

A1.

```
In [62]: import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
```

```
In [18]: data = pd.read_csv("EE627A_HW1_Data.csv")
df = pd.DataFrame(data, columns = ['Mkt-RF', 'SMB', 'HML', 'Mom'])
df.corr()
```

```
Out[18]:
```

	Mkt-RF	SMB	HML	Mom
Mkt-RF	1.000000	0.326863	0.216145	-0.338343
SMB	0.326863	1.000000	0.094113	-0.164023
HML	0.216145	0.094113	1.000000	-0.400635
Mom	-0.338343	-0.164023	-0.400635	1.000000

```
In [22]: df1 = pd.DataFrame(data)
corr = df1.corr()
pd.DataFrame(corr, columns=['Mkt-RF', 'SMB', 'HML', 'Mom'])
```

```
Out[22]:
```

	Mkt-RF	SMB	HML	Mom
Date	-0.015422	-0.016266	0.011771	0.019624
Mkt-RF	1.000000	0.326863	0.216145	-0.338343
SMB	0.326863	1.000000	0.094113	-0.164023
HML	0.216145	0.094113	1.000000	-0.400635
RF	-0.068723	-0.059640	0.012115	0.039130
Mom	-0.338343	-0.164023	-0.400635	1.000000
Food	0.835924	0.201696	0.215132	-0.280269
Beer	0.707673	0.351039	0.214962	-0.200077
Smoke	0.584268	0.103154	0.171809	-0.219165
Games	0.830211	0.412069	0.250387	-0.356992
Books	0.830092	0.408145	0.250608	-0.306399
Hahld	0.816234	0.261863	0.107373	-0.199683
Citls	0.780630	0.460134	0.246719	-0.360482
Hlth	0.804022	0.208896	0.064393	-0.225778
Chems	0.883889	0.208012	0.192392	-0.301203
Txtls	0.823438	0.495356	0.347476	-0.371137
Cnstr	0.913608	0.401464	0.247212	-0.323221
Steel	0.867909	0.363120	0.343723	-0.394459
FabPr	0.925046	0.406788	0.260264	-0.372352
ElcEq	0.904103	0.257442	0.202407	-0.324623
Autos	0.843297	0.285971	0.302626	-0.389308
Carry	0.832370	0.362621	0.337407	-0.343814
Mines	0.706160	0.343889	0.245741	-0.271214
Coal	0.490542	0.267311	0.172241	-0.137068
Oil	0.772597	0.137730	0.302340	-0.249683
Util	0.761767	0.175149	0.354729	-0.331665
Telcm	0.755235	0.160554	0.099079	-0.293874
Serve	0.495606	0.265060	-0.084935	-0.092466
BusEq	0.842644	0.329809	-0.066278	-0.226394
Paper	0.869099	0.275653	0.197336	-0.300174
Trans	0.858688	0.348099	0.430456	-0.423153
Whlsl	0.796012	0.459626	0.209154	-0.260379
Rtail	0.858747	0.302931	0.114997	-0.309368
Meals	0.753859	0.368637	0.163557	-0.250613
Fin	0.915884	0.282566	0.324566	-0.420421
Other	0.842008	0.423908	0.165099	-0.290321

From the above correlation matrix, I infer that market risk free factor correlates highly with every industry. Momentum correlates most negatively with every industry. While RF (risk free rate) correlates negatively for most of the industries

```
In [89]: lagged_data_mkt_RF = pd.Series(sm.tsa.acf(data['Mkt-RF'],nlags = 10))
lagged_data_SMB = pd.Series(sm.tsa.acf(data['SMB'],nlags = 10))
lagged_data_HML = pd.Series(sm.tsa.acf(data['HML'],nlags = 10))
lagged_data_momentum = pd.Series(sm.tsa.acf(data['Mom'],nlags = 10))

print("Mkt-RF", '\n', lagged_data_mkt_RF[1:], '\n')
print("SMB", '\n', lagged_data_SMB[1:], '\n')
print("HML", '\n', lagged_data_HML[1:], '\n')
print("Momentum", '\n', lagged_data_momentum[1:], '\n')
```

```
Mkt-RF
1      0.107165
2     -0.016334
3     -0.108150
4      0.005641
5      0.070126
6     -0.020113
7      0.012570
8      0.036685
9      0.081705
10     0.018962
dtype: float64
```

```
SMB
1      0.075347
2      0.059214
3     -0.054104
4     -0.031584
5     -0.053806
6      0.009881
7      0.022554
8      0.026380
9      0.083590
10     0.024877
dtype: float64
```

```
HML
1      0.178028
2     -0.013279
3     -0.031619
4     -0.080457
5     -0.061377
6      0.007784
7      0.064510
8     -0.002250
9      0.114856
10     0.026137
dtype: float64
```

```
Momentum
1      0.057801
2     -0.077419
3     -0.074536
4     -0.049174
5     -0.038990
6      0.051111
7     -0.036235
8     -0.015936
9      0.012242
10     -0.044102
dtype: float64
```

Referring the above Autocorrelation vs lag for each factor. I conclude Momentum factor to be AR (1) model. Also, I observed that higher the correlation, higher will be the order of ACF. That is, we can consider more observations from the past to predict present of the highly correlated features.

```
In [77]: #Visualization
flag = True
if flag:
    plt.plot(lagged_data_mkt_RF, label = 'Mkt_RF')
    plt.plot(lagged_data_SMB, label = 'SMB')
    plt.plot(lagged_data_HML, label = 'HML')
    plt.plot(lagged_data_momentum, label = 'Momentum')
plt.ylabel('values')
plt.xlabel('lags')
plt.legend(loc = 'upper right')
```

Out[77]: <matplotlib.legend.Legend at 0x1e92625e9e8>

