

# ECON 2250 Final Report

Komei, Lim, Shreyas, Zel

## Introduction

ECON2250 Final Project running simple linear regression hypothesis test on correlation between corporate financial indicators to their stock price performance.

## 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,  $\bar{x}$  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  $\alpha = 0.05$ , allowing us to test:

- **Null hypothesis :** The predictor has no relationship with stock return ( $\beta = 0$ ).
- **Alternative hypothesis :** The predictor is correlated with stock return ( $\beta \neq 0$ ).

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 multicollinearity, 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

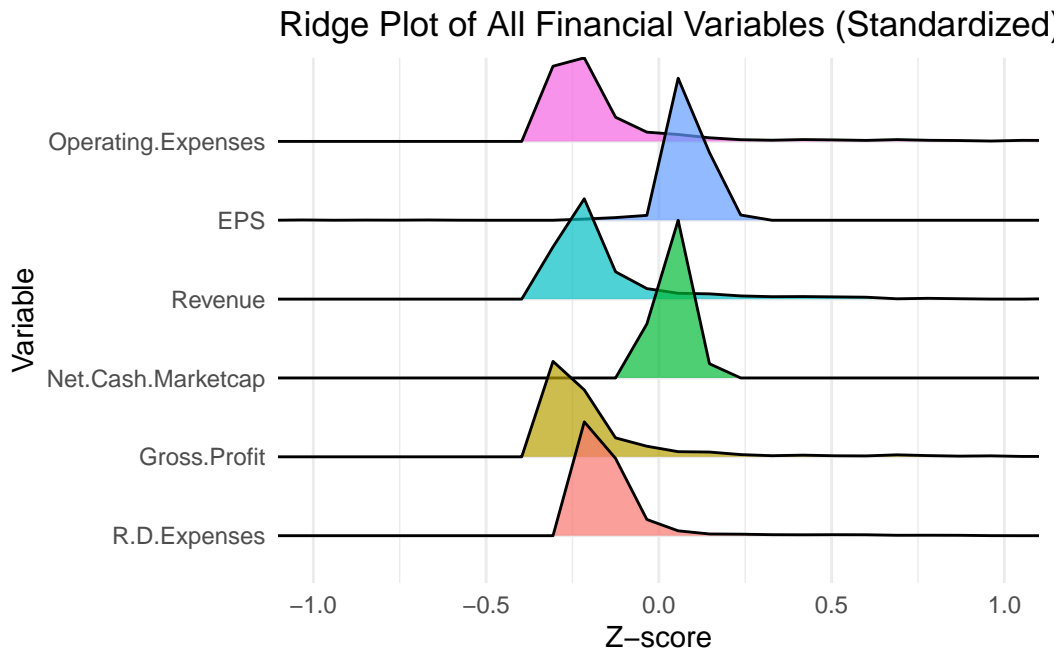


Figure I. Ridge plot showing z-scale normalized variable distributions.

Call:

```
lm(formula = X2015.PRICE.VAR.... ~ ., data = df_scaled)
```

Residuals:

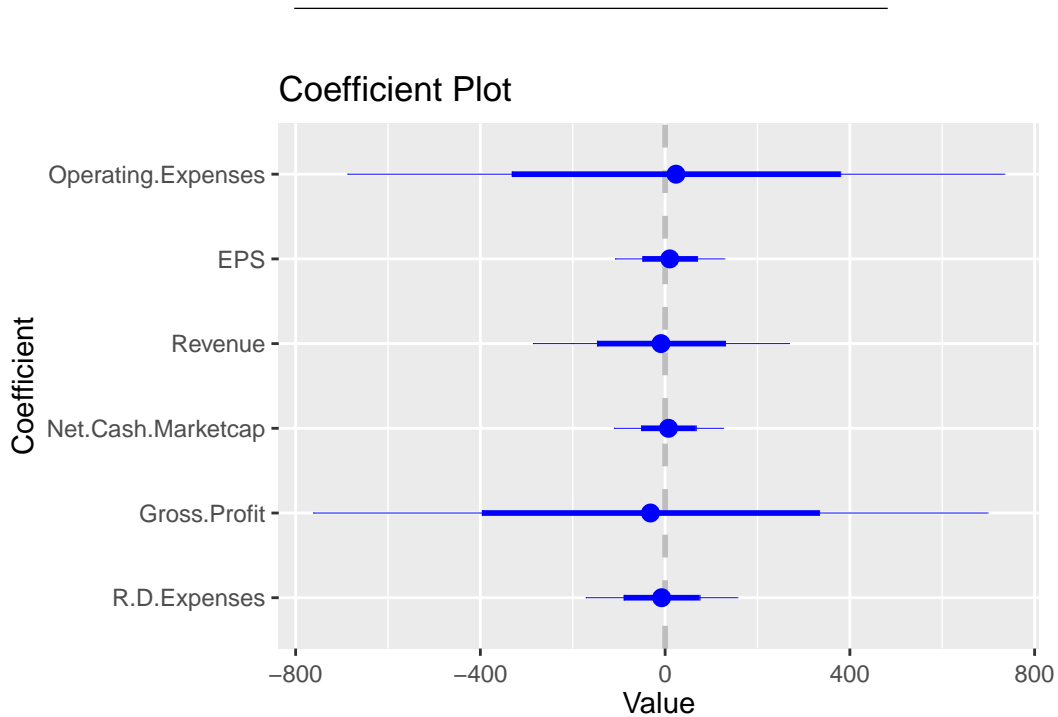
Min	1Q	Median	3Q	Max
-181	-116	-89	-57	40146

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	81.187	59.005	1.376	0.169
R.D.Expenses	-7.553	82.067	-0.092	0.927
Gross.Profit	-31.734	365.008	-0.087	0.931
Net.Cash.Marketcap	7.374	59.051	0.125	0.901
Revenue	-8.971	138.577	-0.065	0.948
EPS	10.074	59.074	0.171	0.865
Operating.Expenses	23.548	355.506	0.066	0.947

Residual standard error: 1570 on 701 degrees of freedom  
Multiple R-squared: 0.0002384, Adjusted R-squared: -0.008319  
F-statistic: 0.02786 on 6 and 701 DF, p-value: 0.9999

**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 shown 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