Final (Individual) Project Guidelines

Analysis of Corporate Dividends in Australia and the UK

Project Due Date: Sunday, 24 October, 2021, by 11:59 pm.

Report due via Turnitin.

Jupyter Notebook should be uploaded to the appropriate folder on Moodle.

For the final project you will conduct analysis of the dividends of companies from the following two countries: Australia and the UK.

Relevant financial data is available on Moodle in one excel file: dividend_au_uk.xlsx. This file is in the "Final Project Documents" folder on Moodle. You will have to download the data, upload to your Jupyter notebook and then conduct your analysis using Pandas and other Python Libraries. You will have to submit the Jupyter notebook you used to do the analysis for this project.

You will have to write up your analysis in a report of up to 2,000 words. Your report should also include tables and graphs from your analysis. These tables and graphs have to be produced using Python and you will have to submit all the relevant codes in a Jupyter notebook.

The objectives of your analysis are as follows:

- Document and discuss the distribution and trends in dividend payout ratio (dividend/net income) and the number and percentage of dividend payers (positive dividend) over time in each country:
    - Use dividend to net income ratio as the measure of dividend payout ratio.
    - Dividend payout ratio attempts to measure what percentage a firm's earnings is paid out in dividends.
    - Dividend payout ratio is not a meaningful measure in the following two cases, so you need to deal with these cases in the data pre-processing step:
    1. When a firm has negative net income, the dividend payout ratio is not a meaningful measure.
        - So exclude observations (rows) with negative net income from your sample.
    2. When a firm has a dividend payout ratio higher than 1 (dividend is higher than net income), the dividend payout ratio is not reliable as a firm cannot pay out more dividend than net income in the long run.
        - So cap the value of dividend payout ratio at 1.0 --- set any value higher than 1.0 to 1.0.

    - You will conduct the analysis for Australia and the UK and you will discuss how the dividend payout ratios of the two countries compare with each other and if they show similar or different trends over time.
    - You should perform similar analysis of dividend payers in the two countries. In two separate graphs, you should show the number and percentage of dividend paying firms in the two counties and how these have changed over time.

o   You will document the distribution of dividend payout ratio in each country in 2007 and 2017 to see if the distribution has changed over time. You can use histograms, boxplots, kernel density plots and percentile plots to show the distributions.

Correlation and Regression Analysis

- Analyse the determinants of dividend payout ratio in each country. So you will have two sets of results.
    - o   Initially, explore the relations between various firm characteristics (such as firm size, profitability, growth opportunity etc.) and dividend payout ratio using scatter plot.
    - o   You will then conduct correlation analysis to determine if there are significant correlations between these characteristics and dividend payout ratio.
    - o   Then use simple linear regressions to quantify the relation between dividend payout ratio and these characteristics one at a time. Here you will use regressions with one independent variable (see lecture 7).
    - o   Finally you will use multiple linear regression analysis to consider the effects of all the different firm characteristics on dividend payout ratio.
    - o   You will compare and contrast the results you get from the above analysis for the two countries in your sample: Australia and the UK.

Machine Learning Analysis

- Finally, you should estimate two Machine Learning models and evaluate the predictive performance of these models.
    - o   The first model will try to predict the dividend payout ratio of a firm. You can use the Boston House Price example as a template for this analysis and do similar analysis on dividend payout ratio (instead of house price).
        - ▪   As X (or independent) variables (features matrix), use the four firm characteristics we used in the group project: Firm size (Log of SALES_USD), Profitability, Tangibility and Market to book ratio.
        - ▪   The y variable or dependent variable (the target vector) in your model would be the dividend payout ratio.
        - ▪   You should to the train-test split and evaluate the model's performance on the test dataset and interpret the results.
    - o   The second model will try to predict whether a firm pays dividends --- that is, whether the dividend of a firm is positive.
        - ▪   Create a variable in your dataframe called PAYER which should be 1 if a firm has positive dividend (and therefore positive dividend payout ratio) and 0 otherwise. This variable will be the categorical dependent variable in your supervised classification model.
        - ▪   Same as in the first model, as X (or independent) variables, use the four firm characteristics we used in the group project: Firm size (Log of SALES_USD), Profitability, Tangibility and Market to book ratio.
        - ▪   Use the K Nearest Neighbor model or KNN model for this analysis
        - ▪   You can use Iris flower example (covered in lecture 8) as a template for this analysis and do similar analysis on dividend PAYER (instead of Iris flower types).

- You should do the train-test split and evaluate the model's performance on the test dataset and interpret the results.

I have posted two papers on Moodle for you to read for this assignment. These papers analyse dividend payout ratio, but they do not use the same firm characteristics as independent variables that your dataset has --- but these papers will help you understand the general research background and how to interpret the results. As independent variables in your analysis, you should use the same firm characteristics that we used in the leverage analysis (such as firm size, Profitability etc.).

You should also do additional research via google on the determinants of dividend payout ratio and use those sources as references in your report.

You will summarise you main finding is a report of 2000 words. The report will:

1. Summarise the relevant literature (research papers) and research question.

2. Report and discuss descriptive data analysis and data visualisation.

3. Report and discuss correlation and regression analysis

4. Report and discuss Machine Learning (ML) analysis of dividend payout ratio using the linear regression model (LinearRegression: covered in lecture 11 in the Boston House Price example). Fit an ML model to predict dividend payout ratio and evaluate the performance of the model.

5. Report and discuss Machine Learning (ML) analysis of dividend payers (if dividend payout ratio > 0 then dividend payer = 1, otherwise dividend payer = 0) using the K Nearest Neighbor model or KNN model (covered in lecture 8). Fit an ML model to predict whether a firm pays dividend and evaluate the performance of the model.

6. Draw inference and conclusion and relate the findings to existing research.

Marks Distribution:

Marks will be distributed as follows:

Report write-up with graphs and tables (2000 words):   25%

80% of the marks will be based on the Juypyter Notebook and the following sections of the notebook (you should comment the Notebook so it is easy to follow your work):

Data cleaning and pre-processing: 10%

Descriptive analysis and data visualisation: 25%

Correlation and regression analysis   15%

Machine learning analysis

     Linear regression   15%

     K Nearest Neighbor   10%