Assignment 2.1: Using SAS and Financial Statement Analysis

Submission Details

- Submit through T-square dropbox

- Use SAS for the programming

- You have to submit ONLY

  - SAS program
  - Output in PDF format
  - You don't need to submit any datasets

Assignment

- Compute the descriptive statistics (see excel spread sheet in Assignments) for the following years: 1971, 1975, 1980, 1985, 1990, 1995, 1998, 2003, 2008 and 2013

- Optional Step: You don't need to read it, but in case you want more back ground, check Frank and Goyal (2003) paper in Assignments (only Tables 1,2, 8, 9 and 10 of the paper). Note that the variable names in the paper are in OLD format. The variable names in the excel spreadsheet and dataset are current. Translation file posted has the variable translations. The paper is useful if you want to check whether your numbers are approximately correct for 1971-1998.

- Descriptive statistics that need to be computed for each of the years listed and for each variable are

  - Number of observations
  - Mean
  - 25th, 50th and 75th percentiles
  - Standard deviation
  - Minimum and Maximum of the variable

- Correlation matrix for the variables as presented in Table 10

Data Extraction

- In order to reduce the time demands for the assignment, I have uploaded the required dataset to tsquare.

- The name of the SAS dataset is FUNDA.SAS7BDAT

SAS commands that may be useful for the Data Analysis

As I mentioned in the class, SAS provides multiple methods to perform any given task. Some of the following commands may be useful (look up the examples from the UCLA SAS web site link that I posted on Tsquare)

- DATA step

- PROC SORT

- PROC FREQ

- PROC UNIVARIATE

- PROC MEANS

- PROC PRINT

- PROC CORR

Data Analysis

Steps in the assignment

1. The data file (funda.sas7bdat) is located on tsquare under resources

2. Download the data to your local computer

3. Use SAS enhanced program editor to write the program. Save the program with extension *.sas*

4. First, set a libname. Say libname comp ' `P:\`' or which ever directory that the data is stored in;

5. Read the data with the following code: *set comp.funda*

6. Read the dataset and subset the data with the *required filters*

7. Be careful with how you write the program. Document the steps and make it readable

8. Check the grading rubric to get a sense of what I am expecting in the assignment

   - There are approximately 1000 variables and a 600,000+ observations. So table size is approximately 600 million items. Data access may take a long time unless you code efficiently

   - I encourage you to write efficient code (remember sequential access in SAS, you don't need to read the data that you don't need)

   - Use macro variables and macros where necessary

   - Write the initial program by reading only 100 observations (use OBS=100 command when you are reading the dataset).

   - Once the program works, you can always remove the 100 obs filter.

9. Use the following filters to subset the data

   - if indfmt='INDL' and datafmt='STD' and popsrc='D' and consol='C'

10. Use PROC MEANS or PROC UNIVARIATE

11. Save the results in PDF format (similar to the output in Tables 1,2,8,9 and 10). I don't want to see hundreds of pages of output.

12. Check the LOG file to see if there are any errors in your code

13. Upload the SAS program and results to tsquare

Notes on Assignment-2.1

1. See details in the excel file for assignment 2

2. For format codes =1, 2 etc., the corresponding variable in the dataset is SCF. The data from COMPUSTAT, for years up to and including 1985, uses format codes 1, 2 and 3; starting in year 1990 format 7 is used. Thus, you shall have something like shown below to filter the dataset:

   - if scf = 4 or scf = 5 or scf = 6 or scf = . then delete;

3. Following stardard practice, financial firms(6000-6999), regulated utilities(4900-4999), and firms involved in major mergers(Compustat footnote code AB) are exclude. Therefore, use the following filters to subset the data (after set comp.funda)

   - if 5999 < sich < 7000, then delete; (One example)

4. for Compustat footnote code AB use variable *COMPST* not equal to AB

5. Optional Steps: not required for the assignment per se.

   - READ the paper. If a particular variable doesn't seem to be defined in the table, the easiest thing is to do a *search* in the PDF document and it would most probably take you to the page in the document where that particular variable is defined.

   - for example, *net assets*: search and it will take you page 222 where net assets is defined as *total assets - current liabilities*

   - when the paper mentions a variable without specifying it further, for example, *current liabilities* it means *current liabilities - total*. Same for other variables that are not specified in more detail. they refer to *total* value of that variable.

6. *logsales* is not a new variable in the database. It is a derived variable and is defined as the natural logarithm of sales.

7. variables in the paper are deflated to constant 1992 dollars. For example, you can use GDP deflator.`http://research.stlouisfed.org/fred2/series/GDPDEF` for more information. Figure out how to do it.

8. See `http://research.stlouisfed.org/fred2/categories/21` for more on deflators

9. UNDERSTAND what the variables means. That is part of the assignment. READ the paper especially the data section and where the variable is defined.

10. I don't expect you to be able to get the numbers right to the fourth decimal. But if you do it correctly your numbers should be of the same order of magnitude and close to the solution

11. Table 1 and 8 of the excel spread sheet / paper are important and useful for your future work.