	> 0.1	> 0.1	> 0.1
F statistic (P value)	< 0.1	< 0.1	< 0.1
AIC		21983	
RMSE		1.995	

kscritvalue: 0.01033, KS statistic: 0.1543

Appendix D: R Programming Code

```
#display numeric value
options (scipen=999, digits=4)

#clear environment
rm(list = ls())

#Load often used Libraries
library(readr)
library(dplyr)
library(psych)
library(ggplot2)
library(caret)
library(lfe)
library(broom)
library(stargazer)
library(reshape2)
library(Matrix)
library(vtable)
library(lubridate)
library(zoo)
library(leaps)
library(fitdistrplus)
library(skimr)
library(forecast)
library(tidyr)

#1.Importing Comp Financial Data
comp_data <- read_csv("comp_financial.csv")
dim(comp_data)
str(comp_data)
summary(comp_data)

#2.Basic data cleaning
comp_data <- arrange(comp_data, gvkey, fyear)

#2.1. analyse an replace missing fyear NA
mean(is.na(comp_data$fyear)) #0.0009467 although very few
NA it is important to remove NAs in fyear

#fill in fyear if missing, based on Compustat's May cutoff
comp_data$fyear <- ifelse(
  is.na(comp_data$fyear),
  ifelse(
    as.numeric(format(comp_data$datadate, format = "%m"))
> 5,
    as.numeric(format(comp_data$datadate, format = "%Y")),
    as.numeric(format(comp_data$datadate, format = "%Y")) -
1), comp_data$fyear)

summary(comp_data$fyear)#no fyear NAs

#verification that each row have a gvkey and fyear
nrow(comp_data) - nrow(subset(comp_data, !is.na(gvkey) &
!is.na(fyear))) #0

#2.2 creating firm-year indices using gvkey and fyear
comp_data <- arrange(comp_data, gvkey, fyear)
comp_data$index <- paste(comp_data$gvkey,
comp_data$fyear, sep = "_")

#2.3 filtering only "INDL" data
comp_data <- filter(comp_data, indfmt == "INDL")

#2.4 identifying and removing duplicate indices
length(unique(comp_data$index)) #291933 different from
the base data set

comp_data_clean <- subset(
  comp_data, !(index %in% subset(comp_data,
duplicated(index) == 1)$index))
nrow(comp_data_clean) #261907 rows
length(unique(comp_data_clean$index)) #261907 firm-year
indices

#3. Importing executive compensation data
exec_data <- read.csv("new_exec_comp.csv")
dim(exec_data)
str(exec_data)
summary(exec_data)

#4. Basic data cleaning
exec_data <- arrange(exec_data, GVKEY, YEAR, CO_PER_ROL)

#4.1 Filtering to only including the data to CEO
exec_data <- exec_data[exec_data$CEOANN == "CEO", ]

#4.2 converting GVKEY to character
exec_data$GVKEY <- as.character(exec_data$GVKEY)
exec_data$GVKEY <- ifelse(nchar(exec_data$GVKEY) == 4,
paste0("00", exec_data$GVKEY), exec_data$GVKEY)
exec_data$GVKEY <- ifelse(nchar(exec_data$GVKEY) == 5,
paste0("0", exec_data$GVKEY), exec_data$GVKEY)

summary(exec_data$GVKEY)

#4.3 analyse missing Year values
mean(is.na(exec_data$YEAR)) #no NAs
nrow(exec_data) - nrow(subset(exec_data, !is.na(GVKEY) &
!is.na(YEAR))) #0

#4.4 creating firm-year indices using gvkey and year
exec_data <- arrange(exec_data, GVKEY, YEAR, CO_PER_ROL)
exec_data$index <- paste(exec_data$GVKEY,
exec_data$YEAR, sep = "_")

#4.5 identifying and removing duplicate indices
```

```
length(unique(exec_data$index)) #44940 different from the base data set

exec_data_clean <- subset(
  exec_data,
  !(index %in% subset(exec_data, duplicated(index) == 1)$index)
)
nrow(exec_data_clean) #44897 rows
length(unique(exec_data_clean$index)) #44897 firm-year indices

#5.combining the two data set using inner join on index
data_comb <- inner_join(comp_data_clean, exec_data_clean, by = "index")

#6. data cleaning on combined data set
data_comb <- arrange(data_comb, index)

#6.1 replace missing values with zero for ni, revt, oiadp, act, lct, ch, lt, invt
data_comb_1 <- data_comb %>% mutate(ni = ifelse(is.na(ni), 0, ni), revt = ifelse(is.na(revt), 0, revt), oiadp = ifelse(is.na(oiadp), 0, oiadp), act = ifelse(is.na(act), 0, act), lct= ifelse(is.na(lct), 0, lct), ch = ifelse(is.na(ch), 0, ch), lt = ifelse(is.na(lt), 0, lt), invt = ifelse(is.na(invt), 0, invt))

#6.2 create required lagged values
data_comb_2 <- arrange(data_comb_1, index)
data_comb_2 <- data_comb_1 %>% group_by(gvkey) %>% mutate(at_lag = ifelse(fyear == lag(fyear) + 1, lag(at, n = 1), NA), invt_lag = ifelse(fyear == lag(fyear) + 1, lag(invt, n = 1), NA), seq_lag = ifelse(fyear == lag(fyear) +1, lag(seq, n = 1), NA)) %>% ungroup()

#6.3 creating financial ratios: profit margin, operating profit, ROE, current ratio, cash ratio, debt ratio, debt to equity ratio, asset turnover, inventory turnover
data_comb_3 <- data_comb_2 %>%
mutate(net_profit_margin=ni/revt,operating_profit=oiadp/revt,ROE=ni/((seq+seq_lag)/2),
current_ratio=act/lct,cash_ratio=ch/lct,debt_ratio=lt/at,debt_to_equity_ratio=lt/seq,
asset_turnover=revt/(at-lt),roa=ni/((at+at_lag)/2),
inventory_turnover=cogs/((invt+invt_lag)/2))

str(data_comb_3)

data_comb_4 <- data_comb_3 %>% group_by(CO_PER_ROL) %>% mutate(ceo_years = YEAR - year(as.Date(BECAMECEO)), count_ceo_years = n()) %>% ungroup() %>%
filter(!ceo_years<0)
summary(data_comb_4)

#7. Regression

#7.1 Select necessary variables for model estimation
data_reg_3 <- data_comb_4 %>% dplyr :: select(gvkey, fyear, index, sic, ceo_years, at, net_profit_margin, debt_to_equity_ratio, asset_turnover, inventory_turnover, roa, current_ratio,
                    SALARY, BONUS, OTHCOMP, RSTKGRNT, OPTION_AWARDS_BLK_VALUE, LTIP)
summary(data_reg_3)
sum(is.infinite(data_reg_3$inventory_turnover))

#7.2 Remove NA and Infinite values in ratios
data_reg_3a <- data_reg_3 %>%
  filter(
    !is.na(at) &
     !is.infinite(inventory_turnover))

summary(data_reg_3a)

#7.3 remove outliers with truncation method
ggplot(data_reg_3a, aes(x = fyear, y = net_profit_margin)) + geom_point() + geom_smooth(method = "lm", se = FALSE)
ggplot(data_reg_3a, aes(x = fyear, y = debt_to_equity_ratio)) + geom_point() + geom_smooth(method = "lm", se = FALSE)
ggplot(data_reg_3a, aes(x = fyear, y = inventory_turnover)) + geom_point() + geom_smooth(method = "lm", se = FALSE)
ggplot(data_reg_3a, aes(x = fyear, y = asset_turnover)) + geom_point() + geom_smooth(method = "lm", se = FALSE)
ggplot(data_reg_3a, aes(x = fyear, y = current_ratio)) + geom_point() + geom_smooth(method = "lm", se = FALSE)
ggplot(data_reg_3a, aes(x = fyear, y = roa)) + geom_point() + geom_smooth(method = "lm", se = FALSE)
ggplot(data_reg_3a, aes(x = ceo_years)) + geom_density()

summary(data_reg_3a)

data_reg_3d_clean <- data_reg_3a %>%
filter(!net_profit_margin > quantile(net_profit_margin, 0.99, na.rm = TRUE) & !net_profit_margin < quantile(net_profit_margin, 0.01, na.rm = TRUE) & !debt_to_equity_ratio > quantile(debt_to_equity_ratio, 0.99, na.rm = TRUE) & !debt_to_equity_ratio < quantile(debt_to_equity_ratio, 0.01, na.rm = TRUE) & !inventory_turnover > quantile(inventory_turnover, 0.99, na.rm = TRUE) & !inventory_turnover < quantile(inventory_turnover, 0.01, na.rm = TRUE) & !asset_turnover > quantile(asset_turnover, 0.99, na.rm = TRUE) & !asset_turnover < quantile(asset_turnover, 0.01, na.rm = TRUE) & !current_ratio > quantile(current_ratio, 0.99, na.rm = TRUE) & !current_ratio < quantile(current_ratio, 0.01, na.rm = TRUE) & !roa > quantile(roa, 0.99, na.rm = TRUE) & !roa < quantile(roa, 0.01, na.rm = TRUE) &!ceo_years > 20) %>% mutate(othcomp_at =
```

```
OTHCOMP/at)  %>% mutate(salary_at = SALARY/at) %>%
mutate(bonus_at = BONUS/at) %>% mutate(LTIP_at =
LTIP/at)

summary(data_reg_3d_clean)

#8 multi-linear regression to analyse the effects of salary on
profit margin, roa, liquidity ratio, debt to equity ratio
#Dependent variable: net_profit_margin, roa, current_ratio,
debt_to_equity_ratio
#Independent variable: salary_at (salary/total assets)

#8.1. choosing training and test dataset
set.seed(1)
train1 <- sample_frac(data_reg_3d_clean, 0.75)
test1 <- anti_join(data_reg_3d_clean, train1)

#8.2. regressing salary_at with profit margin; control
varaibles: at, debt to equity ratio, ceo_years
lm1a_train <- lm(net_profit_margin ~ salary_at + at +
debt_to_equity_ratio + ceo_years, train1)

#8.2.1. perform stepwise regression
data_reg_stepwise_1a <- step(lm1a_train, direction = "both")
summary(data_reg_stepwise_1a)
data_reg_stepwise_pred_1a <-
predict(data_reg_stepwise_1a, test1)
accuracy(data_reg_stepwise_pred_1a,
test1$net_profit_margin)

#8.2.2. perform forward regression
data_reg_forward_1a <- step(lm1a_train, direction =
"forward")
summary(data_reg_forward_1a)
data_reg_forward_pred_1a <- predict(data_reg_forward_1a,
test1)
accuracy(data_reg_forward_pred_1a,
test1$net_profit_margin)

#8.2.3. perform backward regression
data_reg_backward_1a <- step(lm1a_train, direction =
"backward")
summary(data_reg_backward_1a)
data_reg_backward_pred_1a <-
predict(data_reg_backward_1a, test1)
accuracy(data_reg_backward_pred_1a,
test1$net_profit_margin)

#8.2.4. perform regression with fixed effects
data_reg_fixeff_1a <- felm(net_profit_margin ~ salary_at + at
+ debt_to_equity_ratio + ceo_years | gvkey + fyear |0 |
gvkey + fyear, train1)
summary(data_reg_fixeff_1a)

#8.2.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_1a <- car::vif(lm1a_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#8.2.6 resdiual analysis
resid_1a <- lm1a_train$residuals
plot(train1$salary_at, resid_1a) #relationship between
salary_at and residuals

fnorm_1a <- fitdist(resid_1a, "norm")
result_1a <- gofstat(fnorm_1a, discrete = FALSE)
result_1a
kscritvalue_1a <- 1.36/sqrt(length(train1$net_profit_margin))
kscritvalue_1a #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_1a)
plot(fnorm_1a) #residuals are not normally distributed
confint(lm1a_train, level = 0.95)

#8.2.7. Using stargazer for well-formatted regression output
stargazer(lm1a_train, data_reg_forward_1a,
data_reg_backward_1a, data_reg_stepwise_1a,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table1a.txt")

#8.3. regressing salary_at with return of asset; control
variables: total assets, debt to equity ratio, and CEO years
lm1b_train <- lm(roa ~ salary_at + at + debt_to_equity_ratio
+ ceo_years, train1)

#8.3.1. perform stepwise regression
data_reg_stepwise_1b <- step(lm1b_train, direction = "both")
summary(data_reg_stepwise_1b)
data_reg_stepwise_pred_1b <-
predict(data_reg_stepwise_1b, test1)
accuracy(data_reg_stepwise_pred_1b, test1$roa)

#8.3.2. perform forward regression
data_reg_forward_1b <- step(lm1b_train, direction =
"forward")
summary(data_reg_forward_1b)
data_reg_forward_pred_1b <- predict(data_reg_forward_1b,
test1)
accuracy(data_reg_forward_pred_1b, test1$roa)

#8.3.3. perform backward regression
data_reg_backward_1b <- step(lm1b_train, direction =
"backward")
summary(data_reg_backward_1b)
data_reg_backward_pred_1b <-
predict(data_reg_backward_1b, test1)
```

```
accuracy(data_reg_backward_pred_1b, test1$roa)

#8.3.4. perform regression with fixed effects
data_reg_fixeff_1b <- felm(roa ~ salary_at + at + debt_to_equity_ratio + ceo_years | gvkey + fyear |0 | gvkey + fyear, train1)
summary(data_reg_fixeff_1b)

#8.3.5. Checking for multicollinearity of controlled variables with VIF
vif_values_1b <- car::vif(lm1b_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#8.3.6 resdiual analysis
resid_1b <- lm1b_train$residuals
plot(train1$salary_at, resid_1b) #relationship between salary_at and residuals

fnorm_1b <- fitdist(resid_1b, "norm")
result_1b <- gofstat(fnorm_1b, discrete = FALSE)
result_1b
kscritvalue_1b <- 1.36/sqrt(length(train1$roa))
kscritvalue_1b #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_1b)
plot(fnorm_1b) #residuals are not normally distributed
confint(lm1b_train, level = 0.95)

#8.3.7. Using stargazer for well-formatted regression output
stargazer(lm1b_train, data_reg_forward_1b, data_reg_backward_1b, data_reg_stepwise_1b, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table1b.txt")

#8.4. regressing salary_at with current ratio; control variables: total assets, debt to equity ratio, and asset turnover
lm1c_train <- lm(current_ratio ~ salary_at + at + debt_to_equity_ratio + asset_turnover, train1)

#8.4.1. perform stepwise regression
data_reg_stepwise_1c <- step(lm1c_train, direction = "both")
summary(data_reg_stepwise_1c)
data_reg_stepwise_pred_1c <- predict(data_reg_stepwise_1c, test1)
accuracy(data_reg_stepwise_pred_1c, test1$current_ratio)

#8.4.2. perform forward regression
data_reg_forward_1c <- step(lm1c_train, direction = "forward")
summary(data_reg_forward_1c)
data_reg_forward_pred_1c <- predict(data_reg_forward_1c, test1)
accuracy(data_reg_forward_pred_1c, test1$current_ratio)

#8.4.3. perform backward regression
data_reg_backward_1c <- step(lm1c_train, direction = "backward")
summary(data_reg_backward_1c)
data_reg_backward_pred_1c <- predict(data_reg_backward_1c, test1)
accuracy(data_reg_backward_pred_1c, test1$current_ratio)

#8.4.4. perform regression with fixed effects
data_reg_fixeff_1c <- felm(current_ratio ~ salary_at + at + debt_to_equity_ratio + asset_turnover | gvkey + fyear |0 | gvkey + fyear, train1)
summary(data_reg_fixeff_1c)

#8.4.5. Checking for multicollinearity of controlled variables with VIF
vif_values_1c <- car::vif(lm1c_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#8.4.6 resdiual analysis
resid_1c <- lm1c_train$residuals
plot(train1$salary_at, resid_1c) #relationship between salary_at and residuals

fnorm_1c <- fitdist(resid_1c, "norm")
result_1c <- gofstat(fnorm_1c, discrete = FALSE)
result_1c
kscritvalue_1c <- 1.36/sqrt(length(train1$current_ratio))
kscritvalue_1c #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_1c)
plot(fnorm_1c) #residuals are not normally distributed
confint(lm1c_train, level = 0.95)

#8.4.7. Using stargazer for well-formatted regression output
stargazer(lm1c_train, data_reg_forward_1c, data_reg_backward_1c, data_reg_stepwise_1c, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table1c.txt")

#8.5. regressing salary_at with debt to equity ratio; control variables: total assets, net profit margin, and asset turnover
lm1d_train <- lm(debt_to_equity_ratio ~ salary_at + at + net_profit_margin + asset_turnover, train1)

#8.5.1. perform stepwise regression
data_reg_stepwise_1d <- step(lm1d_train, direction = "both")
summary(data_reg_stepwise_1d)
```

```
data_reg_stepwise_pred_1d <-
predict(data_reg_stepwise_1d, test1)
accuracy(data_reg_stepwise_pred_1d,
test1$debt_to_equity_ratio)

#8.5.2. perform forward regression
data_reg_forward_1d <- step(lm1d_train, direction =
"forward")
summary(data_reg_forward_1d)
data_reg_forward_pred_1d <- predict(data_reg_forward_1d,
test1)
accuracy(data_reg_forward_pred_1d,
test1$debt_to_equity_ratio)

#8.5.3. perform backward regression
data_reg_backward_1d <- step(lm1d_train, direction =
"backward")
summary(data_reg_backward_1d)
data_reg_backward_pred_1d <-
predict(data_reg_backward_1d, test1)
accuracy(data_reg_backward_pred_1d,
test1$debt_to_equity_ratio)

#8.5.4. perform regression with fixed effects
data_reg_fixeff_1d <- felm(debt_to_equity_ratio ~ salary_at
+ at + net_profit_margin + asset_turnover | gvkey + fyear |0
| gvkey + fyear, train1)
summary(data_reg_fixeff_1d)

#8.5.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_1d <- car::vif(lm1d_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#8.5.6 resdiual analysis
resid_1d <- lm1d_train$residuals
plot(train1$salary_at, resid_1d) #relationship between
salary_at and residuals

fnorm_1d <- fitdist(resid_1d, "norm")
result_1d <- gofstat(fnorm_1d, discrete = FALSE)
result_1d
kscritvalue_1d <-
1.36/sqrt(length(train1$debt_to_equity_ratio))
kscritvalue_1d #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_1d)
plot(fnorm_1d) #residuals are not normally distributed
confint(lm1d_train, level = 0.95)

#8.5.7. Using stargazer for well-formatted regression output
stargazer(lm1d_train, data_reg_forward_1d,
data_reg_backward_1d, data_reg_stepwise_1d,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table1d.txt")

#9 multi-linear regression to analyse the effects of bonus on
profit margin, roa, liquidity ratio, debt to equity ratio
#Dependent variable: net_profit_margin, roa, current_ratio,
debt_to_equity_ratio
#Independent variable: bonus_at (bonus/total assets)

#9.1. choosing training and test dataset
set.seed(1)
train2 <- sample_frac(data_reg_3d_clean, 0.75)
test2 <- anti_join(data_reg_3d_clean, train2)

#9.2. regressing salary_at with profit margin; control
varaibles: at, debt to equity ratio, ceo_years
lm2a_train <- lm(net_profit_margin ~ bonus_at + at +
debt_to_equity_ratio + ceo_years, train2)

#9.2.1. perform stepwise regression
data_reg_stepwise_2a <- step(lm2a_train, direction = "both")
summary(data_reg_stepwise_2a)
data_reg_stepwise_pred_2a <-
predict(data_reg_stepwise_2a, test2)
accuracy(data_reg_stepwise_pred_2a,
test2$net_profit_margin)

#9.2.2. perform forward regression
data_reg_forward_2a <- step(lm2a_train, direction =
"forward")
summary(data_reg_forward_2a)
data_reg_forward_pred_2a <- predict(data_reg_forward_2a,
test2)
accuracy(data_reg_forward_pred_2a,
test2$net_profit_margin)

#9.2.3. perform backward regression
data_reg_backward_2a <- step(lm2a_train, direction =
"backward")
summary(data_reg_backward_2a)
data_reg_backward_pred_2a <-
predict(data_reg_backward_2a, test2)
accuracy(data_reg_backward_pred_2a,
test2$net_profit_margin)

#9.2.4. perform regression with fixed effects
data_reg_fixeff_2a <- felm(net_profit_margin ~ bonus_at + at
+ debt_to_equity_ratio + ceo_years | gvkey + fyear |0 |
gvkey + fyear, train2)
summary(data_reg_fixeff_2a)
```

```
#9.2.5. Checking for multicollinearity of controlled variables with VIF
vif_values_2a <- car::vif(lm2a_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#9.2.6 resdiual analysis
resid_2a <- lm2a_train$residuals
plot(train2$bonus_at, resid_2a) #relationship between salary_at and residuals

fnorm_2a <- fitdist(resid_2a, "norm")
result_2a <- gofstat(fnorm_2a, discrete = FALSE)
result_2a
kscritvalue_2a <- 1.36/sqrt(length(train2$net_profit_margin))
kscritvalue_2a #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_2a)
plot(fnorm_2a) #residuals are not normally distributed
confint(lm2a_train, level = 0.95)

#9.2.7. Using stargazer for well-formatted regression output
stargazer(lm2a_train, data_reg_forward_2a, data_reg_backward_2a, data_reg_stepwise_2a, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table2a.txt")

#9.3. regressing bonus_at with return of asset; control variables: total assets, debt to equity ratio, and CEO years
lm2b_train <- lm(roa ~ bonus_at + at + debt_to_equity_ratio + ceo_years, train2)

#9.3.1. perform stepwise regression
data_reg_stepwise_2b <- step(lm2b_train, direction = "both")
summary(data_reg_stepwise_2b)
data_reg_stepwise_pred_2b <- predict(data_reg_stepwise_2b, test2)
accuracy(data_reg_stepwise_pred_2b, test2$roa)

#9.3.2. perform forward regression
data_reg_forward_2b <- step(lm2b_train, direction = "forward")
summary(data_reg_forward_2b)
data_reg_forward_pred_2b <- predict(data_reg_forward_2b, test2)
accuracy(data_reg_forward_pred_2b, test2$roa)

#9.3.3. perform backward regression
data_reg_backward_2b <- step(lm2b_train, direction = "backward")
summary(data_reg_backward_2b)
data_reg_backward_pred_2b <- predict(data_reg_backward_2b, test2)
accuracy(data_reg_backward_pred_2b, test2$roa)

#9.3.4. perform regression with fixed effects
data_reg_fixeff_2b <- felm(roa ~ bonus_at + inventory_turnover + asset_turnover + ceo_years | gvkey + fyear |0 | gvkey + fyear, train2)
summary(data_reg_fixeff_2b)

#9.3.5. Checking for multicollinearity of controlled variables with VIF
vif_values_2b <- car::vif(lm2b_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#9.3.6 resdiual analysis
resid_2b <- lm2b_train$residuals
plot(train2$bonus_at, resid_2b) #relationship between salary_at and residuals

fnorm_2b <- fitdist(resid_2b, "norm")
result_2b <- gofstat(fnorm_2b, discrete = FALSE)
result_2b
kscritvalue_2b <- 1.36/sqrt(length(train2$roa))
kscritvalue_2b #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_2b)
plot(fnorm_2b) #residuals are not normally distributed
confint(lm2b_train, level = 0.95)

#9.3.7. Using stargazer for well-formatted regression output
stargazer(lm2b_train, data_reg_forward_2b, data_reg_backward_2b, data_reg_stepwise_2b, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table2b.txt")

#9.4. regressing bonus_at with current ratio; control variables: total assets, debt to equity ratio, and asset turnover
lm2c_train <- lm(current_ratio ~ bonus_at + at + debt_to_equity_ratio + asset_turnover, train2)

#9.4.1. perform stepwise regression
data_reg_stepwise_2c <- step(lm2c_train, direction = "both")
summary(data_reg_stepwise_2c)
data_reg_stepwise_pred_2c <- predict(data_reg_stepwise_2c, test2)
accuracy(data_reg_stepwise_pred_2c, test2$current_ratio)

#9.4.2. perform forward regression
data_reg_forward_2c <- step(lm2c_train, direction = "forward")
summary(data_reg_forward_2c)
```

```
data_reg_forward_pred_2c <- predict(data_reg_forward_2c,
test2)
accuracy(data_reg_forward_pred_2c, test2$current_ratio)

#9.4.3. perform backward regression
data_reg_backward_2c <- step(lm2c_train, direction =
"backward")
summary(data_reg_backward_2c)
data_reg_backward_pred_2c <-
predict(data_reg_backward_2c, test2)
accuracy(data_reg_backward_pred_2c, test2$current_ratio)

#9.4.4. perform regression with fixed effects
data_reg_fixeff_2c <- felm(current_ratio ~ bonus_at + at +
debt_to_equity_ratio + asset_turnover | gvkey + fyear |0 |
gvkey + fyear, train2)
summary(data_reg_fixeff_2c)

#9.4.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_2c <- car::vif(lm2c_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#9.4.6 resdiual analysis
resid_2c <- lm2c_train$residuals
plot(train2$bonus_at, resid_2c) #relationship between
salary_at and residuals

fnorm_2c <- fitdist(resid_2c, "norm")
result_2c <- gofstat(fnorm_2c, discrete = FALSE)
result_2c
kscritvalue_2c <- 1.36/sqrt(length(train2$current_ratio))
kscritvalue_2c #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_2c)
plot(fnorm_2c) #residuals are not normally distributed
confint(lm2c_train, level = 0.95)

#9.4.7. Using stargazer for well-formatted regression output
stargazer(lm2c_train, data_reg_forward_2c,
data_reg_backward_2c, data_reg_stepwise_2c,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table2c.txt")

#9.5. regressing salary_at with debt to equity ratio; control
variables: total assets, net profit margin, and asset turnover
lm2d_train <- lm(debt_to_equity_ratio ~ bonus_at + at +
net_profit_margin + asset_turnover, train2)

#9.5.1. perform stepwise regression
data_reg_stepwise_2d <- step(lm2d_train, direction = "both")
summary(data_reg_stepwise_2d)
data_reg_stepwise_pred_2d <-
predict(data_reg_stepwise_2d, test2)
accuracy(data_reg_stepwise_pred_2d,
test2$debt_to_equity_ratio)

#9.5.2. perform forward regression
data_reg_forward_2d <- step(lm2d_train, direction =
"forward")
summary(data_reg_forward_2d)
data_reg_forward_pred_2d <- predict(data_reg_forward_2d,
test2)
accuracy(data_reg_forward_pred_2d,
test2$debt_to_equity_ratio)

#9.5.3. perform backward regression
data_reg_backward_2d <- step(lm2d_train, direction =
"backward")
summary(data_reg_backward_2d)
data_reg_backward_pred_2d <-
predict(data_reg_backward_2d, test2)
accuracy(data_reg_backward_pred_2d,
test2$debt_to_equity_ratio)

#9.5.4. perform regression with fixed effects
data_reg_fixeff_2d <- felm(debt_to_equity_ratio ~ bonus_at
+ at + net_profit_margin + asset_turnover | gvkey + fyear |0
| gvkey + fyear, train2)
summary(data_reg_fixeff_2d)

#9.5.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_2d <- car::vif(lm2d_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#9.5.6 residual analysis
resid_2d <- lm2d_train$residuals
plot(train2$bonus_at, resid_2d) #relationship between
salary_at and residuals

fnorm_2d <- fitdist(resid_2d, "norm")
result_2d <- gofstat(fnorm_2d, discrete = FALSE)
result_2d
kscritvalue_2d <-
1.36/sqrt(length(train2$debt_to_equity_ratio))
kscritvalue_2d #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_2d)
plot(fnorm_2d) #residuals are not normally distributed
confint(lm2d_train, level = 0.95)

#9.5.7. Using stargazer for well-formatted regression output
```

```
stargazer(lm2d_train, data_reg_forward_2d,
data_reg_backward_2d, data_reg_stepwise_2d,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table2d.txt")

#10 multi-linear regression to analyse the effects of other
compensation on profit margin, roa, liquidity ratio, debt to
equity ratio
#Dependent variable: net_profit_margin, roa, current_ratio,
debt_to_equity_ratio
#Independent variable: othcomp_at (other
compensation/total assets)

#10.1. choosing training and test dataset
set.seed(1)
train3 <- sample_frac(data_reg_3d_clean, 0.75)
test3 <- anti_join(data_reg_3d_clean, train3)

#10.2. regressing othcomp_at with profit margin; control
varaibles: at, debt to equity ratio, ceo_years
lm3a_train <- lm(net_profit_margin ~ othcomp_at + at +
debt_to_equity_ratio + ceo_years, train3)

# 10.2.1. perform stepwise regression
data_reg_stepwise_3a <- step(lm3a_train, direction = "both")
summary(data_reg_stepwise_3a)
data_reg_stepwise_pred_3a <-
predict(data_reg_stepwise_3a, test3)
accuracy(data_reg_stepwise_pred_3a,
test3$net_profit_margin)

# 10.2.2. perform forward regression
data_reg_forward_3a <- step(lm3a_train, direction =
"forward")
summary(data_reg_forward_3a)
data_reg_forward_pred_3a <- predict(data_reg_forward_3a,
test3)
accuracy(data_reg_forward_pred_3a,
test3$net_profit_margin)

# 10.2.3. perform backward regression
data_reg_backward_3a <- step(lm3a_train, direction =
"backward")
summary(data_reg_backward_3a)
data_reg_backward_pred_3a <-
predict(data_reg_backward_3a, test3)
accuracy(data_reg_backward_pred_3a,
test3$net_profit_margin)

#10.2.4. perform regression with fixed effects
data_reg_fixeff_3a <- felm(net_profit_margin ~ othcomp_at
+ at + debt_to_equity_ratio + ceo_years | gvkey + fyear |0 |
gvkey + fyear, train3)
summary(data_reg_fixeff_3a)

#10.2.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_3a <- car::vif(lm3a_train) #no multicollinearity
issues as VIF values remain low across variables

#10.2.6 resdiual analysis
lm3a <- lm(net_profit_margin ~ othcomp_at, train3) #simple
linear regression with only other compensation
summary(lm3a)
plot(train3$othcomp_at, train3$net_profit_margin,
main="Relationship between
    other compensation and profit margin",
    xlab="othcomp_at", ylab="Profit Margin")
abline(lm3a, lwd=3, col="red")

resid_3a <- lm3a$residuals
plot(train3$othcomp_at, resid_3a) #relationship between
othcomp_at and residuals

fnorm_3a <- fitdist(resid_3a, "norm")
result_3a <- gofstat(fnorm_3a, discrete = FALSE)
result_3a
kscritvalue_3a <- 1.36/sqrt(length(train3$net_profit_margin))
kscritvalue_3a #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_3a)
plot(fnorm_3a) #residuals are not normally distributed
confint(lm3a_train, level = 0.95)

#10.2.7. Using stargazer for well-formatted regression output
stargazer(lm3a_train, data_reg_forward_3a,
data_reg_backward_3a, data_reg_stepwise_3a,
type="text",title="Regression Results", digits=4,  no.space =
TRUE, out="table3a.txt")

#10.3. regressing othcomp_at with roa; control varaibles: at,
debt to equity ratio, ceo_years
lm3b_train <- lm(roa ~ othcomp_at + at +
debt_to_equity_ratio + ceo_years, train3)

# 10.3.1. perform stepwise regression
data_reg_stepwise_3b <- step(lm3b_train, direction = "both")
summary(data_reg_stepwise_3b)
data_reg_stepwise_pred_3b <-
predict(data_reg_stepwise_3b, test3)
accuracy(data_reg_stepwise_pred_3b, test3$roa)

# 10.3.2. perform forward regression
data_reg_forward_3b <- step(lm3b_train, direction =
"forward")
summary(data_reg_forward_3b)
```

```
data_reg_forward_pred_3b <- predict(data_reg_forward_3b,
test3)
accuracy(data_reg_forward_pred_3b, test3$roa)

# 10.3.3. perform backward regression
data_reg_backward_3b <- step(lm3b_train, direction =
"backward")
summary(data_reg_backward_3b)
data_reg_backward_pred_3b <-
predict(data_reg_backward_3b, test3)
accuracy(data_reg_backward_pred_3b, test3$roa)

#10.3.4. perform regression with fixed effects
data_reg_fixeff_3b <- felm(roa ~ othcomp_at + at +
debt_to_equity_ratio + ceo_years | gvkey + fyear |0 | gvkey
+ fyear, train3)
summary(data_reg_fixeff_3b)

#10.3.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_3b <- car::vif(lm3b_train) #no multicollinearity
issues as VIF values remain low across variables

#10.3.6 resdiual analysis
lm3b <- lm(roa ~ othcomp_at, train3) #simple linear
regression with only other compensation
summary(lm3b)
plot(train3$othcomp_at, train3$roa, main="Relationship
between
    other compensation and ROA",
    xlab="othcomp_at", ylab="roa")
abline(lm3b, lwd=3, col="red")

resid_3b <- lm3b$residuals
plot(train3$othcomp_at, resid_3b) #relationship between
othcomp_at and residuals

fnorm_3b <- fitdist(resid_3b, "norm")
result_3b <- gofstat(fnorm_3b, discrete = FALSE)
result_3b
kscritvalue_3b <- 1.36/sqrt(length(train3$roa))
kscritvalue_3b #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_3b)
plot(fnorm_3b) #residuals are not normally distributed
confint(lm3b, level = 0.95)

#10.3.7. Using stargazer for well-formatted regression output
stargazer(lm3b_train, data_reg_forward_3b,
data_reg_backward_3b, data_reg_stepwise_3b,
type="text",title="Regression Results", digits=4,  no.space =
TRUE, out="table3b.txt")

#10.4. regressing othcomp_at with current ratio; control
varaibles: at, debt to equity ratio, asset turnover
lm3c_train <- lm(current_ratio ~ othcomp_at + at +
debt_to_equity_ratio + asset_turnover, train3)

# 10.4.1. perform stepwise regression
data_reg_stepwise_3c <- step(lm3c_train, direction = "both")
summary(data_reg_stepwise_3c)
data_reg_stepwise_pred_3c <-
predict(data_reg_stepwise_3c, test3)
accuracy(data_reg_stepwise_pred_3c, test3$current_ratio)

# 10.4.2. perform forward regression
data_reg_forward_3c <- step(lm3c_train, direction =
"forward")
summary(data_reg_forward_3c)
data_reg_forward_pred_3c <- predict(data_reg_forward_3c,
test3)
accuracy(data_reg_forward_pred_3c, test3$current_ratio)

# 10.4.3. perform backward regression
data_reg_backward_3c <- step(lm3c_train, direction =
"backward")
summary(data_reg_backward_3c)
data_reg_backward_pred_3c <-
predict(data_reg_backward_3c, test3)
accuracy(data_reg_backward_pred_3c, test3$current_ratio)

#10.4.4. perform regression with fixed effects
data_reg_fixeff_3c <- felm(current_ratio ~ othcomp_at + at +
debt_to_equity_ratio + asset_turnover | gvkey + fyear |0 |
gvkey + fyear, train3)
summary(data_reg_fixeff_3c)

#10.4.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_3c <- car::vif(lm3c_train) #no multicollinearity
issues as VIF values remain low across variables

#10.4.6 resdiual analysis
lm3c <- lm(current_ratio ~ othcomp_at, train3) #simple linear
regression with only other compensation
summary(lm3c)
plot(train3$othcomp_at, train3$current_ratio,
main="Relationship between
    other compensation and current ratio",
    xlab="othcomp_at", ylab="current ratio")
abline(lm3c, lwd=3, col="red")

resid_3c <- lm3c$residuals
plot(train3$othcomp_at, resid_3c) #relationship between
othcomp_at and residuals

fnorm_3c <- fitdist(resid_3c, "norm")
```

```
result_3c <- gofstat(fnorm_3c, discrete = FALSE)
result_3c
kscritvalue_3c <- 1.36/sqrt(length(train3$current_ratio))
kscritvalue_3c #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_3c)
plot(fnorm_3c) #residuals are not normally distributed
confint(lm3c, level = 0.95)

#10.4.7. Using stargazer for well-formatted regression output
stargazer(lm3c_train, data_reg_forward_3c, data_reg_backward_3c, data_reg_stepwise_3c, type="text",title="Regression Results", digits=4,  no.space = TRUE, out="table3c.txt")


#10.5. regressing othcomp_at with debt to equity ratio; control varaibles: at, profit margin, asset turnover
lm3d_train <- lm(debt_to_equity_ratio ~ othcomp_at + at + net_profit_margin + asset_turnover, train3)
summary(lm3d_train)

# 10.5.1. perform stepwise regression
data_reg_stepwise_3d <- step(lm3d_train, direction = "both")
summary(data_reg_stepwise_3d)
data_reg_stepwise_pred_3d <- predict(data_reg_stepwise_3d, test3)
accuracy(data_reg_stepwise_pred_3d, test3$debt_to_equity_ratio)

# 10.5.2. perform forward regression
data_reg_forward_3d <- step(lm3d_train, direction = "forward")
summary(data_reg_forward_3d)
data_reg_forward_pred_3d <- predict(data_reg_forward_3d, test3)
accuracy(data_reg_forward_pred_3d, test3$debt_to_equity_ratio)

# 10.5.3. perform backward regression
data_reg_backward_3d <- step(lm3d_train, direction = "backward")
summary(data_reg_backward_3d)
data_reg_backward_pred_3d <- predict(data_reg_backward_3d, test3)
accuracy(data_reg_backward_pred_3d, test3$debt_to_equity_ratio)

#10.5.4. perform regression with fixed effects
data_reg_fixeff_3d <- felm(debt_to_equity_ratio ~ othcomp_at + at + net_profit_margin + asset_turnover | gvkey + fyear |0 | gvkey + fyear, train3)
summary(data_reg_fixeff_3d)

#10.5.5. Checking for multicollinearity of controlled variables with VIF
vif_values_3d <- car::vif(lm3d_train) #no multicollinearity issues as VIF values remain low across variables

#10.5.6 resdiual analysis
lm3d <- lm(debt_to_equity_ratio ~ othcomp_at, train3) #simple linear regression with only other compensation
summary(lm3d)
plot(train3$othcomp_at, train3$debt_to_equity_ratio, main="Relationship between
    other compensation and debt to equity ratio",
    xlab="othcomp_at", ylab="Debt to equity ratio")
abline(lm3d, lwd=3, col="red")

resid_3d <- lm3d$residuals
plot(train3$othcomp_at, resid_3d) #relationship between othcomp_at and residuals

fnorm_3d <- fitdist(resid_3d, "norm")
result_3d <- gofstat(fnorm_3d, discrete = FALSE)
result_3d
kscritvalue_3d <- 1.36/sqrt(length(train3$debt_to_equity_ratio))
kscritvalue_3d #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_3d)
plot(fnorm_3d) #residuals are not normally distributed
confint(lm3d, level = 0.95)

#10.5.7. Using stargazer for well-formatted regression output
stargazer(lm3d_train, data_reg_forward_3d, data_reg_backward_3d, data_reg_stepwise_3d, type="text",title="Regression Results", digits=4,  no.space = TRUE, out="table3d.txt")

#11 multi-linear regression to analyse the effects of restricted stock grant (RSTKGRNT) on profit margin, roa, liquidity ratio, debt to equity ratio

#Dependent variable: net_profit_margin, roa, current_ratio, debt_to_equity_ratio
#Independent variable: RSTKGRNT (restricted stock grant)

#11.1. Replacing NAs with 0s
data_reg_4d_clean <- data_reg_3d_clean %>% mutate(RSTKGRNT = ifelse(is.na(RSTKGRNT),0,RSTKGRNT))

#11.1.2. choosing training and test dataset
set.seed(1)
train4 <- sample_frac(data_reg_4d_clean, 0.75)
test4 <- anti_join(data_reg_4d_clean, train4)
```

```
#11.2. regressing RSTKGRNT with profit margin; control varaibles: at, debt to equity ratio, ceo_years
lm4a_train <- lm(net_profit_margin ~ RSTKGRNT + at + debt_to_equity_ratio + ceo_years, train4)

# 11.2.1. perform stepwise regression
data_reg_stepwise_4a <- step(lm4a_train, direction = "both")
summary(data_reg_stepwise_4a)
data_reg_stepwise_pred_4a <- predict(data_reg_stepwise_4a, test4)
accuracy(data_reg_stepwise_pred_4a, test4$net_profit_margin)

# 11.2.2. perform forward regression
data_reg_forward_4a <- step(lm4a_train, direction = "forward")
summary(data_reg_forward_4a)
data_reg_forward_pred_4a <- predict(data_reg_forward_4a, test4)
accuracy(data_reg_forward_pred_4a, test4$net_profit_margin)

# 11.2.3. perform backward regression
data_reg_backward_4a <- step(lm4a_train, direction = "backward")
summary(data_reg_backward_4a)
data_reg_backward_pred_4a <- predict(data_reg_backward_4a, test4)
accuracy(data_reg_backward_pred_4a, test4$net_profit_margin)

#11.2.4. perform regression with fixed effects
data_reg_fixeff_4a <- felm(net_profit_margin ~ RSTKGRNT + debt_to_equity_ratio + ceo_years | gvkey + fyear |0 | gvkey + fyear, train4)
summary(data_reg_fixeff_4a)

#11.2.5. Checking for multicollinearity of controlled variables with VIF
vif_values_4a <- car::vif(lm4a_train) #no multicollinearity issues as VIF values remain low across variables

#11.2.6 resdiual analysis
resid_4a<- lm4a_train$residuals
plot(train4$RSTKGRNT, resid_4a) #relationship between RSTKGRNT and residuals

fnorm_4a <- fitdist(resid_4a, "norm")
result_4a <- gofstat(fnorm_4a, discrete = FALSE)
result_4a
kscritvalue_4a <- 1.36/sqrt(length(train4$net_profit_margin))
kscritvalue_4a #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_4a)
plot(fnorm_4a) #residuals are not normally distributed
confint(lm4a_train, level = 0.95)

#11.2.7. Using stargazer for well-formatted regression output
stargazer(lm4a_train, data_reg_forward_4a, data_reg_backward_4a, data_reg_stepwise_4a, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table4a.txt")

#11.3. regressing RSTKGRNT with ROA ; control variables: ceo_years, inventory turnover, asset turnover; gvkey + fyear

lm4b_train <- lm(roa ~ RSTKGRNT + inventory_turnover + asset_turnover + ceo_years, train4)

# 11.3.1. perform stepwise regression
data_reg_stepwise_4b <- step(lm4b_train, direction = "both")
summary(data_reg_stepwise_4b)
data_reg_stepwise_pred_4b <- predict(data_reg_stepwise_4b, test4)
accuracy(data_reg_stepwise_pred_4b, test4$roa)

# 11.3.2. perform forward regression
data_reg_forward_4b <- step(lm4b_train, direction = "forward")
summary(data_reg_forward_4b)
data_reg_forward_pred_4b <- predict(data_reg_forward_4b, test4)
accuracy(data_reg_forward_pred_4b, test4$roa)

# 11.3.3. perform backward regression
data_reg_backward_4b <- step(lm4b_train, direction = "backward")
summary(data_reg_backward_4b)
data_reg_backward_pred_4b <- predict(data_reg_backward_4b, test4)
accuracy(data_reg_backward_pred_4b, test4$roa)

#11.3.4. perform regression with fixed effects
data_reg_fixeff_4b <- felm(roa ~ RSTKGRNT + inventory_turnover + asset_turnover + ceo_years | gvkey + fyear |0 | gvkey + fyear, train4)
summary(data_reg_fixeff_4b)

#11.3.5. Checking for multicollinearity of controlled variables with VIF
vif_values_4b <- car::vif(lm3a_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables
```

```
#11.3.6 resdiual analysis
resid_4b <- lm4b_train$residuals
plot(train4$RSTKGRNT, resid_4b) #relationship between RSTKGRNT and residuals

fnorm_4b <- fitdist(resid_4b, "norm")
result_4b <- gofstat(fnorm_4b, discrete = FALSE)
result_4b
kscritvalue_4b<- 1.36/sqrt(length(train4$roa))
kscritvalue_4b #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_4b)
plot(fnorm_4b) #residuals are not normally distributed
confint(lm4b_train, level = 0.95)

#11.3.7. Using stargazer for well-formatted regression output
stargazer(lm4b_train, data_reg_forward_4b, data_reg_backward_4b, data_reg_stepwise_4b, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table4b.txt")

 #11.4. regressing RSTKGRNT with Current ratio (liquidity) ; control variables:  asset turnover, at, debt to equity ratio; gvkey + fyear

  lm4c_train <- lm(current_ratio ~ RSTKGRNT + at + debt_to_equity_ratio + asset_turnover, train4)

# 11.4.1. perform stepwise regression
data_reg_stepwise_4c <- step(lm4c_train, direction = "both")
summary(data_reg_stepwise_4c)
data_reg_stepwise_pred_4c <- predict(data_reg_stepwise_4c, test4)
accuracy(data_reg_stepwise_pred_4c, test4$current_ratio)

# 11.4.2. perform forward regression
data_reg_forward_4c <- step(lm4c_train, direction = "forward")
summary(data_reg_forward_4c)
data_reg_forward_pred_4c <- predict(data_reg_forward_4c, test4)
accuracy(data_reg_forward_pred_4c, test4$current_ratio)

# 11.4.3. perform backward regression
data_reg_backward_4c <- step(lm4c_train, direction = "backward")
summary(data_reg_backward_4c)
data_reg_backward_pred_4c <- predict(data_reg_backward_4c, test4)
accuracy(data_reg_backward_pred_4c, test4$current_ratio)

#11.4.4. perform regression with fixed effects
data_reg_fixeff_4c <- felm(current_ratio ~ RSTKGRNT + at + debt_to_equity_ratio + asset_turnover | gvkey + fyear |0 | gvkey + fyear, train4)
summary(data_reg_fixeff_4c)

#11.4.5. Checking for multicollinearity of controlled variables with VIF
vif_values_4c <- car::vif(lm4b_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#11.4.6 resdiual analysis
resid_4c <- lm4c_train$residuals
plot(train4$RSTKGRNT, resid_4c) #relationship between RSTKGRNT and residuals

fnorm_4c<- fitdist(resid_4c, "norm")
result_4c <- gofstat(fnorm_4c, discrete = FALSE)
result_4c
kscritvalue_4c <- 1.36/sqrt(length(train4$current_ratio))
kscritvalue_4c #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_4c)
plot(fnorm_4c) #residuals are not normally distributed
confint(lm4c_train, level = 0.95)

#11.4.7. Using stargazer for well-formatted regression output
stargazer(lm4c_train, data_reg_forward_4c, data_reg_backward_4c, data_reg_stepwise_4c, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table4c.txt")


#11.5. regressing RSTKGRNT with Debt to equity ratio (leverage) ; control variables:  at, net profit margin, asset turnover; gvkey + fyear

lm4d_train <- lm(debt_to_equity_ratio ~ RSTKGRNT + at + net_profit_margin + asset_turnover, train4)

# 11.5.1. perform stepwise regression
data_reg_stepwise_4d <- step(lm4d_train, direction = "both")
summary(data_reg_stepwise_4d)
data_reg_stepwise_pred_4d <- predict(data_reg_stepwise_4d, test4)
accuracy(data_reg_stepwise_pred_4d, test4$debt_to_equity_ratio)

# 11.5.2. perform forward regression
data_reg_forward_4d <- step(lm4d_train, direction = "forward")
summary(data_reg_forward_4d)
```

```
data_reg_forward_pred_4d <- predict(data_reg_forward_4d, test4)
accuracy(data_reg_forward_pred_4d, test4$debt_to_equity_ratio)

# 11.5.3. perform backward regression
data_reg_backward_4d <- step(lm4d_train, direction = "backward")
summary(data_reg_backward_4d)
data_reg_backward_pred_4d <- predict(data_reg_backward_4d, test4)
accuracy(data_reg_backward_pred_4d, test4$debt_to_equity_ratio)

#11.5.4. perform regression with fixed effects
data_reg_fixeff_4d <- felm(debt_to_equity_ratio ~ RSTKGRNT + at + net_profit_margin + asset_turnover | gvkey + fyear |0 | gvkey + fyear, train4)
summary(data_reg_fixeff_4d)

#11.5.5. Checking for multicollinearity of controlled variables with VIF
vif_values_4d <- car::vif(lm4d_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#11.5.6 resdiual analysis
resid_4d <- lm4d_train$residuals
plot(train4$RSTKGRNT, resid_4d) #relationship between RSTKGRNT and residuals

fnorm_4d <- fitdist(resid_4d, "norm")
result_4d <- gofstat(fnorm_4d, discrete = FALSE)
result_4d
kscritvalue_4d <- 1.36/sqrt(length(train4$debt_to_equity_ratio))
kscritvalue_4d #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_4d)
plot(fnorm_4d) #residuals are not normally distributed
confint(lm4d_train, level = 0.95)

#11.5.7. Using stargazer for well-formatted regression output
stargazer(lm4d_train, data_reg_forward_4d, data_reg_backward_4d, data_reg_stepwise_4d, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table4d.txt")

#12 multi-linear regression to analyse the effects of other compensation on profit margin, roa, liquidity ratio, debt to equity ratio
#Dependent variable: net_profit_margin, roa, current_ratio, debt_to_equity_ratio
#Independent variable: OPTION_AWARDS_BLK_VALUE

#12.1.1 Replacing NAs with 0s
data_reg_5d_cleaner <- data_reg_3d_clean %>% mutate(OPTION_AWARDS_BLK_VALUE = ifelse(is.na(OPTION_AWARDS_BLK_VALUE),0,OPTION_AWARDS_BLK_VALUE))

#12.1.2 Choosing Training and Test Dataset
set.seed(1)
train5 <- sample_frac(data_reg_5d_cleaner, 0.75)
test5 <- anti_join(data_reg_5d_cleaner, train5)

#12.2 Net Profit Margin Regression Analysis

#12.2.1 regressing OPTION_AWARDS_BLK_VALUE with net profit margin; control variables: at, debt to equity ratio, ceo_years
lm5a_train <- lm(net_profit_margin ~ OPTION_AWARDS_BLK_VALUE + at + debt_to_equity_ratio + ceo_years, train5)
summary(lm5a_train)

#12.2.2 Stepwise Regression
data_reg_stepwise_5a <- step(lm5a_train, direction = "both")
summary(data_reg_stepwise_5a)
data_reg_stepwise_pred_5a <- predict(data_reg_stepwise_5a, test5)
accuracy(data_reg_stepwise_pred_5a, test5$net_profit_margin)
summary(data_reg_stepwise_pred_5a)

#12.2.3 Forward Regression
data_reg_forward_5a <- step(lm5a_train, direction = "forward")
summary(data_reg_forward_5a)
data_reg_forward_pred_5a <- predict(data_reg_forward_5a, test5)
accuracy(data_reg_forward_pred_5a, test5$net_profit_margin)

#12.2.4 Backward Regression
data_reg_backward_5a <- step(lm5a_train, direction = "backward")
summary(data_reg_backward_5a)
data_reg_backward_pred_5a <- predict(data_reg_backward_5a, test5)
accuracy(data_reg_backward_pred_5a, test5$net_profit_margin)

#12.2.5 perform regression with fixed effects
data_reg_fixeff_5a <- felm(net_profit_margin ~ OPTION_AWARDS_BLK_VALUE + at + debt_to_equity_ratio + ceo_years | gvkey + fyear | 0 | gvkey + fyear, train5)
```

```
summary(data_reg_fixeff_5a)

#12.2.6 Checking for multicollinearity of controlled variables with VIF
vif_values_5a <- car::vif(lm5a_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#12.2.7 residual analysis
resid_5a <- lm5a_train$residuals
plot(train5$OPTION_AWARDS_BLK_VALUE, resid_5a)
#relationship between othcomp_at and residuals

fnorm_5a <- fitdist(resid_5a, "norm")
result_5a <- gofstat(fnorm_5a, discrete = FALSE)
result_5a
kscritvalue_5a <- 1.36/sqrt(length(train5$net_profit_margin))
kscritvalue_5a #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_5a)
plot(fnorm_5a) #residuals are not normally distributed
confint(lm5a_train, level = 0.95)

#12.2.8 Using stargazer for well-formatted regression output
list5a <- list(lm5a_train, data_reg_forward_5a, data_reg_backward_5a, data_reg_stepwise_5a)
stargazer(list5a, type="text",title="Regression Results", digits=4, omit = c("Constant"), no.space = TRUE, out="table5a.txt")

#12.3 ROA Regression Analysis

#12.3.1 regressing OPTION_AWARDS_BLK_VALUE with roa; control variables: at, debt to equity ratio and ceo years
lm5b_train <- lm(roa ~ OPTION_AWARDS_BLK_VALUE + at + debt_to_equity_ratio + ceo_years, train5)
summary(lm5b_train)

#12.3.2 Stepwise Regression
data_reg_stepwise_5b <- step(lm5b_train, direction = "both")
summary(data_reg_stepwise_5b)
data_reg_stepwise_pred_5b <- predict(data_reg_stepwise_5b, test5)
accuracy(data_reg_stepwise_pred_5b, test5$roa)
summary(data_reg_stepwise_pred_5b)

#12.3.3 Forward Regression
data_reg_forward_5b <- step(lm5b_train, direction = "forward")
summary(data_reg_forward_5b)
data_reg_forward_pred_5b <- predict(data_reg_forward_5b, test5)
accuracy(data_reg_forward_pred_5b, test5$roa)

#12.3.4 Backward Regression
data_reg_backward_5b <- step(lm5b_train, direction = "backward")
summary(data_reg_backward_5b)
data_reg_backward_pred_5b <- predict(data_reg_backward_5b, test5)
accuracy(data_reg_backward_pred_5b, test5$roa)

#12.3.5 perform regression with fixed effects
data_reg_fixeff_5b <- felm(roa ~ OPTION_AWARDS_BLK_VALUE + at + debt_to_equity_ratio + ceo_years | gvkey + fyear | 0 | gvkey + fyear, train5)
summary(data_reg_fixeff_5b)

#12.3.6 Checking for multicollinearity of controlled variables with VIF
vif_values_5b <- car::vif(lm5b_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#12.3.7 residual analysis
resid_5b <- lm5b_train$residuals
plot(train5$OPTION_AWARDS_BLK_VALUE, resid_5b)
#relationship between OPTION AWARDS BLK VALUE and residuals

fnorm_5b <- fitdist(resid_5b, "norm")
result_5b <- gofstat(fnorm_5b, discrete = FALSE)
result_5b
kscritvalue_5b <- 1.36/sqrt(length(train5$roa))
kscritvalue_5b #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_5b)
plot(fnorm_5b) #residuals are not normally distributed
confint(lm5b_train, level = 0.95)

#12.3.8 Using stargazer for well-formatted regression output
list5b <- list(lm5b_train, data_reg_forward_5b, data_reg_backward_5b, data_reg_stepwise_5b)
stargazer(list5b, type="text",title="Regression Results", digits=4, omit = c("Constant"), no.space = TRUE, out="table5b.txt")

#12.4 Current Ratio Regression Analysis

#12.4.1 regressing OPTION_AWARDS_BLK_VALUE with current ratio; control variables: at, asset turnover and debt to equity ratio
lm5c_train <- lm(current_ratio ~ OPTION_AWARDS_BLK_VALUE + at + debt_to_equity_ratio + asset_turnover, train5)
summary(lm5c_train)
```

```
#12.4.2 Stepwise Regression
data_reg_stepwise_5c <- step(lm5c_train, direction = "both")
summary(data_reg_stepwise_5c)
data_reg_stepwise_pred_5c <-
predict(data_reg_stepwise_5c, test5)
accuracy(data_reg_stepwise_pred_5c, test5$current_ratio)
summary(data_reg_stepwise_pred_5c)

#12.4.3 Forward Regression
data_reg_forward_5c <- step(lm5c_train, direction =
"forward")
summary(data_reg_forward_5c)
data_reg_forward_pred_5c <- predict(data_reg_forward_5c,
test5)
accuracy(data_reg_forward_pred_5c, test5$current_ratio)

#12.4.4 Backward Regression
data_reg_backward_5c <- step(lm5c_train, direction =
"backward")
summary(data_reg_backward_5c)
data_reg_backward_pred_5c <-
predict(data_reg_backward_5c, test5)
accuracy(data_reg_backward_pred_5c, test5$current_ratio)

#12.4.5 perform regression with fixed effects
data_reg_fixeff_5c <- felm(current_ratio ~
OPTION_AWARDS_BLK_VALUE + at + debt_to_equity_ratio +
asset_turnover | gvkey + fyear | 0 | gvkey + fyear, train5)
summary(data_reg_fixeff_5c)

#12.4.6 Checking for multicollinearity of controlled variables
with VIF
vif_values_5c <- car::vif(lm5c_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#12.4.7 residual analysis
resid_5c <- lm5c_train$residuals
plot(train5$OPTION_AWARDS_BLK_VALUE, resid_5c)
#relationship between OPTION AWARDS BLK VALUE and
residuals

fnorm_5c <- fitdist(resid_5c, "norm")
result_5c <- gofstat(fnorm_5c, discrete = FALSE)
result_5c
kscritvalue_5c <- 1.36/sqrt(length(train5$current_ratio))
kscritvalue_5c #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_5c)
plot(fnorm_5c) #residuals are not normally distributed
confint(lm5c_train, level = 0.95)

#12.4.8 Using stargazer for well-formatted regression output
list5c <- list(lm5c_train, data_reg_forward_5c,
data_reg_backward_5c, data_reg_stepwise_5c)
stargazer(list5c, type="text",title="Regression Results",
digits=4, omit = c("Constant"), no.space = TRUE,
out="table5c.txt")

#12.5 Debt to Equity Ratio Regression Analysis

#12.5.1 regressing OPTION_AWARDS_BLK_VALUE with debt
to equity ratio; control variables: at and net profit margin
lm5d_train <- lm(debt_to_equity_ratio ~
OPTION_AWARDS_BLK_VALUE + at + net_profit_margin +
asset_turnover, train5)
summary(lm5d_train)

#12.5.2 Stepwise Regression
data_reg_stepwise_5d <- step(lm5d_train, direction = "both")
summary(data_reg_stepwise_5d)
data_reg_stepwise_pred_5d <-
predict(data_reg_stepwise_5d, test5)
accuracy(data_reg_stepwise_pred_5d,
test5$debt_to_equity_ratio)
summary(data_reg_stepwise_pred_5d)

#12.5.3 Forward Regression
data_reg_forward_5d <- step(lm5d_train, direction =
"forward")
summary(data_reg_forward_5d)
data_reg_forward_pred_5d <- predict(data_reg_forward_5d,
test5)
accuracy(data_reg_forward_pred_5d,
test5$debt_to_equity_ratio)

#12.5.4 Backward Regression
data_reg_backward_5d <- step(lm5d_train, direction =
"backward")
summary(data_reg_backward_5d)
data_reg_backward_pred_5d <-
predict(data_reg_backward_5d, test5)
accuracy(data_reg_backward_pred_5d,
test5$debt_to_equity_ratio)

#12.5.5 perform regression with fixed effects
data_reg_fixeff_5d <- felm(debt_to_equity_ratio ~
OPTION_AWARDS_BLK_VALUE + at + net_profit_margin +
asset_turnover | gvkey + fyear | 0 | gvkey + fyear, train5)
summary(data_reg_fixeff_5d)

#12.5.6 Checking for multicollinearity of controlled variables
with VIF
vif_values_5d <- car::vif(lm5d_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables
```

```
#12.5.7 residual analysis
resid_5d <- lm5d_train$residuals
plot(train5$OPTION_AWARDS_BLK_VALUE, resid_5d)
#relationship between OPTION AWARDS BLK VALUE and
residuals

fnorm_5d <- fitdist(resid_5d, "norm")
result_5d <- gofstat(fnorm_5d, discrete = FALSE)
result_5d
kscritvalue_5d <-
1.36/sqrt(length(train5$debt_to_equity_ratio))
kscritvalue_5d #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_5d)
plot(fnorm_5d) #residuals are not normally distributed
confint(lm5d_train, level = 0.95)

#12.5.8 Using stargazer for well-formatted regression output
list5d <- list(lm5d_train, data_reg_forward_5d,
data_reg_backward_5d, data_reg_stepwise_5d)
stargazer(list5d, type="text",title="Regression Results",
digits=4, omit = c("Constant"), no.space = TRUE,
out="table5d.txt")

#13 multi-linear regression to analyse the effects of long term
incentives pay on profit margin, roa, liquidity ratio, debt to
equity ratio
#Dependent variable: net_profit_margin, roa, current_ratio,
debt_to_equity_ratio
#Independent variable: LTIP_at (long term incentives
pay/total assets)

data_reg_6d_clean <- data_reg_3d_clean %>%
mutate(LTIP_at = ifelse(is.na(LTIP_at),0,LTIP_at))

#13.1. choosing training and test dataset
set.seed(1)
train6 <- sample_frac(data_reg_6d_clean, 0.75)
test6 <- anti_join(data_reg_6d_clean, train6)

#13.2. regressing LTIP_at with profit margin; control
variables: at, debt to equity ratio, ceo_years
lm6a_train <- lm(net_profit_margin ~ LTIP_at + at +
debt_to_equity_ratio + ceo_years, train6)
summary(lm6a_train)

# 13.2.1. perform stepwise regression
data_reg_stepwise_6a <- step(lm6a_train, direction = "both")
summary(data_reg_stepwise_6a)
data_reg_stepwise_pred_6a <-
predict(data_reg_stepwise_6a, test6)
accuracy(data_reg_stepwise_pred_6a,
test6$net_profit_margin)

# 13.2.2. perform forward regression
data_reg_forward_6a <- step(lm6a_train, direction =
"forward")
summary(data_reg_forward_6a)
data_reg_forward_pred_6a <- predict(data_reg_forward_6a,
test6)
accuracy(data_reg_forward_pred_6a,
test6$net_profit_margin)

# 13.2.3. perform backward regression
data_reg_backward_6a <- step(lm6a_train, direction =
"backward")
summary(data_reg_backward_6a)
data_reg_backward_pred_6a <-
predict(data_reg_backward_6a, test6)
accuracy(data_reg_backward_pred_6a,
test6$net_profit_margin)

#13.2.4. perform regression with fixed effects
data_reg_fixeff_6a <- felm(net_profit_margin ~ LTIP_at + at +
debt_to_equity_ratio + ceo_years | gvkey + fyear |0 | gvkey
+ fyear, train6)
summary(data_reg_fixeff_6a)

#13.2.5. Checking for multicollinearity of controlled variables
with VIF
vif_values_6a <- car::vif(lm6a_train, type = "predictor") #no
multicollinearity issues as VIF values remain low across
variables

#13.2.6 residual analysis
resid_6a <- lm6a_train$residuals
plot(train6$LTIP_at, resid_6a) #relationship between LTIP_at
and residuals

fnorm_6a <- fitdist(resid_6a, "norm")
result_6a <- gofstat(fnorm_6a, discrete = FALSE)
result_6a
kscritvalue_6a <- 1.36/sqrt(length(train6$net_profit_margin))
kscritvalue_6a #KS statistic is more than kscrit value, we can
reject the null hypothesis that residuals are normally
distributed
summary(fnorm_6a)
plot(fnorm_6a) #residuals are not normally distributed
confint(lm6a_train, level = 0.95)

#13.2.7. Using stargazer for well-formatted regression output
stargazer(lm6a_train, data_reg_forward_6a,
data_reg_backward_6a, data_reg_stepwise_6a,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table6a.txt")
```

```
#13.3. regressing LTIP_at with return on assets; control variables: at, debt to equity ratio, ceo_years
lm6b_train <- lm(roa ~ LTIP_at + at + debt_to_equity_ratio + ceo_years, train6)
summary(lm6b_train)

# 13.3.1. perform stepwise regression
data_reg_stepwise_6b <- step(lm6b_train, direction = "both")
summary(data_reg_stepwise_6b)
data_reg_stepwise_pred_6b <-
predict(data_reg_stepwise_6b, test6)
accuracy(data_reg_stepwise_pred_6b, test6$roa)

# 13.3.2. perform forward regression
data_reg_forward_6b <- step(lm6b_train, direction =
"forward")
summary(data_reg_forward_6b)
data_reg_forward_pred_6b <- predict(data_reg_forward_6b,
test6)
accuracy(data_reg_forward_pred_6b, test6$roa)

# 13.3.3. perform backward regression
data_reg_backward_6b <- step(lm6b_train, direction =
"backward")
summary(data_reg_backward_6b)
data_reg_backward_pred_6b <-
predict(data_reg_backward_6b, test6)
accuracy(data_reg_backward_pred_6b, test6$roa)

#13.3.4. perform regression with fixed effects
data_reg_fixeff_6b <- felm(roa ~ LTIP_at + at +
debt_to_equity_ratio + ceo_years | gvkey + fyear |0 | gvkey
+ fyear, train6)
summary(data_reg_fixeff_6b)

#13.3.5. Checking for multicollinearity of controlled variables with VIF
vif_values_6b <- car::vif(lm6b_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#13.3.6 residual analysis
resid_6b <- lm6b_train$residuals
plot(train6$LTIP_at, resid_6b) #relationship between LTIP_at and residuals

fnorm_6b <- fitdist(resid_6b, "norm")
result_6b <- gofstat(fnorm_6b, discrete = FALSE)
result_6b
kscritvalue_6b <- 1.36/sqrt(length(train6$roa))
kscritvalue_6b #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_6b)
plot(fnorm_6b) #residuals are not normally distributed
confint(lm6b_train, level = 0.95)

#13.3.7. Using stargazer for well-formatted regression output
stargazer(lm6b_train, data_reg_forward_6b,
data_reg_backward_6b, data_reg_stepwise_6b,
type="text",title="Regression Results",omit = c("Constant"),
digits=4,  no.space = TRUE, out="table6b.txt")

#13.4. regressing LTIP_at with current ratio; control variables: at, debt to equity ratio, assets turnover
lm6c_train <- lm(current_ratio ~ LTIP_at + at +
debt_to_equity_ratio + asset_turnover, train6)
summary(lm6c_train)

# 13.4.1. perform stepwise regression
data_reg_stepwise_6c <- step(lm6c_train, direction = "both")
summary(data_reg_stepwise_6c)
data_reg_stepwise_pred_6c <-
predict(data_reg_stepwise_6c, test6)
accuracy(data_reg_stepwise_pred_6c, test6$current_ratio)

# 13.4.2. perform forward regression
data_reg_forward_6c <- step(lm6c_train, direction =
"forward")
summary(data_reg_forward_6c)
data_reg_forward_pred_6c <- predict(data_reg_forward_6c,
test6)
accuracy(data_reg_forward_pred_6c, test6$current_ratio)

# 13.4.3. perform backward regression
data_reg_backward_6c <- step(lm6c_train, direction =
"backward")
summary(data_reg_backward_6c)
data_reg_backward_pred_6c <-
predict(data_reg_backward_6c, test6)
accuracy(data_reg_backward_pred_6c, test6$current_ratio)

#13.4.4. perform regression with fixed effects
data_reg_fixeff_6c <- felm(current_ratio ~ LTIP_at + at +
debt_to_equity_ratio + asset_turnover | gvkey + fyear |0 |
gvkey + fyear, train6)
summary(data_reg_fixeff_6c)

#13.4.5. Checking for multicollinearity of controlled variables with VIF
vif_values_6c <- car::vif(lm6c_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#13.4.6 residual analysis
resid_6c <- lm6c_train$residuals
```

```
plot(train6$LTIP_at, resid_6c) #relationship between LTIP_at and residuals

fnorm_6c <- fitdist(resid_6c, "norm")
result_6c <- gofstat(fnorm_6c, discrete = FALSE)
result_6c
kscritvalue_6c <- 1.36/sqrt(length(train6$current_ratio))
kscritvalue_6c #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_6c)
plot(fnorm_6c) #residuals are not normally distributed
confint(lm6c_train, level = 0.95)

#13.4.7. Using stargazer for well-formatted regression output
stargazer(lm6c_train, data_reg_forward_6c, data_reg_backward_6c, data_reg_stepwise_6c, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table6c.txt")

#13.5. regressing LTIP_at with debt to equity ratio; control variables: at, net profit margin, assets turnover
lm6d_train <- lm(debt_to_equity_ratio ~ LTIP_at + at + net_profit_margin + asset_turnover, train6)
summary(lm6d_train)

# 13.5.1. perform stepwise regression
data_reg_stepwise_6d <- step(lm6d_train, direction = "both")
summary(data_reg_stepwise_6d)
data_reg_stepwise_pred_6d <- predict(data_reg_stepwise_6d, test6)
accuracy(data_reg_stepwise_pred_6d, test6$debt_to_equity_ratio)

# 13.5.2. perform forward regression
data_reg_forward_6d <- step(lm6d_train, direction = "forward")
summary(data_reg_forward_6d)
data_reg_forward_pred_6d <- predict(data_reg_forward_6d, test6)
accuracy(data_reg_forward_pred_6d, test6$debt_to_equity_ratio)

# 13.5.3. perform backward regression
data_reg_backward_6d <- step(lm6d_train, direction = "backward")
summary(data_reg_backward_6d)
data_reg_backward_pred_6d <- predict(data_reg_backward_6d, test6)
accuracy(data_reg_backward_pred_6d, test6$debt_to_equity_ratio)

#13.5.4. perform regression with fixed effects
data_reg_fixeff_6d <- felm(debt_to_equity_ratio ~ LTIP_at + at + net_profit_margin + asset_turnover | gvkey + fyear |0 | gvkey + fyear, train6)
summary(data_reg_fixeff_6d)

#13.5.5. Checking for multicollinearity of controlled variables with VIF
vif_values_6d <- car::vif(lm6d_train, type = "predictor") #no multicollinearity issues as VIF values remain low across variables

#13.5.6 residual analysis
resid_6d <- lm6d_train$residuals
plot(train6$LTIP_at, resid_6d) #relationship between LTIP_at and residuals

fnorm_6d <- fitdist(resid_6d, "norm")
result_6d <- gofstat(fnorm_6d, discrete = FALSE)
result_6d
kscritvalue_6d <- 1.36/sqrt(length(train6$debt_to_equity_ratio))
kscritvalue_6d #KS statistic is more than kscrit value, we can reject the null hypothesis that residuals are normally distributed
summary(fnorm_6d)
plot(fnorm_6d) #residuals are not normally distributed
confint(lm6d_train, level = 0.95)

#13.5.7. Using stargazer for well-formatted regression output
stargazer(lm6d_train, data_reg_forward_6d, data_reg_backward_6d, data_reg_stepwise_6d, type="text",title="Regression Results",omit = c("Constant"), digits=4,  no.space = TRUE, out="table6d.txt")
```