

The Capstone Project – ECON 5105

The capstone of this course is analyzing a publicly traded company. The goal is to demonstrate your understanding of data analysis, as presented. You do this project with your partner. Create a PowerPoint deck, that is, a set of slides that contain the following sections¹. The presentation should be a maximum of 15 slides (and hopefully less). Don't write anything in the notes section. If it is important enough to say, it is important enough to have on the actual slide.

Part 1 - Introduction:

In the first part of the introduction, explain to the reader the situation. Are you an investment advisor recommending someone should buy or sell stock in this company? Are you a CEO/CFO/CIO analyzing your own company? Are you a takeover specialist thinking of purchasing the entire company? These are only a few suggestions. Feel free to be creative. Being creative improves your grade. Once you decide, use the situation to make the whole report flow together in an exciting way. Even the professor does not want to see just page after page of regressions! Give the report context and meaning. Keep the audience interested. Make them want to read all the way through.

The second part of the introduction should describe the company. It need not be a USA company, but it MUST be publicly traded. Avoid Starbucks, Amazon, Twitter, and Apple. Generally, avoid recent start-ups. In a few paragraphs, please describe what the company does and why it is of interest. Include where the stock is traded, its ticker symbol and any other information your intended audience wants to know about the company. Do NOT cut and paste this from a financial website. Write it in your own words.

Part 2 - Return:

The capital asset pricing model asserts that an asset's (share price, portfolio, etc.) expected return is related to investments that could be made in a risk-free asset or a market basket of assets. The formula is

$$E(\text{Return} - \text{risk-free rate}) = \alpha + \beta (\text{Market basket return} - \text{risk-free rate})$$

β describes how the asset varies with the market basket. The market basket is a broad-based collection of diversified assets. For example, the S&P 500 index represents such a market basket. β is negative, its returns move in the opposite direction of the market basket's return. For example, the returns on utility shares often vary countercyclically to the market's performance. If β is positive, then the asset's return follows that of the market. If $|\beta| > 1$, the asset's returns exhibit greater volatility than the market. α measures the asset's unique return regardless of how the market basket performs. It is not uncommon to find that an asset's α is not statistically different from zero.

In this section of the presentation, you will calculate the parameters of the simple version

¹ Unless there is a significant problem, please use PowerPoint to create the presentation.

of the capital asset pricing model (CAPM), calculating alpha and beta. In this section of the presentation, you need to run at least one alpha/beta regression and possibly more. The goal is to demonstrate that you understand how to run the CAPM model and explain the choices you made and what the regression(s) mean. If you decide that the data do not support the classic approach to such calculations, explain why.

There are many choices to be made when doing an alpha/beta regression, such as the data frequency (daily, weekly, monthly), the length of time covered (six months, one year, three years, five years, etc.), and the type of stock market index used (NASDAQ, Wilshire 5000, Standard and Poor's). Explain clearly to the reader which choices you made and why these choices are appropriate.

There are many ways to calculate returns in R. If you have daily share prices stored in the variable "price," then

```
logreturns <- diff(log(price))
```

calculates $\log(p_{t-1}/p_t)$, that is, the log-returns on the share. You can also change the lag in the diff() function. For example,

```
logwkreturns <- diff(log(price), lag = 5)
```

gives you a running weekly return rate. Remember, you lose as many observations as your lag. The lm() won't include missing values in the calculation. You can also use R's [] capability to eliminate the missing values.

Your most easily acquired source of data is Yahoo finance.

<https://finance.yahoo.com/?guccounter=1>

For share prices, regardless of the time interval you use, you'll want to use adjusted closing prices which include dividends. You can also use Yahoo finance for the risk-free rate. Typically, the risk-free rate to use is the 90-day U.S. Treasury bill.

Part 3 - Stock Price:

The next section of your presentation analyzes the stock price, not the return. Stock price regressions sometimes use the overall stock market as an explanatory variable but, in general, have several company-specific variables as predictors. You should gather the predictors and explain why you chose them. Then graph the key variables, create descriptive statistics, develop your model, which may require a few regressions, and make sure you explain the graphs, statistics and regressions.

You will make predictions about the company's stock price. You should acquire fifteen years of quarterly data. It's best to use end-of-quarter prices. You should estimate your model over the first fourteen years, then forecast the fifteenth year, then compare your forecast to actuals. Provide the 95% confidence interval for this predicted stock price. (that is, 95% of the time, we expect the stock price on Feb. 1st of next year to be between what two

values.)

You can use the various techniques from class for this part of the project. I am also willing to accept new approaches if they are in your statistics toolset, and you can explain their rationale.

STOCK PRICE = function (Market Variables, Company Specific Variables)

Company Specific Variables:

Typically, every company has a key driver. For example, at Northeastern University, student enrollment is the key driver behind departments' success or failure. Departments that have growing enrollments get resources; departments with falling enrollments wither and die. Your job is to determine the key driver for your company, graph the critical driver, create descriptive statistics, and ensure the necessary driver is in the stock price regressions. Key driver examples include:

Critical Driver	Company
Number of employees	McDonald's
Oil Reserves	Exxon, Shell
Number cars sold	Toyota, Ford
Number of computers sold	Dell, HP

The data may include market trends (such as the Dow Jones Industrial Average or S&P 500 for USA companies) and relevant company-specific data series such as the company's stock price, earnings, earnings per share, sales, number of employees, number of patents, R&D expenditures, and other factors.

You can find the data in at least three places. First, many companies provide extensive data on their web site. Look for pages titled "Investor Relations." Second, all public US companies must file data with the Securities and Exchange Commission. Go to www.sec.gov and use the Edgar database to look at 10-k and 10-q reports. You must use BOTH to get a full four quarters of data. In the USA, relatively few publicly traded companies issue all four quarters of data. Instead, they issue three-quarters of data and, in the fourth quarter, release the annual totals. To see the fourth-quarter data, you must subtract the three prior quarters from the yearly figures. Third, many accounting courses teach you how to use the Mergent database. Mergent is an excellent source of data for this project if you know how to use the program. An excellent source for macroeconomic data is the St. Louis Federal Reserve Bank's FRED database

<https://fred.stlouisfed.org/>

Part 4 - Conclusions:

This last part of the project summarizes what you have found and recommendations for what the reader should do. For example, if you are creating a stock market report, do you recommend buying the stock today?

Merely a sell or buy recommendation is unacceptable. Explain why you are making a particular choice.

In your project submission, besides the presentation, please include the R code and the data set you've used.

Other Thoughts:

You can use an appendix to report the details of the statistical results without crowding the presentation.

If your formatted coefficients are all zero, then either show more decimal places or adjust the company-specific data. For example, modifying company data from thousands to millions causes the regression coefficients to be shifted by three decimal places and becomes far easier to interpret.

Make sure that in your project, it is clear you have explicated/performed all of the following for each of the regressions:

- R-squared, Adjusted R-squared

- The standard error of the regression ANOVA table/F-statistic/Significance F Degrees of freedom/Residuals

Explain in simple language what each predictor coefficient means. (A most important step!)

Explain the statistical significance of each predictor coefficient

Explain if there are any outliers and, if there are, examine the effect of eliminating them.

You can try both the LASSO and standard regression approaches and see if it improves the regression results or ameliorates any problems. See the [Introduction to Statistical Learning](#) book for a discussion of how to a LASSO regression.