In [10]:
```
import numpy as np
import pandas as pd
import datetime
```

In [11]:
```
Sample_data=pd.read_csv("D:/Dropbox/Lecture notes/Estimation of Asset Pricing Models/2021 Fall/
Students folder/SAS codes and assignment data/Assignment 2/assignment2_sample_data.csv")
```

In [12]:
```
Sample_data
```

Out[12]:

	permno	date	year	exchcd	siccd	retadj	eretadj	altprc_lag1	ME_lag1	b_mkt	...	size_20
0	10001	20121231	2012	2	4925	-0.015231	-0.015331	9.520000	77.654644	0.136698	...	6.255015
1	10002	20121231	2012	3	6020	0.010946	0.010846	2.749900	49.404705	1.251151	...	6.255015
2	10025	20121231	2012	3	3081	-0.020992	-0.021092	60.500000	334.625500	0.732402	...	6.255015
3	10026	20121231	2012	3	2052	0.018173	0.018073	62.900002	1179.312129	0.959478	...	6.255015
4	10032	20121231	2012	3	3670	0.114471	0.114371	23.150000	812.125137	1.747265	...	6.255015
...	...	...	...	...	...	...	...	...	...	...	...	...
3192	93428	20121231	2012	3	9999	0.148593	0.148493	31.629999	880.421027	1.520446	...	6.255015
3193	93429	20121231	2012	3	9999	0.007672	0.007572	29.980000	2616.414520	0.491209	...	6.255015
3194	93433	20121231	2012	3	9999	-0.517704	-0.517804	0.850100	39.250816	1.810744	...	6.255015
3195	93434	20121231	2012	3	9999	0.037634	0.037534	7.440000	58.575120	0.393096	...	6.255015
3196	93436	20121231	2012	3	9999	0.001478	0.001378	33.820000	3848.005745	1.482579	...	6.255015

3197 rows × 24 columns

In [13]:
```
print('\n Name : Sample_data \n')
print(Sample_data.info())
```

```
 Name : Sample_data

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3197 entries, 0 to 3196
Data columns (total 24 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   permno       3197 non-null   int64
 1   date         3197 non-null   int64
 2   year         3197 non-null   int64
 3   exchcd       3197 non-null   int64
 4   siccd        3197 non-null   int64
 5   retadj       3197 non-null   float64
 6   eretadj      3197 non-null   float64
 7   altprc_lag1  3197 non-null   float64
 8   ME_lag1      3197 non-null   float64
 9   b_mkt        3197 non-null   float64
 10  size         3197 non-null   float64
 11  size_CPI     3197 non-null   float64
 12  BM           3197 non-null   float64
 13  log_BM       3197 non-null   float64
 14  size_20      3197 non-null   float64
 15  size_40      3197 non-null   float64
 16  size_60      3197 non-null   float64
 17  size_80      3197 non-null   float64
 18  p1           3197 non-null   int64
 19  BM_20        3197 non-null   float64
 20  BM_40        3197 non-null   float64
 21  BM_60        3197 non-null   float64
 22  BM_80        3197 non-null   float64
 23  p2           3197 non-null   int64
dtypes: float64(17), int64(7)
memory usage: 599.6 KB
None
```

In [14]:
```
nstocks = Sample_data.sort_values(['date','p1','p2']).groupby(['date','p1','p2'])['permno'].cou
nt()
```

In [15]:
```
nstocks
```

Out[15]:
```
date      p1  p2
20121231  1   1     336
              2     336
              3     336
              4     336
              5     336
          2   1     111
              2     112
              3     112
              4     112
              5     112
          3   1      68
              2      69
              3      68
              4      69
              5      69
          4   1      63
              2      64
              3      64
              4      64
              5      64
          5   1      59
              2      59
              3      59
              4      59
              5      60
Name: permno, dtype: int64
```

In [ ]: