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	-0.016286	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	0.201698	0.215132	-0.280289
Beer	0.707673	0.351039	0.214982	-0.200077
Smoke	0.584268	0.103154	0.171809	-0.219165
Games	0.830211	0.412069	0.250367	-0.356992
Books	0.830092	0.408145	0.250608	-0.306399
Hshld	0.816234	0.261863	0.107373	-0.199683
Clths	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
Servs	0.495606	0.285060	-0.084935	-0.092466
BusEq	0.842644	0.329809	-0.066278	-0.228394
Paper	0.869099	0.275653	0.197336	-0.300174
Trans	0.858668	0.348099	0.430456	-0.423153
Whlsl	0.796012	0.459826	0.209154	-0.260379
Rtail	0.858747	0.302931	0.114997	-0.309088
Meals	0.753859	0.368637	0.163557	-0.250813
Fin	0.915884	0.282586	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

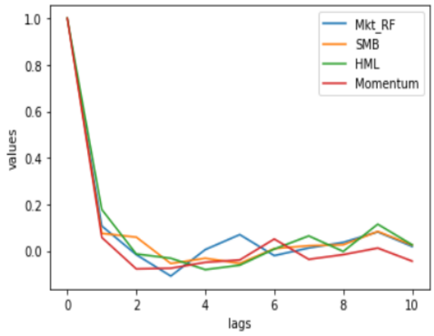
In [89]:
```
lagged_data_mkt_RF = pd.Series(sm.tsa.acf(data['Mkt-RF'],nlags = 10))
lagged_data_SMB = pd.Series(sm.tsa.acf(data['SMB'],nlags = 10))
lagged_data_HML = pd.Series(sm.tsa.acf(data['HML'],nlags = 10))
lagged_data_momentum = pd.Series(sm.tsa.acf(data['Mom'],nlags = 10))

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]:
```python
#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>
```



Tuesday, September 15, 2020 9:18 PM

$x_t = 0.01 + 0.2\, x_{t-2} + a_t$

$\{a_t\}$ is gaussian noise with mean $= 0$,
variance $= 0.02$

Acc to AR(2) model :

$x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + e_t$

$a_0 = 0.01$
$a_1 = 0$
$x_{t-1} = 0$
$a_2 = 0.2$
$e_t = a_t$

Part (a) : Mean & Variance of x_t

Taking expectation on both sides of Equation ②

$$E[x_t] = a_0 + a_1 E[x_{t-1}] + a_2 E[x_{t-2}] + E[e_t]$$
$$= 0.01 + 0.2\, E[x_{t-2}]$$

Under Stationary Conditions :

$E[x_t] = E[x_{t-1}] = E[x_{t-2}] = \mu$

$\mu = a_0 + \mu a_1 + \mu a_2$
$a_0 = \mu - \mu a_1 - \mu a_2$
$a_0 = \mu(1 - a_1 - a_2)$

$$\boxed{\mu = \frac{a_0}{1 - a_1 - a_2}} \Rightarrow \text{Mean}$$

$\mu = 0.01 + 0.2\mu$
$\mu - 0.2\mu = 0.01$
$0.8\mu = 0.01$
$\mu = \frac{0.01}{0.8}$
$= \frac{1}{80} = 0.0125$

Variance :

$a_0 = \mu(1 - a_1 - a_2)$

$x_t = \mu(1 - a_1 - a_2) + a_1 x_{t-1} + a_2 x_{t-2} + a_t$
$x_t = \mu - a_1\mu - a_2\mu + a_1 x_{t-1} + a_2 x_{t-2} + e_t$
$x_t - \mu = a_1(x_{t-1} - \mu) + a_2(x_{t-2} - \mu) + e_t$ — ①

Squaring & taking Expectation

$Var(x_t) = a_1^2\, Var(x_{t-1}) + a_2^2\, Var(x_{t-2}) + \sigma_e^2$

Under Stationary Condition :

$Var(x_t) = Var(x_{t-1}) = Var(x_{t-2})$

$Var(x_t) - a_1^2\, Var(x_{t-1}) - a_2^2\, Var(x_{t-2}) = \sigma_e^2$
$Var(x_t)[1 - a_1^2 - a_2^2] = \sigma_e^2$

$$\boxed{Var(x_t) = \frac{\sigma_e^2}{1 - a_1^2 - a_2^2}} = \frac{0.02}{1 - 0 - (0.2)^2} = \frac{0.02}{0.96} = \frac{2}{96} = \boxed{0.020833}$$

Part (b): Lag 1 and Lag 2 of auto-correlation of X_t

Multiplying Eq ① by $(X_{t-l} - \mu)$

$$(X_{t-l} - \mu)(X_t - \mu) = a_1 (X_{t-1} - \mu)(X_{t-l} - \mu) + a_2 (X_{t-2} - \mu)(X_{t-l} - \mu) + e_t (X_{t-l} - \mu)$$

Taking Expectation, where,

$$E[(X_{t-l} - \mu) e_t] = 0$$

$$v_l = a_1 v_{l-1} + a_2 v_{l-2}$$

when divided by $v_0 \Rightarrow$

$$\frac{v_l}{v_0} = a_1 \frac{v_{l-1}}{v_0} + a_2 \frac{v_{l-2}}{v_0}$$

$$r_l = a_1 r_{l-1} + a_2 r_{l-2} \qquad [l > 0]$$

for lag = 1

$$r_1 = a_1 r_0 + a_2 r_{(-1)}$$

$$r_1 = a_1 + a_2 (r_{(-1)})$$

$$r_1 = \frac{a_1}{1 - a_2} = \frac{0}{1 - 0.2} = 0$$

$$r_2 = a_1 r_1 + a_2 r_0$$

$$r_2 = 0 + 1(0.2)$$

$$r_2 = 0.2$$

Part (c)

Equation : $X_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + e_t$

$t = 101$

$$\begin{aligned} X_{101} &= a_0 + a_1 X_{100} + a_2 X_{99} + e_t \\ &= 0.01 + 0 + 0.2(0.02) + e_t \\ &= 0.01 + 0.2(0.02) + e_t \\ &= 0.01 + 0.004 + e_t \\ &= 0.014 + e_t \end{aligned}$$

$$\begin{aligned} X_{102} &= a_0 + a_1 X_{101} + a_2 X_{100} + e_t \\ &= 0.01 + 0 + 0.2(-0.01) + e_t \\ &= 0.01 + (-0.002) + e_t \\ &= (8 \times 10^{-3}) + e_t \end{aligned}$$