

FIN 557 Problem Set 3

Due on 03/26/2023

Instruction: format your answers to includes SAS code, brief descriptions of the code, and output results in a single PDF file. For example, copy the SAS code and descriptions in a word document and add SAS results using Screenshot tools. Organize the answers by question numbers.

1 Downloading the data

Get data/Compustat - Capital IQ/North America/Fundamentals Annual

Step 1: Choose your date range

Date range: 01/2010-12/2014

Step 2: Apply your company codes

SIC: 2085 2834 2911 3674 4512 4812 5331 6020 7372

Step 2: Apply your company codes.

☐ TIC ☐ GVKEY ☐ CUSIP ☒ SIC ☐ NAICS ☐ CIK ☐ GIC Sub-Industry

Select an option for entering company codes



2085 2834 2911 3674 4512 4812 5331 6020 7372

Please enter Company codes separated by a space.

Example: IBM MSFT DELL [[Code Lookup](#)]



Code List Name

Save code list to Saved Codes

In the screening variables section, keep the default options. As a result, the output data will include variables on consolidation level, industry format, data format, population source, currency, and company status.

Step 3: Choose variables types

ACT: Current Assets – Total

AT: Assets – Total

CHE: Cash and Short-Term Investments

COGS: Cost of Goods Sold

CONM: Company Name

DATADATE: Data date (this is the default variable, which is automatically downloaded)

DLC: Debt in Current Liabilities – Total

DLTT: Long-Term Debt – Total

EBIT: Earnings Before Interest and Taxes

FYEAR: Data Year – Fiscal

GP: Gross Profit (Loss)

GVKEY: Global Company Key (this is the default variable, which is automatically downloaded)

INVT: Inventories – Total

LCT: Current Liabilities – Total

NI: Net Income (Losses)

RECT: Receivables – Total

REVT: Revenue – Total

SALE: Sales/Turnover (Net)

SIC: Standard Industry Classification Code

TEQ: Stockholders Equity Total

TIC: Ticker Symbol

Apply conditional statements

Fiscal year <= 2013 AND

Fiscal year >= 2011

The goal is to keep observations with the fiscal year equals 2011, 2012, or 2013. Since the calendar year is different from the fiscal year, we chose a wider range of calendar dates in step 1.

Step 4: Select query output

Output format: SAS Windows_64 (for 64-bit processors) or SAS Windows_32 (for 32-bit processors)

Date format: MMDDYY10.

Submit the query

Save the table and name it as firm (make sure that the name of the data set is in lowercase letters; this data set is also available on Canvas)

Create a new folder in SAS Studio, name it hw3, and upload the firm dataset

Use a LIBNAME statement to specify the path of the library hw3

Write a PROC CONTENTS procedure to list the variable attributes and information about the dataset hw3.firm

2 Creating the sample

The goal is to keep only the top 10 firms in terms of having the highest sales in the fiscal year of 2011 in each of the 9 industries. This process will end up with 88 firms because one industry has fewer than 10 firms.

2.1 Keep observations in the fiscal year of 2011

In a DATA step, use hw3.firm as the input data set and apply a conditional statement to keep observations in the fiscal year of 2011. Use a KEEP statement to select the following variables: FYEAR, SIC, TIC, and SALE. Write the output to a new data set named list1 in the work library.

2.2 Sort the data

Use a PROC SORT procedure to sort the observations in the list1 data set by the ascending order of SIC and then the descending of SALE: by SIC descending SALE.

2.3 Create an accumulating column

In a DATA step, use list1 as the input data set. Define an accumulating column N that creates a running total that counts the number of firms accumulated in each industry. Reset N to 0 each time that SAS reaches the first row within a new SIC group. Note that you need to use a BY statement, if FIRST.SIC then N=0, and a sum statement in this DATA step. Write the output to a new data set named list2 in the work library.

The purpose of this step is to rank firms within each industry by the sales in 2011. For example, N=1 means that the firm has the highest sales in 2011 in an industry.

2.4 Keep only 10 firms in each industry

In a DATA step, use list2 as the input data set. Filter the observations and include only the top 10 firms in each of the 9 industries based on the variable N defined in step 2.3. Add a KEEP statement to select the variable TIC only. Write the output to a new data set named list3 in the work library.

Now we have identified TICKER of the top 10 firms with the highest sales in 2011 from 9 industries. Given that one industry has 8 firms in total, this process will end up with 88 firms.

2.5 Merge tables and include matching rows

Sort and merge hw3.firm and list3 and write the output to a new data set named firm1. Use hw3.firm and list3 as the input data sets. The common column used in the BY statement is TIC. Include only the matching rows from both input tables. Note that you need to use IN=VARIABLE in the MERGE statement and use an IF statement to filter the rows.

This step aims to create a panel data that includes fundamental variables of these 88 firms during the fiscal years of 2011 to 2013.

2.6 Clean the data

Notice that some firms have multiple observations per year, and we need to drop the duplicate rows. For example, the following table lists TIC, FYEAR, and CHE of observations 25 to 30. In this case, we would like to keep the 25th, 27th, and 29th observations and drop the 26th, 28th, and 30th observations.

Obs	TIC	FYEAR	CHE
25	BAC	2011	357289.0000
26	BAC	2011	.
27	BAC	2012	349370.0000
28	BAC	2012	.
29	BAC	2013	333190.0000
30	BAC	2013	.

Use a PROC SORT procedure to sort the observations in the firm1 data set by the ascending order of TIC, the ascending order of FYEAR, and the descending of CHE: by TIC FYEAR descending CHE.

In a DATA step, use firm1 as the input data set. Add a BY statement, which is the same as the BY statement in the above PROC SORT procedure. Keep only the first occurrence for each unique value of FYEAR. Write the output to a new data set named firm2 in the work library.

After cleaning the data, the following table lists TIC, FYEAR, and CHE for observations 25 to 27.

Obs	TIC	FYEAR	CHE
25	BAC	2011	357289.0000
26	BAC	2012	349370.0000
27	BAC	2013	333190.0000

Congratulations, you are done with sampling.

3. Creating new variables

In this section, we will normalize the variables and create new common-size items. Scaling helps compare financial variables across firms.

3.1 Create common-size income statement items

In a DATA step, use firm2 as the input data set. Scale the following accounting variables by the variable SALE (sales) and assign the values to new variables. Use the DIVIDE function to divide the first variable by the second variable. One advantage of using the DIVIDE function is to avoid dividing the numerator by zero. Use a LABEL statement to add labels to the new variables created. Write the output to a new data set named firm3 in the work library.

NEW VARIABLE NAME	FORMULA	LABEL
SALE_P	divide(SALE,SALE)	Sales
COGS_P	divide(COGS,SALE)	Cost of Goods Sold
EBIT_P	divide(EBIT,SALE)	Operating Profit
GP_P	divide(GP,SALE)	Gross Profit
NI_P	divide(NI,SALE)	Net Income

3.2 Create common-size balance sheet items

In a DATA step, use firm3 as the input data set. Scale the following accounting variables by the variable AT (total assets) and assign the values to new variables. Use the DIVIDE function to divide the first variable by the second variable. Use a LABEL statement to add labels to the new variables created. Write the output to a new data set named firm4 in the work library.

NEW VARIABLE NAME	FORMULA	LABEL
CHE_P	divide(CHE,AT)	Cash & Short-Term Investment
RECT_P	divide(RECT,AT)	Net Receivables
INVT_P	divide(INVT,AT)	Inventories
ACT_P	divide(ACT,AT)	Total Current Assets
AT_P	divide(AT,AT)	Total Assets
DLC_P	divide(DLC,AT)	Total Current Debt
LCT_P	divide(LCT,AT)	Total Current Liabilities
DLTT_P	divide(DLTT,AT)	Long Term Debt
TEQ_P	divide(TEQ,AT)	Stockholders Equity Total

3.3 Create financial ratios

In a DATA step, use firm4 as the input data set. Define the following financial ratios on liquidity, asset management, financial leverage, and profitability. Use the DIVIDE function to divide the first variable by the second variable. Use a LABEL statement to add labels to the new variables created. Write the output to a new data set named firm5 in the work library.

NEW VARIABLE NAME	FORMULA	LABEL
CASH_RATIO	divide(CHE,AT)	Current and Marketable Securities to Total Assets
ACID_TEST_RATIO	divide(CHE+RECT,LCT)	Acid Test Ratio
CURRENT_RATIO	divide(ACT,LCT)	Current Ratio
DAY_RECEIVABLE	divide(365*RECT,SALE)	Day's Receivable
DAY_INVENTORY	divide(365*INVT,COGS)	Day's Inventory
ASSET_TURNOVER	divide(SALE,AT)	Asset Turnover
DEBT_ASSETS	divide(DLC+DLTT,AT)	Debt to Total Assets
GROSS_MARGIN	divide(GP,SALE)	Gross Margin Ratio
RETURN_SALE	divide(NI,SALE)	Return on Sales
RETURN_ASSET	divide(NI,AT)	Return on Assets
RETURN_EQUITY	divide(NI,TEQ)	Return on Equity

Congratulations, you are done with the data preparation process.

4. Compare firm characteristics across industries

In this section, we will compare the accounting variables and financial ratios of firms across industries. The analysis is based on the data set firm5.

4.1 Generate summary statistics

Use a PROC MEANS procedure to group the data by SIC with a CLASS statement and generate the mean of DAY_RECEIVABLE, INVT_P, and GROSS_MARGIN for firms in each of the 9 industries. Include the output table of summary statistics generated by the PROC MEANS procedure in the solutions. Based on the industry average during the fiscal years of 2011 to 2013, answer the following questions.

Which industry has the longest day's receivable (DAY_RECEIVABLE), and what's the industry average?

Which industry has the largest Inventories (INVT_P), and what's the industry average?

Which industry has the highest gross margin ratio (GROSS_MARGIN), and what's the industry average?

You can look up SIC codes and find the corresponding industry names. For example, 6020 represents commercial banks, 5331 is for retail grocery companies, and 2834 stands for pharmaceutical manufacturers.