Decoding Firm Profitability: The Role of Financial Indicators

Monika Baloda
A. Gary Anderson Graduate School of Management, University of California, Riverside

Abstract

Finance and Profit are the two of the most important words talked about in the corporate sector. I delve into some fascinating ideas on the intersection of these topics in this project. I do panel data analysis to investigate this. My research question is "what matters for profits". I aim to find out the main variables that affect profitability out of around 75 financial indicators. I use LASSO and random forest methods to achieve this. One appealing feature of my panel data analysis is that I deal with big data (around 2.3 million observations) containing information on over 18000 firms for over 40 years. Therefore, my results become important for someone interested in the general economy in the united states. I discuss the steps of data wrangling, pre-processing, missing value treatment, and exploratory data analysis (EDA). I discuss the merits and shortcomings of my work within each section to correctly interpret my results.

Keywords: Financial Ratios, Big data, Random Forest, Variable Selection, LASSO, Model Evaluation, R

1 Introduction

Everyone in the corporate sector talk about Finance, Profit, and Technology. Naturally, they are very attractive topics to work on. I, in this paper dig deeper to find out some insights on these topics. This work contains two broadly separate sections. The work deals with *big-data*, covering over 40 years on around 18000 firms with quarterly frequency, thereby around 2.3 million observations. I use the standard methods developed for panel data in this section to analyze our research questions. In the remaining introduction, I discuss motivation, narrowing down the research question, and paper organization.

Motivation– The big panel data work is motivated by the fact that there is no ambiguity in saying that profit is the single most important variable a firm care about. An investor in the firm care about its profitability. The fundamental source of stock returns of the company is its profitability. However, profits are realized ex-post, i.e. we know it after all the business activities have happened. Therefore, the interesting problem is to predict future profits and know about the factors that affect it significantly. The 'fundamentals' of a company give a good idea about the future profitability of a firm. However, how to judge the fundamentals is a question worth asking. Investors look at the financial ratio indicators of a firm to assess its fundamentals and future prospects. For example, a firm with a lower debt-to-market-value ratio can easily be seen as having a lesser risk of failure relative to one with a high debt-to-market-value ratio firm. By exploring the patterns and trends, this research aims to provide valuable insights into the relationship between corporate profits and financial indicators to help investors make more informed decisions when allocating their capital. We do variation selection to pick the important variables using LASSO. We use the *random forest* method to further rank the importance of variables according to how they matter for profitability. By selecting the best variation among a set of options, businesses can minimize risk, increase efficiency, and improve their competitive position in the market. Some non-linear methods may be computationally very time taking in big-data regime, therefore, we need to learn dealing with this problem. I learned this new methods through this project.

Concrete Research Questions:– I collect information on around 75 financial indicators and more than 18000 firms. I define my target or response variable to be *gross-profit.* I apply the various methods to get to know what set of variables matters the most for my response variable. In particular, I use the LASSO method to shrink down the coefficient of not-so-important variables to zero and hence get the list of remaining important variables for my response variable.

Paper Organization:– Continuing further, in section-3, we investigate the crucial role of financial indicators in decoding firm profitability. Here, we start with describing data, definitions of variables, data wrangling, and pre-processing(to treat missing values) in section 3.1. We further did Exploratory Data Analysis of the target and predictor variable where we discussed their histograms, box plots, summary statistics, and correlation analysis in section 3.2. Further, in section 3.3, we conducted variable screening through the utilization of the Lasso method and discussed its results. After the screening, we moved to section 3.4, where we ranked the variables according to their importance score using random forest and discussed the results. Looking at all the results that we got from the above sections, we wrote inferences and discussion under section 3.5. Finally, in section 4, we concluded the two parts of this paper after a discussion of important insights for the US economy.

2 Data

2.1 Data Source

We downloaded the data from the Financial Ratios Firm Level by WRDS(Wharton Research Data Services). WRDS is a web-based platform designed to provide researchers access to a wide range of financial, accounting, economic, and marketing data from various sources. Academic institutions and corporate researchers use it for their financial research needs.

2.2 Variables and Definitions

our original data comprises of 2379290 observations and 76 variables. We treated missing values and removed outliers, which finally gives us 1942369 observations and 43 variables in total. Since, we are interested in profitability, we choose we chose *Gross profit* as our main variable representing profitability of a firm. Some other factors may be particular to firms or states, therefore we don't want to deal with those therefore gross-profit is our choice for profitability variable. This variable gross profit helps companies to track their financial performance, make informed decisions about pricing and cost control, to remain competitive in the marketplace. It is difficult to write all the variables from the data so we are writing a few variables that can describe most of the information. We have the following variables: Profit Variables like *Gross Profit/Total Assets(GProf), gpm(Gross Profit Margin)* etc. Liability or debt related ratio variables like *curr_debt(Current Liabilities/Total Liabilities), lt_debt(Long-term Debt/Total Liabilities)* etc. Liquidity variables like *Price/Cash flow(pcf), Cash ratio,* etc. Return variables like *Return on Capital Employed(roce), roe (Return on Equity)* etc. Other variables like tax, sales, and aggregate variables. Detailed information including abbreviation about all the variables is available in the supplementary section.

2.3 Data Wrangling and Pre-Processing

This section is mainly dedicated to the treatment of missing values in our data. Our original dataset had 2,379,290 observations and 71 variables. We find the missing values (NAs) by column to see which variable is better to drop. We see that there are a disproportionately high number of missing values in certain columns. Therefore, it may be a wise thing to drop a variable having lot of missing values rather than including it, because it may decrease our overall complete cases. We choose a threshold equal to 7.5% i.e. we drop a variable if its missing values are more than 7.5 percent of total. This choice of threshold could be 5% or 10%, but we choose it in a way such that maximum number of variables can be retained and minimum missing values remain in the data. We found 7.5% to be an appropriate value for this goal.One can do treatment of missing value by company ID (gvkey) as well, but for simplicity, we leave this exercise. From this point, we just drop the missing values because they are little portion of total and are likely to not affect the analysis.

Finally, we get complete observation by deleting all NAs in the dataframe. After treating missing values, our new dataset now contains 1942369 observations and 43 variables. The number of unique firms in our data are 18391 .Total percentage of deleted observations from missing value treatment are 18.36.

3 Exploratory Data Analysis (EDA)

In this section, we get to know our variables more closely through looking at their distributions and summary statistics.

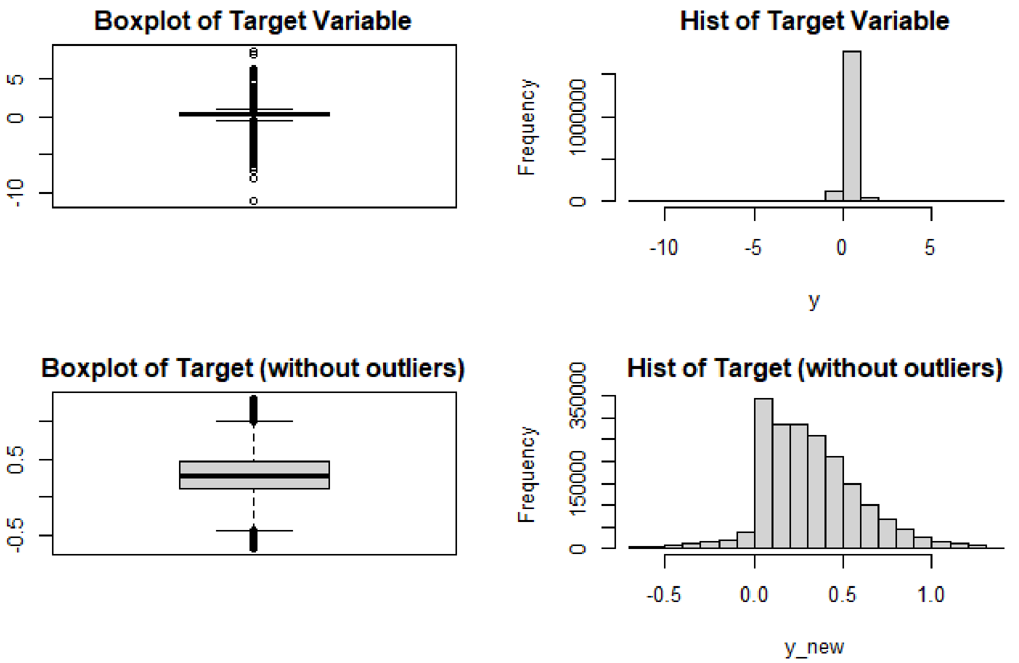
3.1 Target variable

Target variables are important because they help to establish the purpose of the model, and guide the selection of appropriate input variables or features. Our target variable is *GProf* i.e. gross profit. We will see the box plot of the target variable showing the distribution of values for the target variable, then histogram that provides a visual representation of the frequency distribution of the values in the target variable.

Table 1: Summary Statistics of Target variable

Variable	Min	1st Q	Median	Mean	3rd Q	Max.
Gross Profit(original)	-11.26	0.10	0.28	0.31	0.47	8.72
Gross Profit(without outliers)	-0.68	0.11	0.28	0.31	0.47	1.31

Figure 1: Boxplot and Histogram of Gross Profit

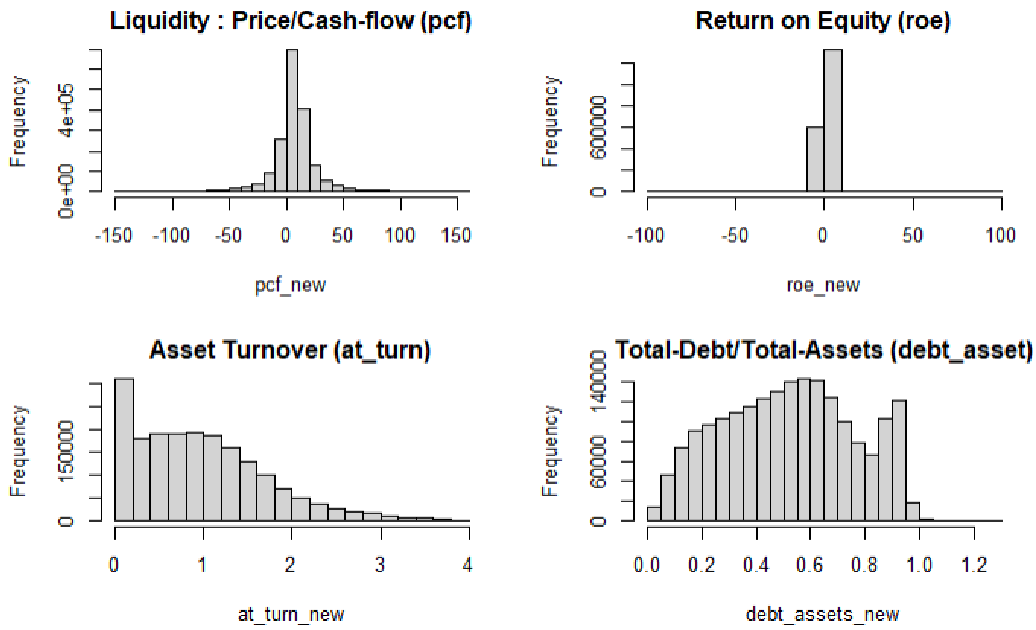


The box plot and histogram of the target variable after removing outliers show the distribution of values for the target variable, with the outliers removed based on a threshold of 3 standard deviations from the mean. By removing the outliers, the new box plot and histogram can provide a clearer picture of the distribution of the data that is not influenced by any extreme values.

3.2 Predictors

These are the variables that are used to make predictions or to explain variation in a dependent variable. We present he histograms of some of the predictors after deleting the outlier observations. We define outliers to be falling outside the mean plus-minus three standard deviation range. We have many predictor variables like *Return on Equity(roe), Return on Capital Employed(roce), Price/Cash flow(pcf),* etc.

Figure 2: Histogram of Predictor variables (after removing outliers)



Interpretation of graphs– *pcf* graph shows normal distribution and important property of the normal distribution is that it is characterized by two parameters: the mean (average) and the standard deviation (a measure of the spread of the data around the mean). If a variable in a business project follows a normal distribution, then we can use these parameters to make predictions and estimate probabilities about the variable. *roe* graph falls between dirac-delta and normal distribution. If a variable has a direct-delta histogram, it means that the variable takes on only a few distinct values, with most of the data points falling into one of those categories. *at_turn* histogram shows pareto/power-law, which means that the variable exhibits a specific pattern of frequency of occurrence. Here a few extreme values occur very frequently, while most other values occur rarely. *debt* shows asset-bimodal distribution as variable has two distinct peaks in its distribution, indicating that it has two different modes or values that occur most frequently.

Table 2: Summary Statistics of Predictor variable

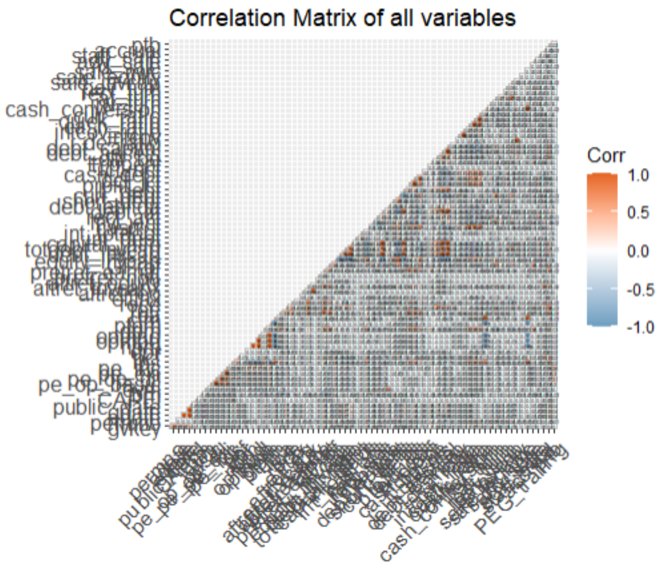
Variable	Min	1st Q	Median	Mean	3rd Q	Max.
pcf	-663.100	-0.998	6.830	7.003	14.660	650.279
roe	-14400	-0.059	0.076	-0.009	0.150	10532.800
at_turn	0.00	0.391	0.899	1.052	1.456	69.284
debt_assets	0.00	0.3290	0.5260	0.5217	0.7030	14.3170

Interpretation of table-2–From the table for the variable pcf, the minimum value is -663.1, the first quartile (25th percentile) is -0.998, the median (50th percentile) is 6.83, the mean is 7.003, the third quartile (75th percentile) is 14.66, and the maximum value is 650.279. This suggests that the distribution of pcf is positively skewed (i.e., has a longer right tail) as the mean is greater than the median. Whereas, for the variable roe, the minimum value is -14400, the first quartile is -0.059, the median is 0.076, the mean is -0.009, the third quartile is 0.150, and the maximum value is 10532.8. This suggests that the distribution of roe is also positively skewed as the mean is less than the median. For the variable at turn and debt assets, the distribution is positively skewed.

3.3 Correlation Analysis

Correlation analysis measures the degree to which two or more variables are related to each other. It helps us to determine whether the relationship between two variables is positive or negative, and the strength of the relationship. It can be used to identify patterns and relationships between variables in market research that is we can gain valuable insights into consumer behavior, preferences, and trends.

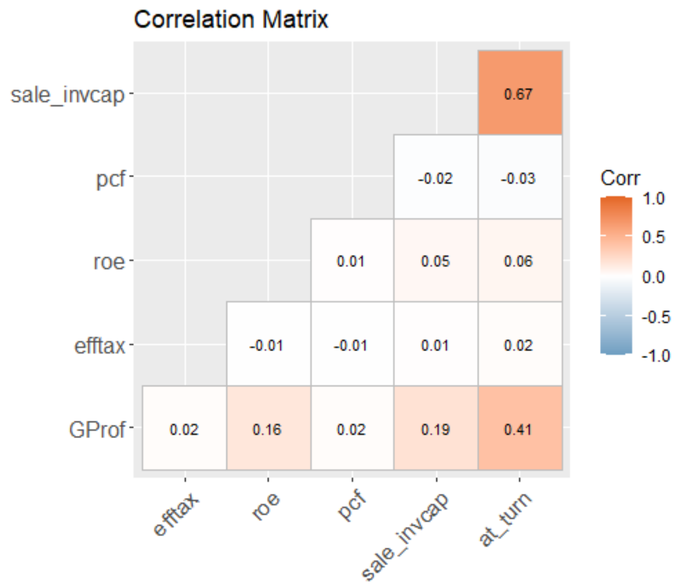
Figure 3: Correlation matrix of all variables with main variable(Gross Profit)



In Figure-3, the resulting plot shows the correlation between all numeric variables in the original data-set. Here, correlation matrix is not clear that is why it is convenient to use some selected predictor variables that can help in making visualization easy to interpret.

In Figure-4, In order to get better correlation we select five predictive variables that is *efftax, roe, pcf, sale_invcap, at_turn*. *G_prof* as our response variable.We can see the correlation between at_turn and sale_invcap to be 0.67 which is highly positive and correlation between pcf and at_turn to be -0.03 showing negative correlation.

Figure 4: Correlation matrix:Gross Profit with predictor variables



4 Main Analysis : Variable Screening Using LASSO

4.1 Model

Linear regression is a type of supervised learning algorithm that helps to model the relationship between an independent variable (also called as the predictor variable) and a dependent variable (also known as the response variable). In linear regression, our goal is to find the linear function that best fits the data. The equation for linear regression can be written as:

$$\min_{\{\beta_0,\beta_1,\ldots,\beta_p\}} \quad \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2$$

This equation best fits the data but doesn't provide most variables. To do so, we use LASSO, where use of a penalty term in the loss function has the effect of shrinking the coefficients of less important variables towards zero, effectively removing them from the model. This results in a sparse model with only the most important variables included.

Our optimization problem can be written as:

$$\min_{\{\beta_0,\beta_1,\ldots,\beta_p\}} \quad \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

Here, $|\beta_j|$ the penalty will shrink all of the coefficients towards zero. Lasso shrinks the coefficient estimates towards zero. In the case of the lasso, the penalty has the effect of forcing some of the coefficient estimates to be exactly equal to zero when the tuning parameter λ is sufficiently large. Hence, much like best subset selection, the lasso performs variable selection.

When $\lambda = 0$ then the lasso simply gives the least squares fit, and when λ becomes sufficiently large, the lasso gives the null model in which all coefficient estimates equal zero.

Variable selection- To illustrate the situation, the least squares solution is marked as $\widehat{\beta}$. Here blue diamond represent lasso constraints.

The ellipses centered around around β shows the regions of constant RSS. A common value of RSS is shared by all the points on a given ellipse. When the ellipses expand and moved away from the least squares coefficient estimates, the value of RSS also increases. The lasso constraint has corners at each of the axes, and therefore the ellipse will often intersect the constraint region at an axis. In such a situation, one of the coefficients will have a value of zero. In higher dimensions, it is possible that multiple coefficient estimates could be equal to zero at the same time.

4.2 Result

We used the *glmnet* package in R to perform the Lasso regression, and used cross-validation to select the optimal value of the penalty parameter. The Lasso regression identified twelve predictor variables that were significantly associated with gross profit. These variables are are given in the table of regression coefficients.

Figure 5: Counters of error and constraint functions for the lasso

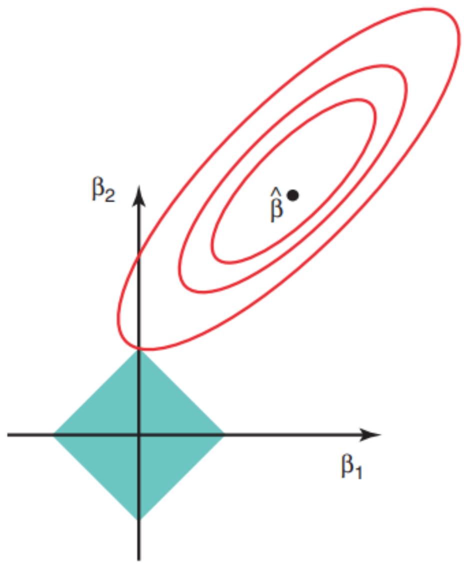


Table 3: Shrinked Regression Coefficients Given by LASSO

Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
(Intercept)	0.3123	bm	0	evm	0
pe_exi	0	pe_inc	0	ps	-0.0309
pcf	0	npm	0	opmbd	0
opmad	0	gpm	0.0022	ptpm	0
cfm	0	roa	0.1370	roe	0
roce	0	aftret_eq	0	aftret_invcapx	0
aftret_equity	0	equity_invcap	0	debt_invcap	0
totdebt_invcap	0	capital_ratio	-0.0048	cash_lt	-0.0108
debt_at	0	debt_ebitda	0	lt_debt	-0.0210
cash_debt	0.0044	lt_ppent	0	dltt_be	0
debt_assets	-0.0456	debt_capital	-0.0013	de_ratio	0
at_turn	0.1304	sale_invcap	-0.0007	sale_equity	0
rd_sale	0	adv_sale	0	staff_sale	0
accrual	-0.000006	ptb	0.0411		

The coefficient for the intercept is 0.312. Number of non-zero variables after lasso comes to be 14.This means that the Lasso regression has selected a subset of the independent variables that are most strongly related to the dependent variable and set the rest to zero, effectively performing feature selection.These non-zero variables are given in the table. The size and direction of the coefficients indicate the strength and direction of the relationship between each independent variable and the dependent variable. For example, the coefficient for roa is 0.137, which suggests that a one-unit increase in roa is associated with a 0.137-unit increase in the dependent variable, all else being equal.

5 Ranking Variables According to their Importance for Target

LASSO gives us the variables which matter the most for our target variables. However, one might be interested in knowing the ranking of importance of these variables. We use *random-forest*'s importance score. We discuss the method and results below:

5.1 What is Random Forest

In a Random Forest, a collection of decision trees are built on randomly selected subsets of the original data. Each decision tree in the forest is built independently of the others, and they are combined to produce the final prediction. Random forest builds many decision trees on randomly selected data subsets, which helps to prevent overfitting and improves the model's ability to generalize to new data. Combining the results of multiple trees can reduce errors and biases in individual trees, making the model more reliable. The importance score of each variable can be calculated in random forest based on its frequency of appearance in node splits across all the decision trees

Importance score: The importance score in a random forest is often based on the occurrence of a variable in the node splits of the decision trees in the forest. This score is calculated by counting how many times a variable is chosen as the best one to split a node across all the trees in the forest. Once the importance score is calculated for each variable, they can be ranked in order of importance, with the most important feature having the highest score. This ranking can help to identify the most relevant features in a dataset and improve the interpretability of the model.

5.2 Results

In order to identify the key predictors in our model, we utilized the Random Forest to generate importance scores for each variable. As a result of this analysis, we have obtained a comprehensive table that displays the importance scores for all variables considered in our study. We refer to the table below to see the importance scores for each variable.

Table 4: Importance score of variables

Variable	Importance score	Variable	Importance score	Variable	Importance score
gpm	57.51	at_turn	25.22	roa	17.37
sale_invcap	14.99	debt_assets	10.57	cash_lt	10.39
lt_debt	10.32	cash_debt	9.5	capital_ratio	8.29
ps	7.77	debt_capital	7.35	ptb	7.28
accrual	3.48				

These results appear to be the variable importances generated by a random forest regression model. The number next to each variable indicates its relative importance in predicting the target variable, which is not provided in the given information. The variable with the highest importance score is *at_turn* with a score of 25.22, followed by *roa* with a score of 17.37, *sale_invcap* with a score of 14.99, and so on. The lowest variable importance score is *accrual* with a score of 3.48. The variable importance can be used to identify the most important predictors of the target variable, which can help in variable selection and identifying the most relevant factors for further analysis.

6 Inference and Discussion

We got a total of twelve important variables from LASSO regression. Identifying these 12 predictor variables suggests that they are likely to be essential drivers of gross profit from the business point of view. Therefore, focusing on these variables and understanding their impact on gross profit can help businesses make more informed decisions about resource allocation, marketing strategies, product development, and pricing policies, among others.

After using random forest we got the variable with the highest importance score is *gpm*, indicating that it is the most important variable in predicting the target variable. But this variable is a gross-profit margin which is an obvious indicator. However, the next important variables are asset turnover and return on assets. Followed by sales, debt, and cashflow variables. In particular, other variables that are relatively important in the model include *sale invcap, debt assets, cash lt, lt debt, cash debt, capital ratio, ps, debt capital, ptb*, and *accrual* in that order. Overall, the importance scores suggest that the variables *gpm, at turn, and roa*, are the most important predictors of the target variable, while the other variables also make a significant contribution to the model. These results can be used to guide further analysis and modeling efforts to improve the accuracy of the predictions.

7 Conclusion

In this project, I work on two different data regimes to test two different types of research questions. In the first part, I test the relationship between two-time series while in the second, I do analysis keeping a target variable in mind under the panel data regime.

In the time series analysis of testing whether technology leads the other sectors, I learned several lessons. This time series analysis have been a great learning experience for me. I realized how much pre-processing is needed before jumping to the main model. I started by saying that my goal in this exercise it to test whether the technology sector leads the non-technology sector. In order to do this test, we needed to do a lot of pre-processing. When you start an analysis, sometimes, the results guide you to do the next things. In my case, this is at least partially true. When I plotted the two series together, I found that there is an interesting decoupling after 2016 which further widened in 2019. This result motivated me to consider this project to be a full-fledged research paper later on working on volatility change. Well, I can consider that for later. The key takeaways can be very briefly summarized in the following two points: First, the technology index series (y_t) follows ARIMA(3,1,3) model. On the other hand, the non-technology index follows an ARIMA(2,1,4) model. The I=1 order implies that both the series are non-stationary but their first differenced series becomes stationary. Second, we have enough evidence that the technology index can be said a leading indicator for the remaining sectors of the economy. However, we caution the reader not to interpret these results as causal. Third, There is a very close relationship between tech (y_t) and non-tech (x_t) indices, however, we observed a decoupling after 2016 which further widened with covid shock in 2019. Fourth, There is a structural break in the tech index series in 2019 due to covid shock.

In the second part of the work, several insights are worth discussing. We got a total of twelve important variables from LASSO regression. Identifying these 12 predictor variables suggests that they are likely to be essential drivers of gross profit from the business point of view. Therefore, focusing on these variables and understanding their impact on gross profit can help businesses make more informed decisions about resource allocation, marketing strategies, product development, and pricing policies, among others. To learn more, we make use of random forests to rank the variables according to their importance in predicting the response variable. After using random forest, we found that the variables with the highest importance for profitability are: gross-profit margin, asset turnover, and return on asset. While gross-profit margin could be an obvious indicator for the gross-profit variable, the remaining variables give us meaningful insights. We note that sales, debt, and cashflow-related variables come in that order of importance. Therefore, the asset is the most important variable to focus on for profitability. If one wants to focus just on one variable then it should be an asset and then sales, debt, and cashflow. These results can be used to guide further analysis and modeling efforts to improve the accuracy of the predictions.

Overall, we do not claim our results to be causal, however, given the number of years and firms they cover, our results represent important insights for the US economy. I'm happy to find several ideas for further research to make causal claims. I also learned that sometimes it may be computationally very difficult to deal with big-data, therefore, we need to learn how to deal with it unlike usual data case.

References

[1] Hamilton, J. D. (2020). "Time series analysis". *Princeton university press*, hyperlink.
[2] Harvey, A. (2020). "Time series models for epidemics: Leading indicators, control groups and policy assessment". *NIESR UK's Disussion Papers*, 517. hyperlink
[3] Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). "The elements of statistical learning: data mining, inference, and prediction" (Vol. 2, pp. 1-758). *New York: Springer*.
[4] Romer, P. M. (1994). "The origins of endogenous growth". *Journal of Economic perspectives*, 8(1), 3-22.

Supplementary Material

Variable definitions of LASSO Analysis

- **Profit Variables**: Companies that generate consistent profits are better positioned for growth and long-term success, while investors and stakeholders are more likely to be attracted to companies with a strong profitability track record.
  1. *GProf*: Gross Profit/Total Assets, financial ratio that measures a company's ability to generate profit from its assets.
  2. *gpm*: Gross Profit Margin, financial ratio that measures a company's profitability by comparing its gross profit to its revenue.
  3. *npm*:Net Profit Margin, financial ratio that measures a company's profitability by comparing its net profit to its revenue.
  4. *opmad*: Operating Profit Margin After Depreciation, financial ratio that measures a company's profitability by comparing its operating profit after depreciation to its revenue
  5. *opmbd*: Operating Profit Margin Before Depreciation, financial ratio that measures a company's profitability by comparing its operating profit before depreciation to its revenue.
  6. *ptpm*: Pre-tax Profit Margin, is the amount of money a company has left over after deducting all expenses, except for income taxes, from its revenue.
  7. *Profit-let*: Profit Before Depreciation/Current Liabilities, financial metric that measures a company's ability to pay off its short-term obligations using its profits before accounting for non-cash expenses such as depreciation.

- **Liability or debt related ratios**: It refers to an obligation that a company or individual owes to another party. It is important for companies and individuals to maintain a solid financial position and manage their liabilities/debt effectively to ensure long-term financial stability.
  1. *curr_debt*: Current Liabilities/Total Liabilities
  2. *lt_ppent*: Total Liabilities/Total Tangible Assets
  3. *de_ratio*: Total Debt/Equity
  4. *debt_assets*: Total Debt/Total Assets
  5. *debt_at*: Total Debt/Total Assets
  6. *debt_capital*: Total Debt/Capital
  7. *lt_debt*: Long-term Debt/Total Liabilities
  8. *debt_invcap*: Long-term Debt/Invested Capital
  9. *totdebt_invcap*: Total Debt/Invested Capital
  10. *debt_ebitda*: Total Debt/EBITDA
  11. *dltt_be*: Long-term Debt/Book
- **Liquidity**: It is a measure of the ability of an individual or business to quickly convert an asset or security into cash without causing a significant change in its value. Companies maintaining strong liquidity positions are better able to meet their financial obligations, manage unexpected events, and take advantage of growth opportunities, which can help with sustained success over the long term.
  1. *pcf*: Price/Cash flow
  2. *ps*: Price/Sales
  3. *ptb*: Price/Book
  4. *fcf_ocf*: Free Cash Flow/Operating Cash Flow
  5. *ocf_lct*: Operating CF/Current Liabilities
  6. *cash_debt*: Cash Flow/Total Debt
  7. *cash_lt*: Cash Balance/Total Liabilities
  8. *cfm*: Cash Flow Margin
  9. *cash_conversion*: Cash Conversion Cycle
  10. *cash_ratio*: Cash Ratio
- **Capital**
  1. *capital_ratio*: Capitalization Ratio
- **Return**: It refers to the profit or loss on an investment or business activity, calculated as a percentage of the initial investment or capital. It helps to measure the performance of an investment or business activity over a given period of time.
  1. *aftret_eq*: After-tax Return on Average Common Equity
  2. *aftret_equity*: After-tax Return on Total Stockholders Equity
  3. *aftret_invcapx*: After-tax Return on Invested Capital
  4. *pretret_earnat*: Pre-tax Return on Total Earning Assets
  5. *pretret_noa*: Pre-tax return on Net Operating Assets
  6. *roa*: Return on Assets
  7. *roce*: Return on Capital Employed
  8. *roe*: Return on Equity
- **Tax**: Taxes are typically calculated as a percentage of income, property, goods, or services. Taxes are an important consideration for businesses, both as a legal obligation and as a factor that can impact cash flow, reputation, and competitive advantage.
  1. *efftax*: Effective Tax Rate
  2. *ptpm*: Pre-tax Profit Margin
  3. *intcov*: After-tax Interest Coverage

- **Sales**: The process of selling products or services in exchange for money or other consideration. These ratios are important financial metrics that provide insights into various aspects of a company's financial performance. They are a useful tool for investors, analysts, and management to assess a company's performance and make informed decisions.

  1. *sale_equity*: Sales/Stockholders Equity
  2. *sale_invcap*: Sales/Invested Capital
  3. *sale_nwc*: Sales/Working Capital

- **Aggregate**: It provides idea on scale or size of the company, e.g. if Company A had a net income of \$10,000 and Company B had a net income of \$1 million, their profit margin ratios would be 10% and 10%, respectively. This would suggest that both companies are equally profitable, even though Company B generates significantly more revenue in absolute amount. Hence, aggregate variables becomes important. In statistical jargon, aggregate variables are not normalized therefore gives a different information that are lost in ratios.

  1. *at_turn*: Asset Turnover
  2. *inv_turn*: Inventory Turnover
  3. *pay_turn*: Payables Turnover
  4. *rect_turn*: Receivables Turnover