University of Waterloo

CFM 301

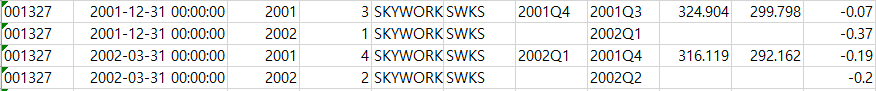
Data Assignment 2

Jeongseop Yi (j22yi)

**Q1.** The downloaded data is q1.sas7dbat file in the assignment package. It can be also found in q1 sheet of DA2.xlsx file.

**Q2.** The downloaded data is q2.sas7dbat file in the assignment package. It can be also found in q2 sheet of DA2.xlsx file.

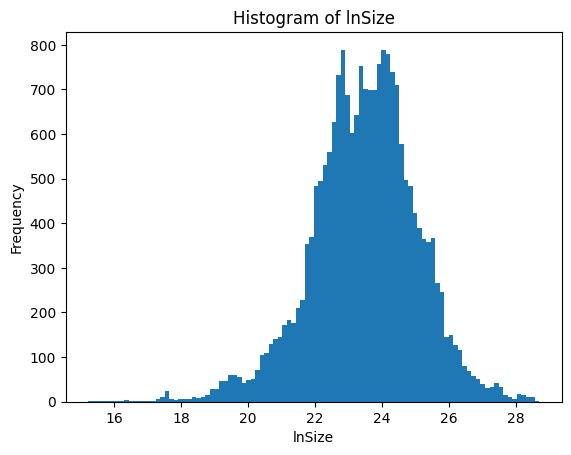
**Q3.** The merged dataset is in the q3 sheet of the DA2.xlsx. There are 6 duplicated entries in terms of MONTHID added to the merged dataset due to the duplicated DATADATE from Q2 for SKYWORKS SOLUTIONS INC. As there exist some values exist for the duplicated entries, the entries were not dropped. Therefore, there exist 20848 entries exist in the merged dataset instead of 20842 entries.

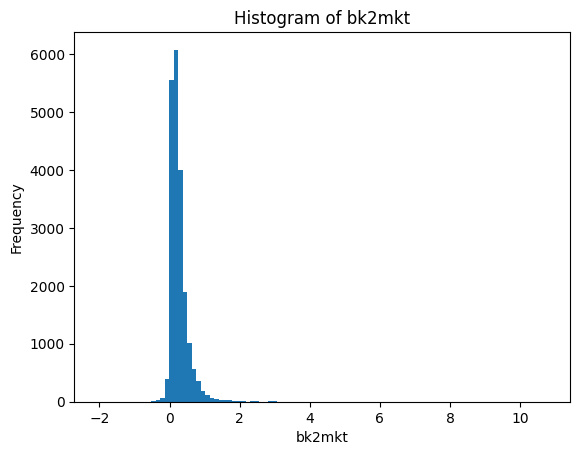


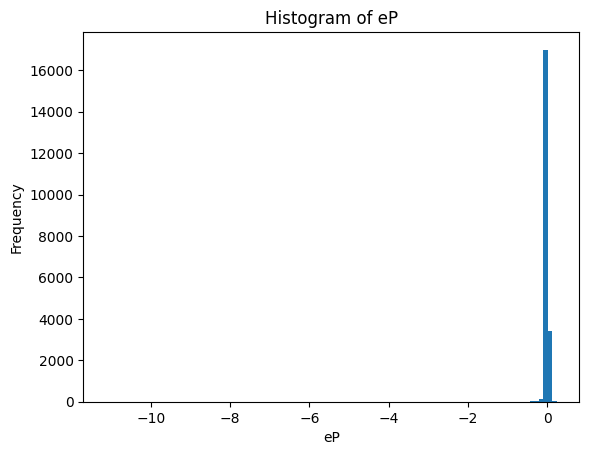
**Q4.** The data can be found in the q4 sheet of the DA2.xlsx.

c) i). I decided to use IBQ / Market equity for the definition of eP. From the data dictionary from WRDS, EPSPXQ is calculated with the number of shares including the conversion of common/ordinary stock equivalents. The number of the shares including the conversion of common/ordinary stock equivalents is usually bigger than the value of SHROUT, which disregards the conversion. The current stock price is determined by SHROUT, which is the number of publicly held shares and does not include the number of the conversions which are not stocks currently. As the PRC and EPSPXQ are determined from the different number of stocks, it may not be correct to use the value from EPSPXQ method over IBQ method. The eP value from EPSPXQ would be diluted from some stocks that do not exist. Therefore, I will be using IBQ / Market equity method for the definition of eP.

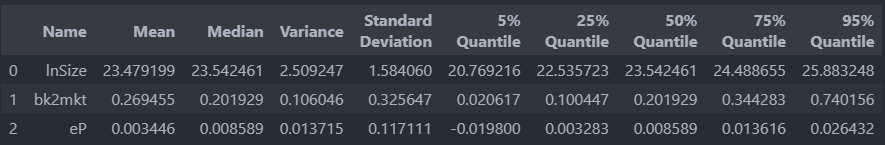
Histograms of each variable are as follows:



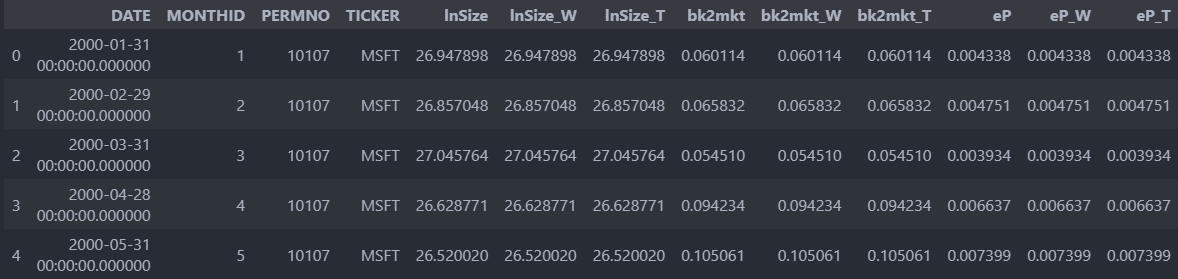




**Q5**. The summary statistic data are as follows:



**Q6.** The data can be found in the q6 sheet of the DA2.xlsx file. A sneak peek of the data and the summary statistics are as follows.



Number of total outliers of lnSize: 165

Number of total outliers of bk2mkt: 688

Number of total outliers of eP: 723

Statistics for winsorized data for lnSize:

Mean: 23.4789

Median: 23.54246

Standard deviation: 1.57831

Min: 16.1331

Max: 28.20077

1% percentile: 19.20504

99% percentile: 27.17433

Statistics for truncated data for lnSize:

Mean: 23.47274

Median: 23.53915

Standard deviation: 1.54521

Min: 16.42112

Max: 28.15999

1% percentile: 19.24006

99% percentile: 26.90796

Statistics for winsorized data for bk2mkt:

Mean: 0.26046

Median: 0.20193

Standard deviation: 0.25339

Min: -1.208

Max: 4.43973

1% percentile: -0.09317

99% percentile: 1.09799

Statistics for truncated data for bk2mkt:

Mean: 0.24481

Median: 0.19794

Standard deviation: 0.21315

Min: -0.56474

Max: 2.97543

1% percentile: -0.09141

99% percentile: 0.94145

Statistics for winsorized data for eP:

Mean: 0.00555

Median: 0.00859

Standard deviation: 0.04957

Min: -4.08768

Max: 0.15519

1% percentile: -0.07256

99% percentile: 0.05202

Statistics for truncated data for eP:

Mean: 0.00726

Median: 0.00864

Standard deviation: 0.01676

Min: -0.46053

Max: 0.15519

1% percentile: -0.04775

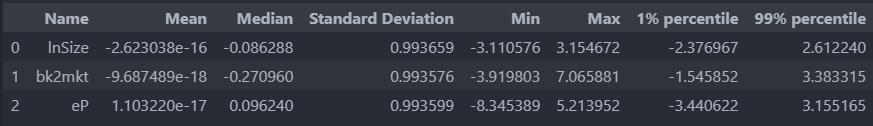
99% percentile: 0.04253

**Q7.** The data can be found in the q7 sheet of the DA2.xlsx file. A sneak peek of the data and the summary statistics are as follows. As the question states that z-score is defined at each point of time, monthly calculations of z-scores are done similarly as in Q6. The full monthly statistics can be found in the q7\_summary\_month sheet of the DA2.xlsx file.

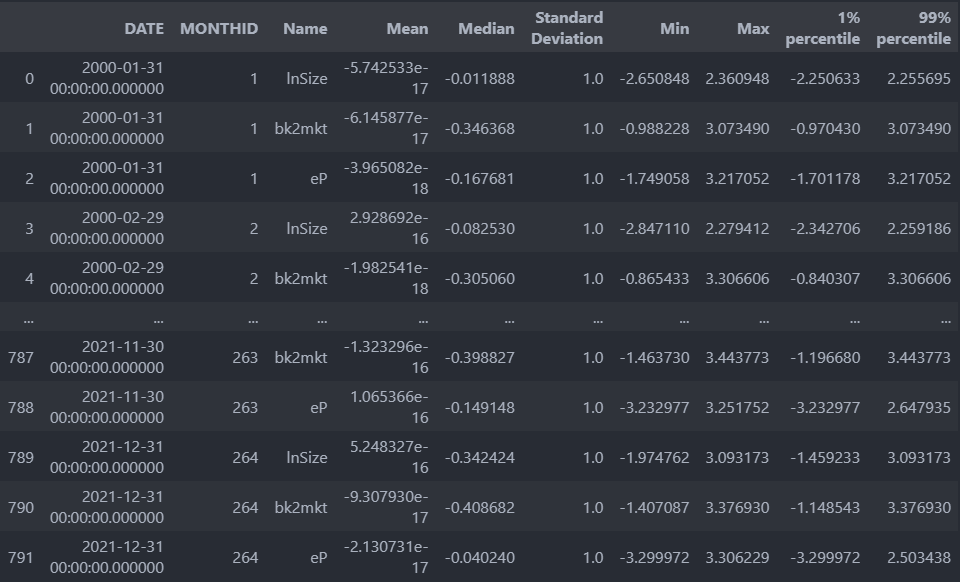
Sneak peek of q7 data:



Total statistics for z-scores:

****

Monthly statistics for z-scores:

****