

BC2406 Analytics I Computer Lab Based Assessment (Submission 2)

AY2017 Semester 1

Duration: 11 Days

Instructions

1. This paper contains 7 questions and 5 printed pages, including this cover page.
2. The second submission is due on **Mon 6 Nov 2017, 11pm.**
3. Answer all parts of the question, in the order listed. **Set your seed based on the random number 2710.**
4. This is an open-book **individual** assessment. All communication, in any way, about the questions and answers contained herein are not allowed during and after the lab assessment. You may only discuss with your course instructor, if necessary. If you use any code (in significant amt) or explanations from a website, book, paper or person, please cite the source and reference accordingly.
5. You are only allowed to use the following software to answer the questions. **Do not use Microsoft Excel or any other software.**
 - R or/and Rstudio (for all analysis)
 - Microsoft Word (for your answers)
6. You are required to type all your answers, and copy & paste (if necessary) screenshots of software outputs/charts into Microsoft Word.
7. For submission, you must name and email the following two files to neumann.chew@ntu.edu.sg.
 - *studentname2.R*
 - *studentname2.docx*Do not zip, pdf or use any cloud or online repository e.g. google drive, one drive etc. Attach the two files directly, in the file format specified, in your email.
8. In email subject, type *GrpX studentname lab2*. Eg: Grp2 sidneysheldon lab2.
Note: Tuesday class is Grp1, Wed class is Grp2, Thu class is Grp3.
9. If you did not receive email acknowledgement of your submission from your instructor, please check with him. Marks will be deducted if you submit after the deadline. Please submit before the deadline.
10. The final marks awarded will be a weighted average of your first submission (in the lab) and second submission.

Dividends and Earnings

Introduction

Public companies pay dividends as a way to provide return to their shareholders. Many researchers and practitioners believe that a company's dividend policy provides information about the firm's earnings prospects. Some believe that dividends are “sticky” – that managers are reluctant to increase dividends unless they believe that dividends payout can be sustained.¹ As an implication of the “information content of dividends” hypothesis, investors often prefer stocks that have stable dividend payout. Over the past three decades, there have been significant changes in corporate payout policy. Some observed that fewer companies are paying regular dividends and stock repurchase has become an increasingly popular alternative payout to shareholders. The main question is: **Does paying dividends still provide information about earnings prospect?**

Dataset Description

The dataset contains one table with the following variables:

Table Name: Dividend	
File Name: dividend.csv	
Description: Public Company Information Date Range: 2007-07 to 2017-07	
Variable	Description
gvkey	Global Company Key, a unique key assigned to each company
datadate	Date of the data recorded.
fyear	Fiscal year of the record.
indfmt	This code indicates whether a company reports in a Financial Services or Industrial format. INDL – Industrial FS -- Financial Services
consol	Consolidation Level. There is only one value: “C”.
popsrc	Population Source. There is only one value: “D”.

¹ Lintner, J. (1956). Distribution of incomes of corporations among dividends, retained earnings, and taxes. American Economic Review, 46, 97–113.

datafmt	Data Format. There is only one value: “STD”.
tic	Ticker Symbol of the company.
conm	Company Name.
curcd	Currency. There are two values: USD and CAD.
at	Assets - Total
dvc	Dividends Common
ib	Income Before Extraordinary Items
idit	Interest and Related Income - Total
prstk	Purchase of Common and Preferred Stock
spi	Special Items
xint	Interest and Related Expense - Total
exchg	Stock Exchange Code of the exchange where the company is listed (11 = New York Stock Exchange; 12 = American Stock Exchange; 14 = NASDAQ-NMS Stock Market, etc...)
costat	Company status (A = Active; I = Inactive)
sic	Standard Industry Classification Code.

Questions:

1. Conduct data exploration. Show and explain your top 3 most interesting data exploration findings.

(9 marks)

2. Clean the data as follows:

- a. Consider only companies listed on the three US stock exchanges. Keep records with Stock Exchange Code = 11, 12, or 14. What is the number of records after executing 2(a)?
- b. To facilitate comparison, we focus only on US denominated companies. Keep records with currency = USD. What is the number of records after executing 2(a) to 2(b)?

- c. It is a common practice to exclude utilities and financial companies, as they often have different financial structures. Financial companies are identified by SIC sector code from 6000 – 6999. Utilities companies are identified by SIC sector code from 4900 – 4999. Drop these records. What is the number of records after executing 2(a) to 2(c)?
- d. There are also several missing values. We will just delete observations that do not report income (ib = NA). What is the number of records after executing 2(a) to 2(d)?
- e. Bearing in mind the main question, what other data cleaning steps should be executed that you consider necessary and obvious? Explain your choice. What is the number of records after executing 2(a) to 2(e)?

(20 marks)

3. Generate the following summary table. What patterns do you see? Is the number of firms reporting losses [based on ib only] increasing or decreasing? How about the number of companies paying dividends, or repurchasing stock?

Fiscal Year	Number of companies	Number of loss making companies	Number of dividend paying companies	Number of companies with stock repurchase
2007				
2008				
2009				
2010				
2011				
2012				
2013				
2014				
2015				
2016				
2017				

(10 marks)

4. Is dividend payout policy affected by earnings?
- a. State your definition of earnings.
 - b. Conduct Linear Regression to answer this question. What are your conclusion and insights?
 - c. Conduct logistic regression to answer this question. What are your conclusion and insights?
 - d. Conduct decision tree (CART) to answer this question. What are your conclusion and insights?
 - e. Segregate the data into three time frames (fiscal years): 2007 to 2010, 2011 to 2013, and 2014 to 2017. Run 4(b), 4(c), 4(d) separately for each time frame. What are your conclusion and additional insights?

(25 marks)

5. One business hypothesis is that companies are reluctant to cut dividends if the losses² are temporary. That is, the relationship between reporting losses and paying dividends might be less strong if the losses are attributable to special items. Explain how you will test this business hypothesis³. Execute your analysis and explain your conclusion.

(20 marks)

6. Does your answer in Q4 address fully the main question is: Does paying dividends still provide information about earnings prospect? Explain. If it does not fully address the main question, explain what else needs to be done.

(8 marks)

7. Instead of selecting the decision tree (CART) with the lowest testset cross-validation error, Breiman et al (1983) recommends selecting, if possible, the simplest tree that is within 1 standard error of the lowest testset cross-validation error i.e. the 1SE rule. Explain the value of such a recommendation. Is there any difference if the 1SE rule is used in 4(c)? Show and explain.

(8 marks)

² losses are based on ib only.

³ We are using hypothesis in a general sense and does not mean statistical hypothesis testing.