ECON 2250 Final Report: Analysis of Financial Indicators of Stocks in 2014

Komei, Lim, Shreyas, Zel

Introduction

To start off it is important to provide a quick introduction into the project's motivations, dataset, description, and key variables (both outcome and indicator), and research question. Starting off by looking into our group's motivations, we were interested in finding out how companies are able to recover in a post recession economy. This question on the surface may seem boring, yet if answered can play a key role in policy decisions and how companies are able to navigate uncertain times. The way in which companies and governments respond to these dramatic changes play a big role in a country's well being, overall stability, and future investment decisions on both an individual and institutional level making this study both valuable and intriguing, hence serving as our motivation. After this, we realized that we needed to find a data set that we could use that preferably was both recent and directly followed a well known economic downturn. This was when we decided to settle on the 2009 financial crisis as this was an event that we all had lived through and had genuine interest in. Following this we found a data set that contained at least ten quantitative financial variables on 2000+ publicly traded companies in 2014. This data is primarily centered around key metrics such as revenue, profit margin, gross profit, R&D expenditure, Earnings Per Share (EPS), market capitalization, and operating expenses. As we started our analysis we soon realized that we needed to clean our data due to duplicates, missing values, de-listed companies, among other reasons that left us with a final dataset of 788+ companies which we would analyze.

With this new dataset, we found that it was important that we clearly define all variables that we planned to discuss within this project and report:

Outcome Variables:

Stock Return - The percentage change in a stock over the year of 2014, which we are using to predict movement of post rescission economies.

Predictor variables:

Revenue - The total income that a company makes before subtracting expenses or any other operating costs.

Profit - The net earnings of a company after all the expenses have been subtracted. This is basically Revenue - Total Expenses.

Profit Margin - This is similar to profit, but instead is the percentage of profit that is left after expenses.

Gross Profit - This is revenue minus the cost of goods sold, which is better explained as the amount that is left after production and delivery costs.

R&D Expenditure - This is the amount of money it takes to research and develop a product. This helps with long term growth and is included in the per piece sold cost.

Market Capitalization - This is the total value of the company which is basically the share price times the amount of shares available.

Earnings Per Share (EPS) - This is the total income per quarter divided by the shares available and is a key value for investors and company well being

Operating Expenses - This is the amount of money it takes for the business to maintain its operations and production with expenses like marketing being included.

Now that we fully understood the data provided to us along with the variables we needed to conduct a brief descriptive data analysis that found that the stock returns had a wide skewness with most companies having gains that were pretty small or within historical trends. There were also a couple of companies that were outliers with either extreme gains or losses. Additionally, the predictor variables themselves had substantial variation. Things like revenue and market cap had a wide variety of data across various industries while things like R&D and were more focused in technology based fields. All together after understanding our motivations, cleaning our data, and understanding some trends we made the following research question: "How do financial indicators like R&D expenditure, revenue, and profit relate to stock returns in 2014 and what can it tell us about post-recession economies?".

Methodology

Our analysis uses a financial dataset containing over 700 publicly traded companies from the year 2014. As described in the project materials, the dataset includes more than ten quantitative financial indicators such as revenue, profit margin, gross profit, earnings per share (EPS), operating expenses, market capitalization, and R&D expenditure. We restricted the sample to firms that were active and reporting financial data in 2014, and we removed duplicate entries and observations with missing values to ensure consistency and comparability across firms. These preprocessing steps produced a cleaned dataset of 788 observations prior to model selection. After cleaning the dataset, we standardized all continuous predictor variables using **z-score normalization**:

$z_i = \frac{x_i - \bar{x}}{s}$

where x_i is the original value, x is the sample mean of variable i, and s is the standard deviation. This transformation places the predictors on a common scale (mean 0, standard deviation 1), which facilitates numerical stability of the regressions and makes the magnitude of the estimated coefficients more directly comparable across financial indicators. We defined stock return (%) as the outcome variable for the regression analysis, following the project framing on page 4. Predictor variables were chosen to represent a diverse set of financial performance indicators, including: R&D expenditure, profit, market capitalization, revenue, EPS, operating expenses. These indicators were selected because they capture both long-term strategic investment (e.g., R&D) and short-term profitability and liquidity measures. Their inclusion allows us to test whether firm fundamentals were informative of stock returns in the immediate post-recession environment of 2014. All data preprocessing and regression analysis were conducted in RStudio, ensuring reproducibility through documented code and GitHub version tracking. After cleaning the dataset, we examined descriptive statistics—including skewness—for all variables to understand distributional properties that might influence model performance or interpretation.

To evaluate the relationship between financial indicators and stock returns, we employed simple linear regression models. For each predictor variable, we estimated the model:

$y = \alpha + \beta x$

where y is stock return (%) and x is one of the six financial indicators listed above. The approach outlined on page 4 guided this choice of model. For each coefficient, we computed the standard error and associated t-statistic:

$t = \frac{\hat{\beta}}{SE(\hat{\beta})}$

We used a significance level of =0.05, allowing us to test:

- **Null hypothesis :** The predictor has no relationship with stock return (=0).

- **Alternative hypothesis :** The predictor is correlated with stock return (0).

Each predictor was tested individually to isolate its marginal effect on returns. This iterative approach was selected because preliminary correlation checks did not indicate strong multi-collinearity, and the project's objective emphasized understanding the standalone relevance of each indicator.

For each regression model, we examined:

1. **Estimated coefficient size and sign**
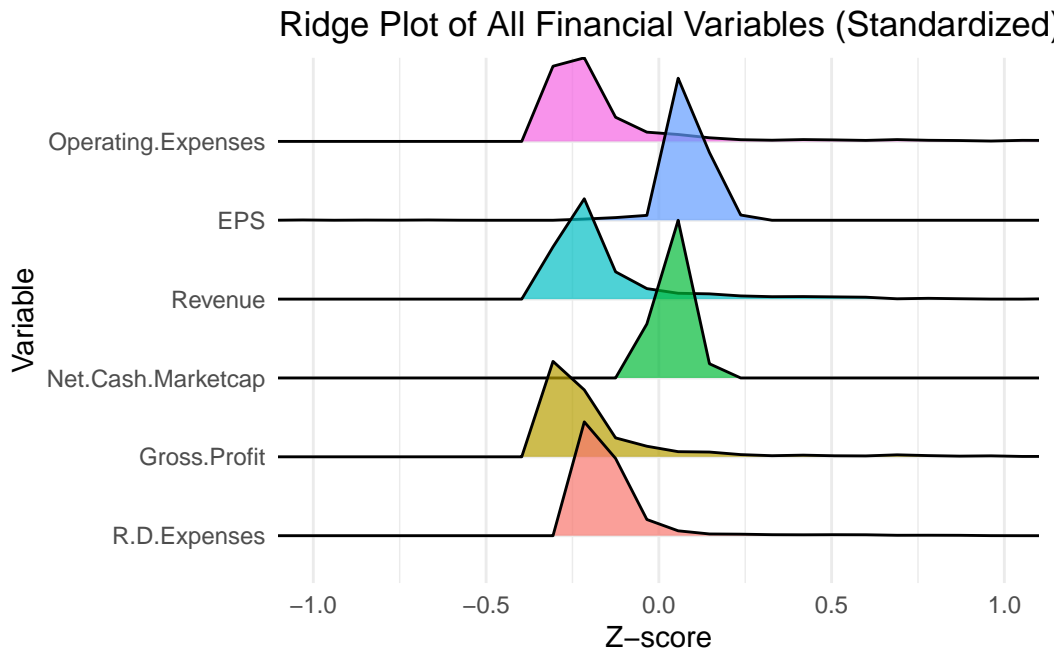2. **Confidence intervals**
3. **p-values**
4. R^2

Results

Figure I. Ridge plot showing z-scale normalized variable distributions.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	81.187424	59.00532	1.37593395	0.1692815
R.D.Expenses	-7.553160	82.06747	-0.09203598	0.9266957
Gross.Profit	-31.733905	365.00777	-0.08694035	0.9307438
Net.Cash.Marketcap	7.374059	59.05134	0.12487539	0.9006580
Revenue	-8.970612	138.57650	-0.06473401	0.9484042
EPS	10.073860	59.07427	0.17052874	0.8646435
Operating.Expenses	23.548167	355.50581	0.06623849	0.9472068

Figure II. Table of predictor coefficients used in linear regression model and respective estimates, standard error, t-value, and p-value. Multiple R-squared is also found to be extremely low at **0.0002**.
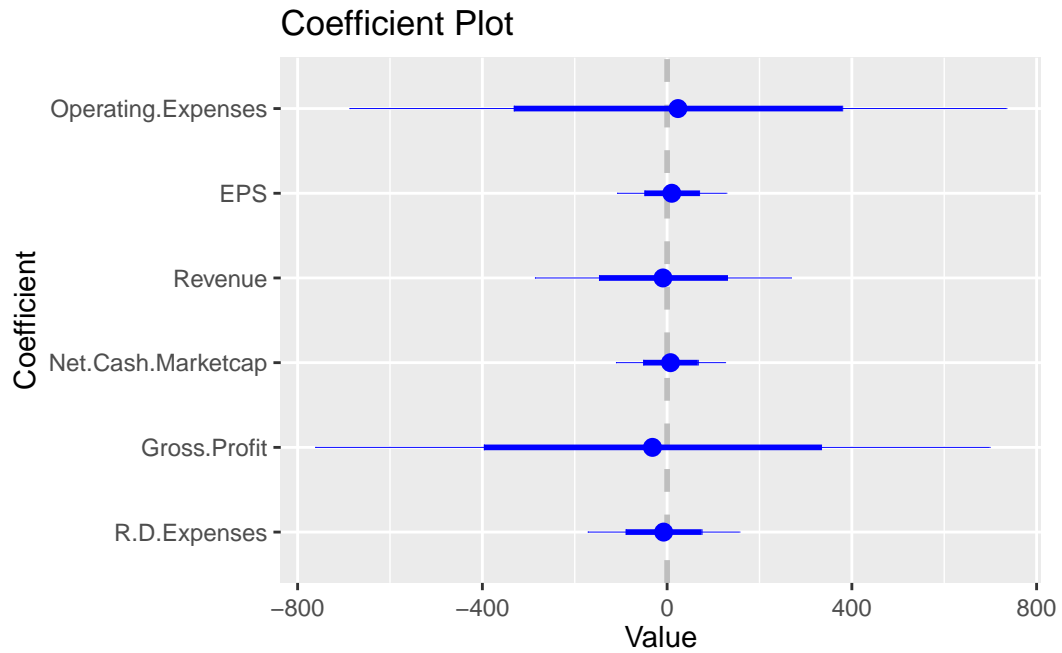

Figure III. Coefficient plot showing estimates of predictor coefficients (dot), along with the 50% confidence interval (bolded lines) and 95% confidence interval (lines). We can see that the null hypothesis (x=0 axis) lies well within the confidence intervals.

Discussion & Conclusion