Statistical Consulting Service

Report for
Mrs Samanmali Dassanayake

7 February 2022

Dear Ms. Dassanayake,

We would be truly grateful if you could follow the below given set of guidelines when compiling the final data set. The below guidelines will aid you to compile your data into a *tidy data* format which is extremely helpful to understand the structure of the dataset and extract the required variables being more time efficient by reducing the excessive time spent on data cleaning.

The standard criteria to follow for the overall data structure is as follows:

- Every column is a variable
- Every row is an observation
- Every cell is a single value

All the guidelines mentioned below will help you to compile the final data set adhering to the above criteria.

1 Structure of data set (worksheets)

The data set can be compiled into two worksheets namely;

1.1. Company specific data and;

1.2. Macro-economic data.

Please refer the format given below when entering the data.

1.1 Company (Firm) specific data

The more convenient format out of the following two formats can be chosen when entering the firm specific data:

1. Entering all the data from 2007-2018 of a single firm at once followed by the other firms.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Year	Firm	TDER	LTDE	STDE	ROE	ROA	TA	SIZE			
2	2018	A	xx	xx	xx	xx	xx	xx	xx			
3	2017	A	xx	xx	xx	xx	xx	xx	xx			
4	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮			
5	2008	A	xx	xx	xx	xx	xx	xx	xx			
6	2007	A	xx	xx	xx	xx	xx	xx	xx			
7	2018	B	xx	xx	xx	xx	xx	xx	xx			
8	2017	B	xx	xx	xx	xx	xx	xx	xx			
9	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮			
10	2008	B	xx	xx	xx	xx	xx	xx	xx			
11	2007	B	xx	xx	xx	xx	xx	xx	xx			
12												
13												
14												
15												
16												

2. Entering all the data of all the firms for a particular year at once followed by the other years.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Year	Firm	TDER	LTDE	STDE	ROE	ROA	TA	SIZE			
2	2018	A	xx	xx	xx	xx	xx	xx	xx			
3	2018	B	xx	xx	xx	xx	xx	xx	xx			
4	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮			
5	2018	Y	xx	xx	xx	xx	xx	xx	xx			
6	2018	Z	xx	xx	xx	xx	xx	xx	xx			
7	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮			
8	2007	A	xx	xx	xx	xx	xx	xx	xx			
9	2007	B	xx	xx	xx	xx	xx	xx	xx			
10	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮			
11	2007	Y	xx	xx	xx	xx	xx	xx	xx			
12	2007	Z	xx	xx	xx	xx	xx	xx	xx			
13												
14												
15												
16												

1.2. Macro-economic data

	A	B	C	D	E	F	G	H	I	J	K	L
1	Year	GDP	INF	MKTCAP								
2	2018	xx	xx	xx								
3	2017	xx	xx	xx								
4	⋮	⋮	⋮	⋮								
5	2008	xx	xx	xx								
6	2007	xx	xx	xx								
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												

2 Additional Guidelines

Additionally, please kindly follow the below specifications when entering the data.

- **Avoid using colors:** Do not color the fonts or the cells which can make the data set visually noisy
- **Be uniform:** Use the same format and units for same type of data *e.g: If capital letters are used for the name of a firm, use capital letters for all the other names of firms, if a certain observation of a variable is in percentage format make sure all values of the variable follows that format.*
- **Use abbreviations:** To avoid the variable names being too long, use meaningful abbreviations which can be then further described through the data dictionary. *Please refer section on data dictionary for more information.*

3 Data Dictionary

A data dictionary is a descriptions of variables and their definitions and properties in a given data set. They are very useful to give an overall idea about the data set to a data analyst and hence enabling much more meaningful interpretations of results.

The following tabular format can be followed when creating a simple data dictionary.

Abbreviation	Variable	Description	Notes	
TDER	Total Debt to Equity Ratio		Xxx	Xxx
STDR	Short Term Debt to Equity		Xxx	Xxx

Note: The above abbreviations are used as basic examples of creating the data dictionary. Please do feel free to use any such meaningful abbreviation

- **Abbreviation:** The variable name of a given column
- **Variable:** The expanded name of the variable (column)
- **Description:** A brief description about the variable, ensure to mention the specific formula used if many forms of formulas can be used to obtain the same measure
- **Notes:** Any additional information with regards to the observations of the variable that might be important when analyzing the data

4 Summary

According to the above guidelines and specification, the final three worksheets to be shared are as follows;

1. Company specific data
2. Macro-economic data
3. Data dictionary

Hope the above documentation is easily comprehensible and if there is need for any further clarification, please do reach out to us.

Thank you.

Kind regards,

Statistical Consulting Service – USJ