University of Waterloo

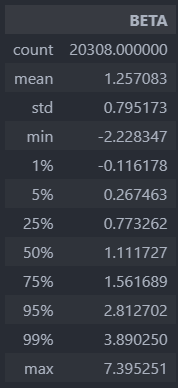
CFM 301

Data Assignment 3

Jeongseop Yi (j22yi)

**Q1.** The data for Q1 is in Q1 sheet in the DA3.xlsx file in the assignment package.

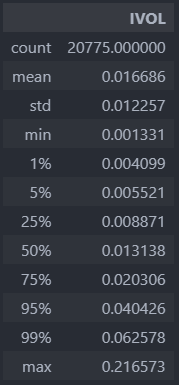
The summary statistic of BETA is as follows:

****

**Q2.** The data for Q2 is in Q2 sheet in the DA3.xlsx file in the assignment package.

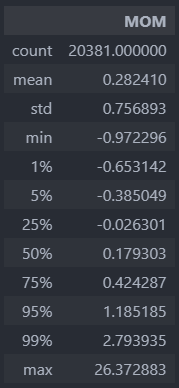
I used the built-in std() function of pandas DataFrame. I found out that the function uses sample standard deviation method which uses n-1 for the divisor. The IVOL values are a bit higher than the values calculated using population standard deviation method which uses n for the divisor.

The summary statistic of IVOL is as follows:



**Q3.** The data for Q3 is in Q3 sheet in the DA3.xlsx file in the assignment package.

The summary statistic of MOM is as follows:

****

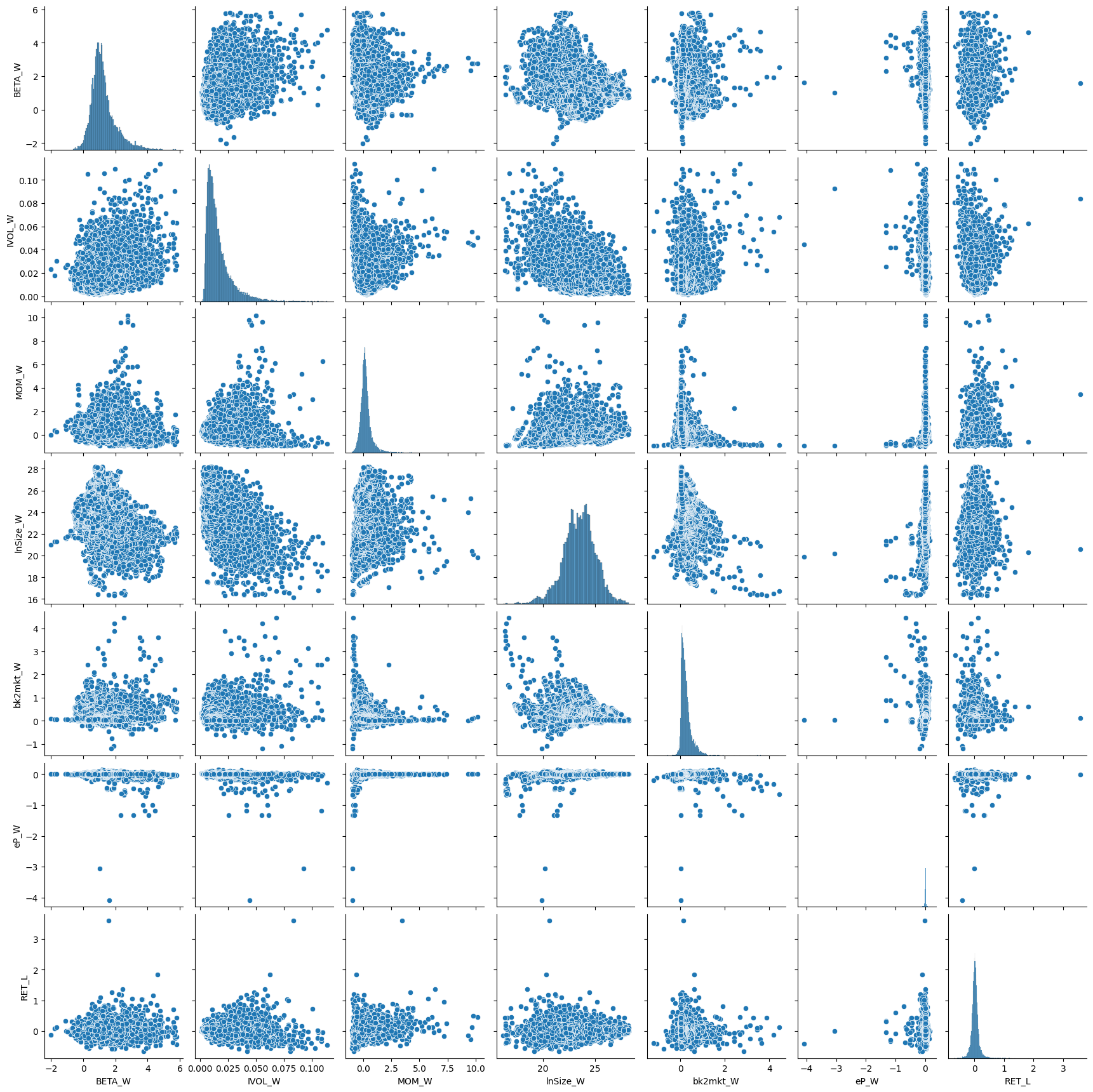
**Q4.** The data for Q4 is in Q4 sheet in the DA3.xlsx file in the assignment package.

The summary statistic of winsorized BETA, IVOL, and MOM are as follows:



**Q5.** The data for Q5 is in Q5 sheet in the DA3.xlsx file in the assignment package. In my solution of DA2, I included the 6 duplicate data entries from SKYWORKS SOLUTIONS INC. Therefore, there are 6 duplicate values for BETA\_W, IVOL\_W, MOM\_W, which may add some errors to my results.

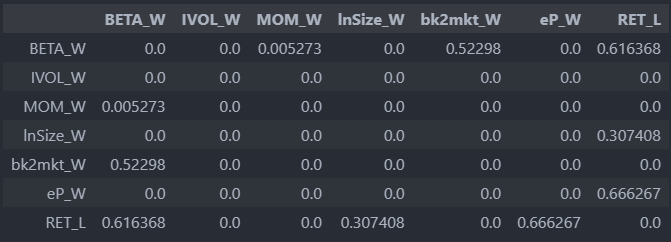
1. The scatterplot between the explained variables and each of the explanatory variables is as follows:



1. The correlation matrix between the explained variables and each of the explanatory variables is as follows:



The p-value matrix is as follows:



Based on the p-values, the null hypothesis of no correlation between the variables must be rejected for all pairs of variables except for between BETA\_W and bk2mkt\_W, BETA\_W and RET\_L, InSize\_W and RET\_L, and eP\_W and RET\_L, where the two variables have higher p-value than 0.05. The result follows in the correlation coefficient matrix, where the values are greater than 0 except for the pairs of variables which have p-value greater than 0.05.