Part I

```
import pandas as pd
file = 'EE627A_HW1_Data.csv'
data = pd.read_csv(file)
data.head()
```

```
     Date  Mkt-RF   SMB   HML    RF   Mom  Food   Beer  Smoke
Games  ...  \
0  192701   -0.10 -0.09  4.72  0.25  0.36 -0.70   0.57  -0.33
2.46  ...
1  192702    4.32  0.31  3.40  0.26 -1.67  4.29  12.83   1.58
1.43  ...
2  192703    0.33 -1.77 -2.42  0.30  2.97  1.98 -13.56   5.55
0.57  ...
3  192704    0.42  0.30  1.03  0.25  4.53  2.60   2.85   4.09  -
3.34  ...
4  192705    5.36  0.67  3.41  0.30  3.41  6.14  11.62  11.87  -
0.50  ...

   Telcm  Servs  BusEq  Paper  Trans  Whlsl  Rtail  Meals    Fin
Other
0   1.88   2.08  -1.45  -2.60   1.44 -17.93  -3.34   1.53  -2.48  -
4.13
1   3.97   8.90   4.85   5.21   5.20   3.49   4.48   6.81   2.77
0.30
2   5.56  -7.80   4.30  -8.39   1.06 -20.47   3.05  -2.44   1.41
2.28
3  -2.08   3.44   3.10   4.43   0.77 -10.75   2.09   6.02   3.76
4.71
4   3.35  18.33   5.10   5.66   6.69  -4.01   0.49   4.69  10.25
1.40

[5 rows x 36 columns]
```

```
import numpy as np

data_without_date = data.drop(columns=['Date'])

# correlation matrix
correlation_matrix = data_without_date.corr()

# correlations of the four factors with the industries
four_factors = ['Mkt-RF', 'SMB', 'HML', 'Mom']
industry_columns = [col for col in data_without_date.columns if col
not in four_factors and col != 'RF']

highest_correlation = correlation_matrix.loc[industry_columns,
four_factors].idxmax(axis=1)
lowest_correlation = correlation_matrix.loc[industry_columns,
```

```
four_factors].idxmin(axis=1)

rf_correlation_with_industries = correlation_matrix.loc['RF',
industry_columns]

highest_correlation, lowest_correlation,
rf_correlation_with_industries.abs().mean()
```

```
(Food      Mkt-RF
 Beer      Mkt-RF
 Smoke     Mkt-RF
 Games     Mkt-RF
 Books     Mkt-RF
 Hshld     Mkt-RF
 Clths     Mkt-RF
 Hlth      Mkt-RF
 Chems     Mkt-RF
 Txtls     Mkt-RF
 Cnstr     Mkt-RF
 Steel     Mkt-RF
 FabPr     Mkt-RF
 ElcEq     Mkt-RF
 Autos     Mkt-RF
 Carry     Mkt-RF
 Mines     Mkt-RF
 Coal      Mkt-RF
 Oil       Mkt-RF
 Util      Mkt-RF
 Telcm     Mkt-RF
 Servs     Mkt-RF
 BusEq     Mkt-RF
 Paper     Mkt-RF
 Trans     Mkt-RF
 Whlsl     Mkt-RF
 Rtail     Mkt-RF
 Meals     Mkt-RF
 Fin       Mkt-RF
 Other     Mkt-RF
 dtype: object,
 Food      Mom
 Beer      Mom
 Smoke     Mom
 Games     Mom
 Books     Mom
 Hshld     Mom
 Clths     Mom
 Hlth      Mom
 Chems     Mom
 Txtls     Mom
 Cnstr     Mom
```

```
Steel     Mom
FabPr     Mom
ElcEq     Mom
Autos     Mom
Carry     Mom
Mines     Mom
Coal      Mom
Oil       Mom
Util      Mom
Telcm     Mom
Servs     Mom
BusEq     Mom
Paper     Mom
Trans     Mom
Whlsl     Mom
Rtail     Mom
Meals     Mom
Fin       Mom
Other     Mom
dtype: object,
0.029056167675253056)
```

Highest Correlation with Industries:

- The 'Market minus Risk-Free' (Mkt-RF) factor correlates most highly with every industry. This indicates that the market factor (after adjusting for the risk-free rate) has a strong influence on the returns of all the industries in the dataset.

Lowest (or Negative) Correlation with Industries:

- The 'Momentum' factor has the lowest correlation with every industry. This suggests that the momentum factor, which reflects the tendency of assets to continue moving in their recent direction, does not have a strong positive correlation with the returns of these industries.

Correlation of Risk-Free Rate with Industries:

- The Risk-Free Rate (RF) does not correlate highly with the 30 industry time series. The average absolute correlation of the Risk-Free Rate with the industries is approximately 0.029, which is quite low.

```
from statsmodels.tsa.stattools import acf

# Auto-Correlation Function (ACF)
lags = 10  # Number of lags

acf_results = {}
for factor in four_factors:
    acf_results[factor] = acf(data[factor], nlags=lags, fft=True)

acf_df = pd.DataFrame(acf_results, index=[f'Lag {i}' for i in
```

```
range(lags + 1)])

acf_df
```

```
          Mkt-RF       SMB       HML       Mom
Lag 0   1.000000  1.000000  1.000000  1.000000
Lag 1   0.107165  0.075347  0.178028  0.057801
Lag 2  -0.016334  0.059214 -0.013279 -0.077419
Lag 3  -0.108150 -0.054104 -0.031619 -0.074536
Lag 4   0.005641 -0.031584 -0.080457 -0.049174
Lag 5   0.070126 -0.053806 -0.061377 -0.038990
Lag 6  -0.020113  0.009881  0.007784  0.051111
Lag 7   0.012570  0.022554  0.064510 -0.036235
Lag 8   0.036685  0.026380 -0.002250 -0.015936
Lag 9   0.081705  0.083590  0.114856  0.012242
Lag 10  0.018962  0.024877  0.026137 -0.044102
```

Market-RF (Mkt-RF):

- The ACF shows a moderate auto-correlation at lag 1 (0.107), but this correlation is not strong enough to conclusively indicate an AR(1) model. The correlation at other lags is low and sometimes negative, suggesting inconsistency and the lack of a clear pattern indicative of an AR(1) model.

SMB (Small Minus Big):

- The auto-correlation at lag 1 is relatively low (0.075). The correlations at higher lags fluctuate without showing a consistent pattern, further indicating the absence of a strong AR(1) model in the SMB time series.

HML (High Minus Low):

- This factor exhibits a somewhat higher auto-correlation at lag 1 (0.178) compared to the others. However, the auto-correlation is not sufficiently high to strongly suggest the presence of an AR(1) model, especially given the variability in correlations at higher lags.

Momentum:

- The auto-correlation at lag 1 is low (0.058). Coupled with low and sometimes negative correlations at higher lags, this indicates that the Momentum factor does not exhibit a significant AR(1) model pattern.

There is some level of auto-correlation at the first lag for these factors, the values are not sufficiently strong or consistent across lags to suggest the presence of an AR(1) model in any of the four time series. The fluctuations in correlation values at higher lags further support the absence of a clear AR(1) model in the Market-RF, SMB, HML, and Momentum time series.

Part II

$E(X_t) = 0.01 + 0.2 X_{t-2} + a_t$

$\Rightarrow E(X_t) = \dfrac{0.01}{1-0.2} = 0.0125.$

$\Rightarrow$ Taking variance

$Var(X_t) = 0.04\, Var(X_{t-2}) + \sigma_a^2$

$\therefore Var(X_t) = \dfrac{0.02}{1-0.96} = 0.0208$

Now, For Autocorrelation function,
drop constant term & multiply by X_{t-l}

$\Rightarrow X_t X_{t-l} = 0.2 X_{t-2} X_{t-l} + a_t X_{t-l}$

expectation of $l > 0$, $\Rightarrow \gamma_l = 0.2 \gamma_{l-2}$

$\therefore P_l = 0.2 P_{l-2}$ for $l > 0$

$\because P_0 = 1 \Rightarrow P_2 = 0.2$, $P_4 = 0.2^4$ .... (if even no.)

For $l = 1$, $P_l = P_{-l}$.

$\Rightarrow P_1 = 0.2 P_{-1} = 0.2 P_1$

$\Rightarrow 0.8 P_1 = 0$

$\therefore P_l = 0$ (if l is an odd no.)

1-step ahead forecast at $t = 100$, we have

$X_{101} = 0.01 + 0.2 X_{t\,99} + a_{101}$

$\Rightarrow X_{100}(1) = 0.01 + 0.2 X_{99}$

$= 0.01 + 0.2 \times (0.02)$

$= 0.014$

$e_{100}(1) = a_{101}$

std. deviation $\sqrt{0.02} = 0.141$.

2-step ahead forecast X_{102}

$\Rightarrow X_{102} = 0.01 + 0.2 \cancel{X}_{100} + a_{102}$

$\Rightarrow X_{100}(2) = 0.01 + 0.2(-0.01)$

$= 0.008$.

$\Rightarrow$ forecast error is a_{102}

with std. deviation $\sqrt{0.02} = 0.141$

$\Rightarrow$ 1) Mean $= 0.0125$, Variance $= 0.0208$

2) Auto correlation: $0.2, 0.2^2, 0.2^3 \ldots$, if even no.

0 if odd no.

3) 1-step ahead forecast $= 0.014$

2-step ahead forecast $= 0.141$