timeserieslab7-rohramehak-251524

April 17, 2024

[12]:
```
import datetime
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
```

1 Correlations in AR models

An autoregressive model of order p is defined as yt = c + phi1 yt−1 + phi2 yt−2 + … + phi_p yt−p + t,

Write a function that calculates the values of AR(p) model. The function must have a parameter burnin that determines how many initial values are discarded.

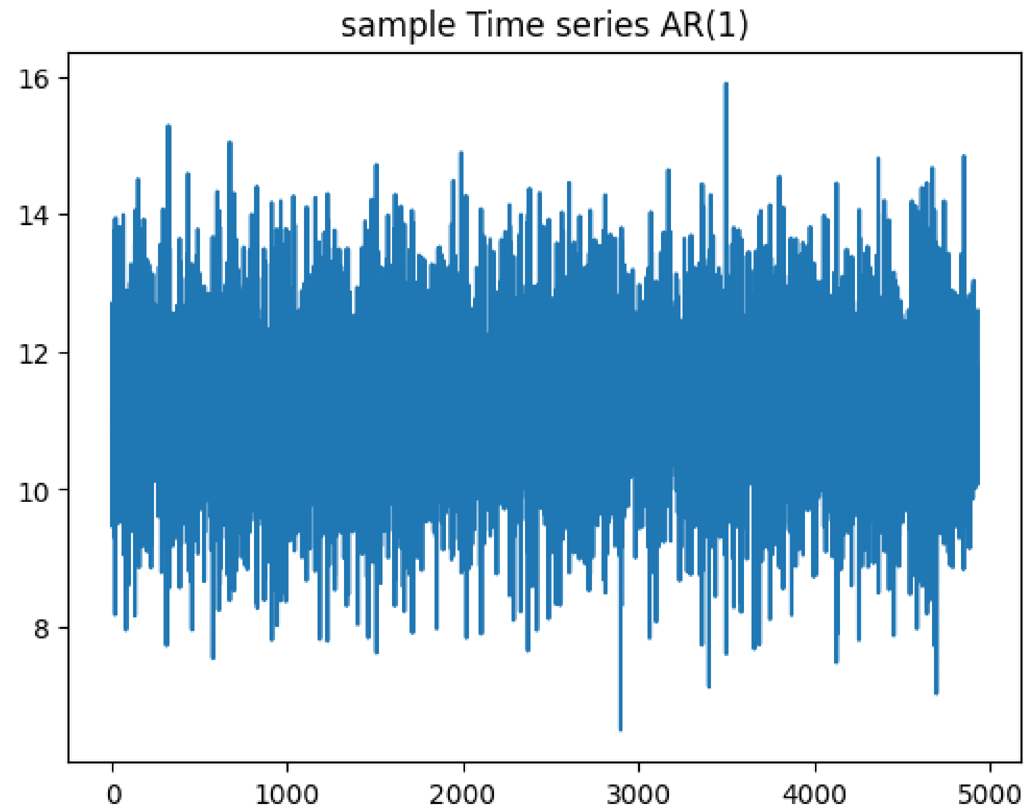
[21]:
```
def ar_model(p, ar_coeffs, c, n_samples, burnin=10):
    et = np.random.normal(size=n_samples)
    ar_values = np.zeros_like(et, dtype=float)
    for t in range(p, n_samples):
        ar_values[t] = c + np.sum(ar_coeffs[::-1] * ar_values[t-p:t]) + et[t]
    return ar_values[burnin:]
```

Calculate n = 5000 values of the AR(1) model yt = 18 − 0.6yt−1 + t.

[97]:
```
order = 1
coefficients = [-0.6]
num_samples = 5000
c = 18
```

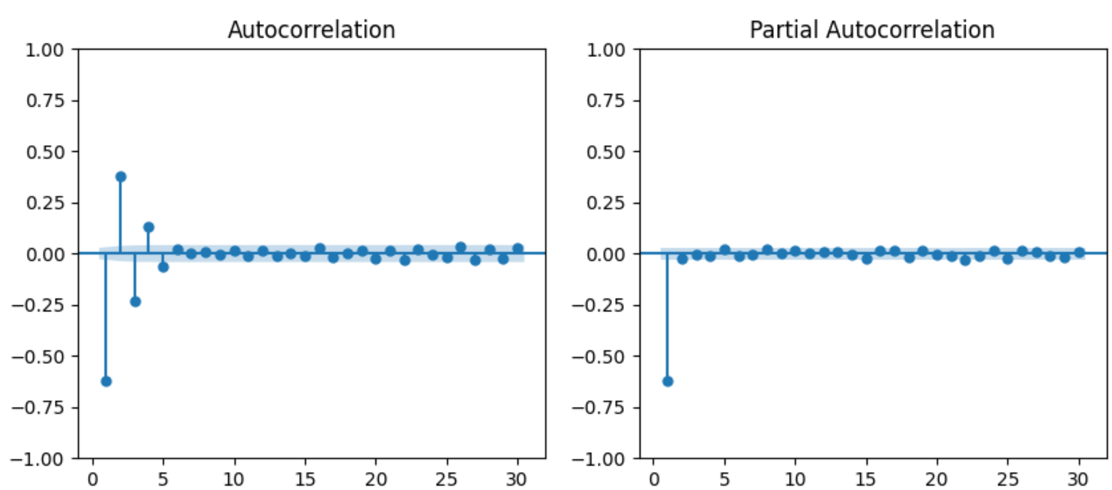
[98]:
```
ar_values = ar_model(order, coefficients, c, num_samples, burnin=70)
```

[99]:
```
plt.plot(ar_values)
plt.title("sample Time series AR(1)")
plt.show()
```



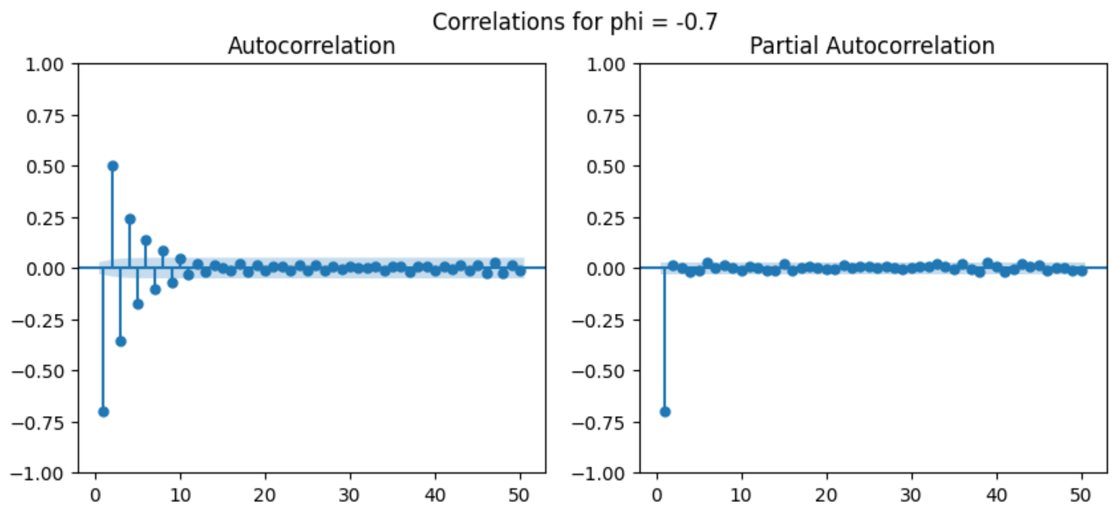
- Calculate the autocorrelation (ACF) and partial autocorrelation (PACF) function for this time series.

[100]:
```
fig, axs = plt.subplots(1, 2, figsize=(10, 4))
plot_acf(ar_values, lags=30, zero=False, ax=axs[0])
plot_pacf(ar_values, lags=30, zero=False, ax=axs[1])
plt.show()
```

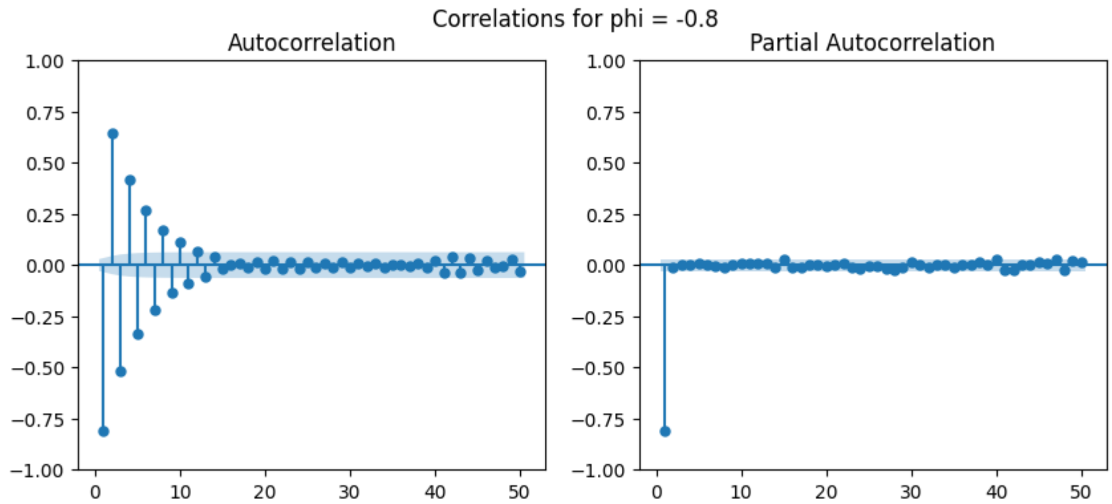


- Repeat the calculations for phi_1 = −0.7, −0.8, −0.9.

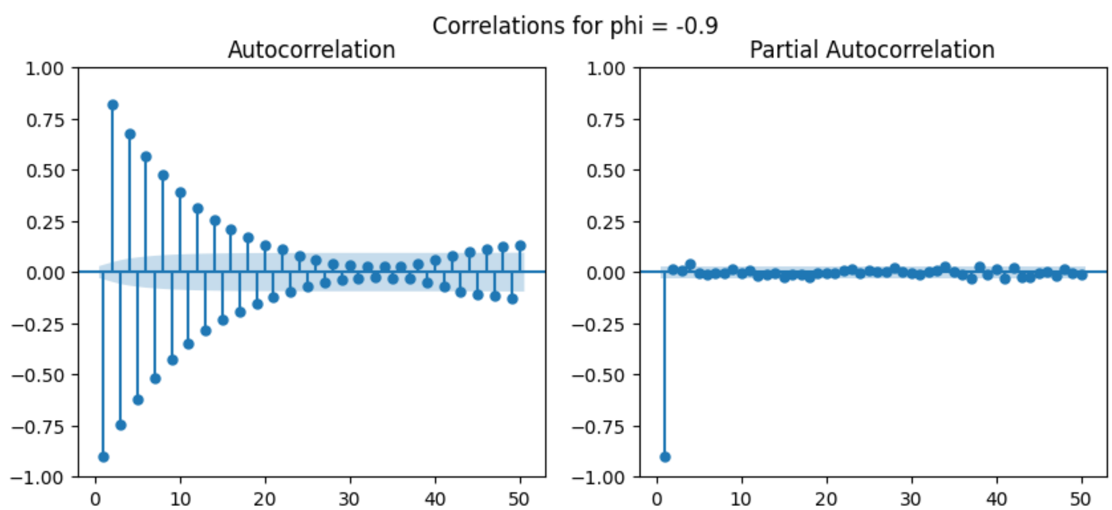
[101]:
```
order = 1
coefficients = [-0.7]
num_samples = 5000
c = 18
ar_values = ar_model(order, coefficients, c, num_samples, burnin=500)
fig, axs = plt.subplots(1, 2, figsize=(10, 4))
plt.suptitle("Correlations for phi = -0.7")
plot_acf(ar_values, lags=50, zero=False, ax=axs[0])
plot_pacf(ar_values, lags=50, zero=False, ax=axs[1])
plt.show()
```



[102]:
```
order = 1
coefficients = [-0.8]
num_samples = 5000
c = 18
ar_values = ar_model(order, coefficients, c, num_samples, burnin=500)
fig, axs = plt.subplots(1, 2, figsize=(10, 4))
plt.suptitle("Correlations for phi = -0.8")
plot_acf(ar_values, lags=50, zero=False, ax=axs[0])
plot_pacf(ar_values, lags=50, zero=False, ax=axs[1])
plt.show()
```



[103]:
```
order = 1
coefficients = [-0.9]
num_samples = 5000
c = 18
ar_values = ar_model(order, coefficients, c, num_samples, burnin=500)
fig, axs = plt.subplots(1, 2, figsize=(10, 4))
plt.suptitle("Correlations for phi = -0.9")
plot_acf(ar_values, lags=50, zero=False, ax=axs[0])
plot_pacf(ar_values, lags=50, zero=False, ax=axs[1])
plt.show()
```



What happens when $| 1| > 1$

[ ]:
```
order = 1
coefficients = [-1.9]
num_samples = 5000
c = 18
ar_values = ar_model(order, coefficients, c, num_samples, burnin=500)
ar_values
```

```
<ipython-input-125-25dd3f363950>:4: RuntimeWarning: overflow encountered in
multiply
  ar_values[t] = c + np.sum(ar_coeffs[::-1] * ar_values[t-p:t]) +
np.random.normal()
```

[ ]:
```
array([-1.27835765e+140,  2.42887954e+140, -4.61487112e+140, …,
                    inf,             -inf,              inf])
```

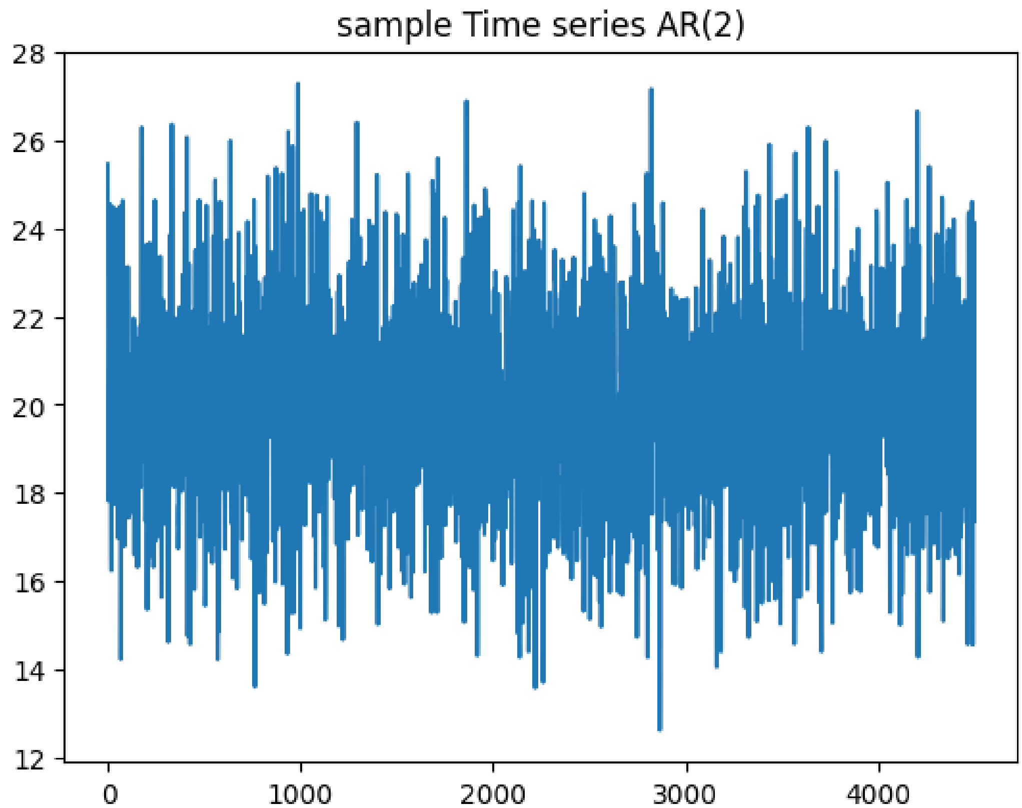
If we set |phi| > 1 , our time series goes to infinity

2 AR(2)

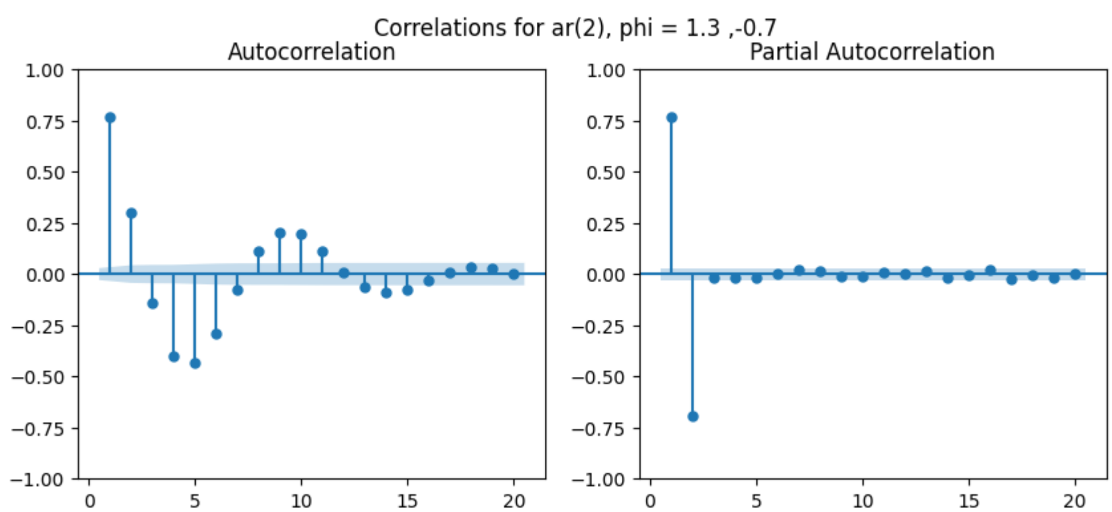
Calculate n = 5000 values of the AR(2) model yt = 8 + 1.3yt−1 − 0.7yt−2 + t. Compare the structure of PACFs for AR(1) and AR(2) models.

[104]:
```
order = 2
coefficients = [1.3, -0.7]
num_samples = 5000
c = 8
ar_values = ar_model(order, coefficients, c, num_samples, burnin=500)
```

[105]:
```
plt.plot(ar_values)
plt.title("sample Time series AR(2)")
plt.show()
```



[106]:
```
fig, axs = plt.subplots(1, 2, figsize=(10, 4))
plt.suptitle("Correlations for ar(2), phi = 1.3 ,-0.7")
plot_acf(ar_values, lags=20, zero=False, ax=axs[0])
plot_pacf(ar_values, lags=20, zero=False, ax=axs[1])
plt.show()
```



For an AR(1) model, the PACF will have a sharp drop after lag 1 because it directly measures the correlation between a variable and its lagged values, and after the first lag, all other lags are indirectly related through the first lag. So, the PACF will be significant at lag 1 and then drop to approximately zero for all other lags.

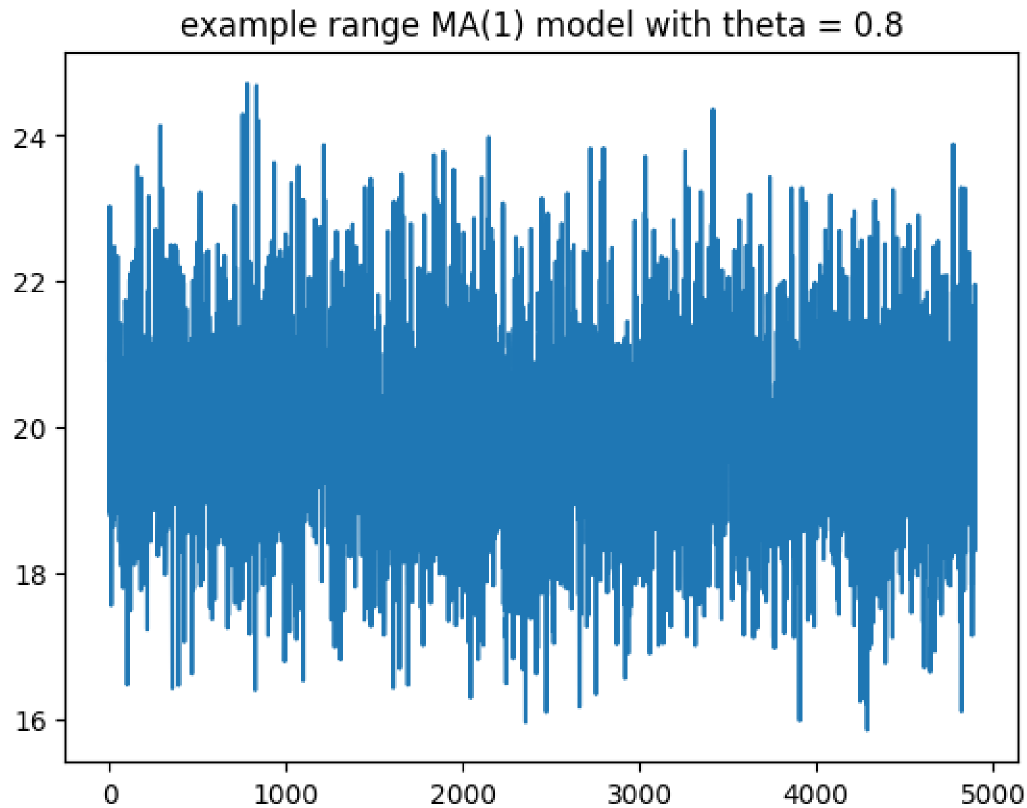
For an AR(2) model, the PACF will be significant at lags 1 and 2 because the variable is directly related to its first and second lagged values

3 MOVING AVERAGES

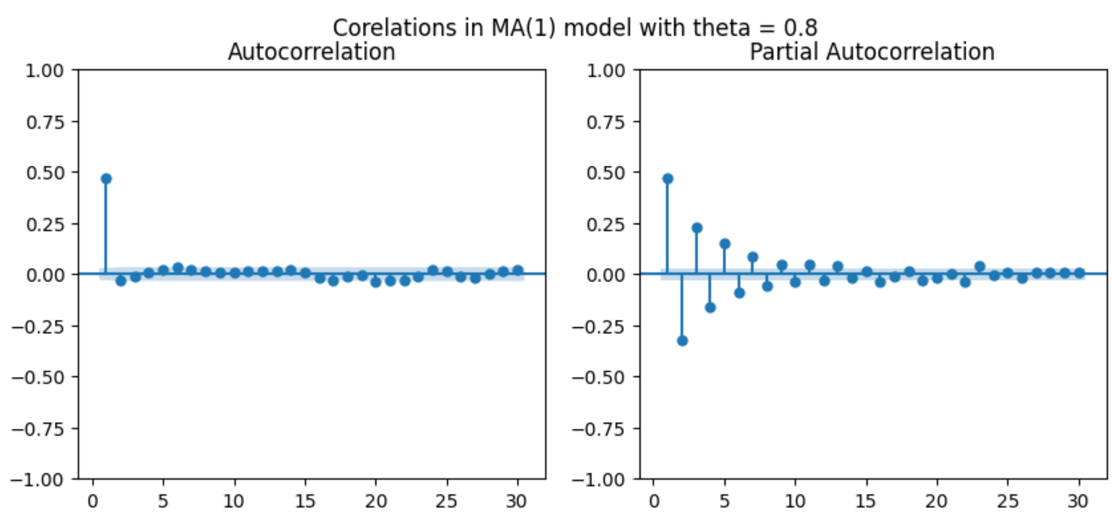
[107]:
```
def ma_model(q, ma_coeffs, c, n_samples, burnin=10):
    ma_values = np.zeros(n_samples)
    et_values = np.random.normal(size=n_samples)
    for t in range(q, n_samples):
        ma_values[t] = c + np.dot(ma_coeffs[::-1], et_values[t-q:t]) + et_values[t]
    return ma_values[burnin:]
```

[108]:
```
q = 1
coefficients = [0.8]
num_samples = 5000
c = 20
ma_values = ma_model(q, coefficients, c, num_samples, burnin=100)
```

[109]:
```
plt.plot(ma_values)
plt.title("example range MA(1) model with theta = 0.8")
plt.show()
```

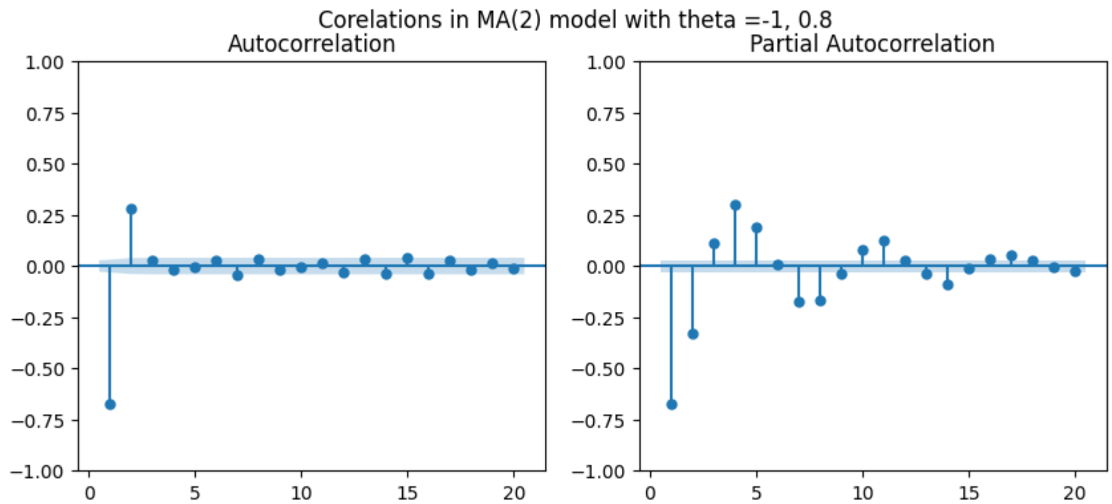


[111]:
```
fig, axs = plt.subplots(1, 2, figsize=(10, 4))
plt.suptitle("Corelations in MA(1) model with theta = 0.8")
plot_acf(ma_values, lags=30, zero=False, ax=axs[0])
plot_pacf(ma_values, lags=30, zero=False, ax=axs[1])
plt.show()
```



[112]:
```python
q = 2
coefficients = [-1, 0.8]
num_samples = 5000
c = 0
ma_values = ma_model(q, coefficients, c, num_samples, burnin=100)
```

[113]:
```python
fig, axs = plt.subplots(1, 2, figsize=(10, 4))
plt.suptitle("Corelations in MA(2) model with theta =-1, 0.8")
plot_acf(ma_values, lags=20, zero=False, ax=axs[0])
plot_pacf(ma_values, lags=20, zero=False, ax=axs[1])
plt.show()
```



In acf :

1. for AR model : Slowly decaying. It will show significant autocorrelation values for several lags.
2. for MA Model : Sharp drop after the lag corresponding to the order of the MA model. All subsequent lags should be approximately zero.

in Pacf :

1. for AR model : Sharp drop after the lag corresponding to the order of the AR model. All subsequent lags should be approximately zero
2. for MA model : Slowly decaying. It will show significant partial autocorrelation values for several lags.

4 AR(p) model fitting

[60]:
```
order = 2
coefficients = [1.3, -0.7]
num_samples = 5000
c = 8
ar2_values = ar_model(order, coefficients, c, num_samples, burnin=20)
```

[61]:
```
from statsmodels.tsa.arima.model import ARIMA
aic_values= []
models = []
p_values = [1, 2, 3, 4]
for p in p_values:
    model = ARIMA(ar2_values, order=(p, 0, 0))
    model_fit = model.fit()
    models.append(model_fit)
    estimated_params = model_fit.params
    print(f"\n------------------Model AR({p})------------------\n")
    print(model_fit.summary())
    aic_values.append(model_fit.aic)
```

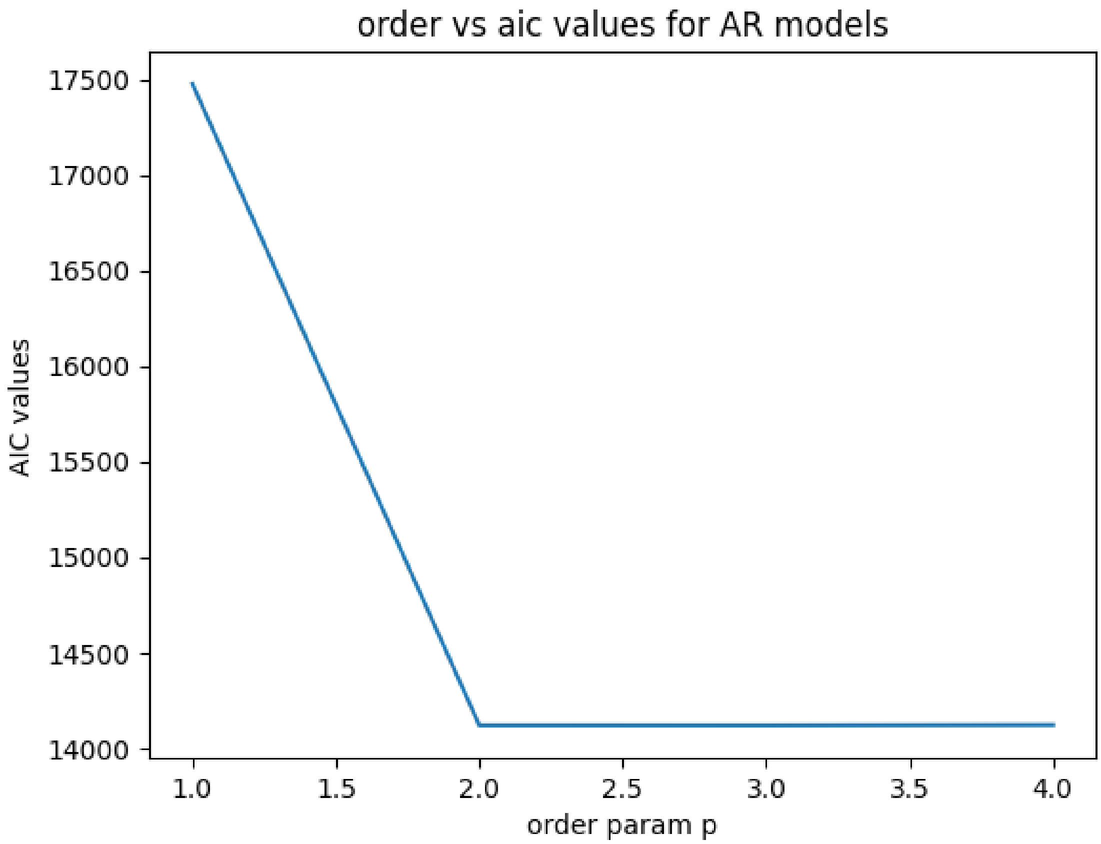
```
------------------Model AR(1)------------------

                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 4980
Model:                 ARIMA(1, 0, 0)   Log Likelihood               -8735.110
Date:                Wed, 17 Apr 2024   AIC                          17476.219
Time:                        15:45:29   BIC                          17495.759
Sample:                             0   HQIC                         17483.069
                               - 4980
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         20.0188      0.086    233.468      0.000      19.851      20.187
ar.L1          0.7691      0.009     84.789      0.000       0.751       0.787
sigma2         1.9544      0.039     50.211      0.000       1.878       2.031
================================================================================
===
Ljung-Box (L1) (Q):                1444.84   Jarque-Bera (JB):
0.73
Prob(Q):                              0.00   Prob(JB):
0.69
Heteroskedasticity (H):               0.87   Skew:
0.03
```

```
Prob(H) (two-sided):                  0.00   Kurtosis:
3.03
================================================================================
===

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).

------------------Model AR(2)------------------

                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 4980
Model:                 ARIMA(2, 0, 0)   Log Likelihood               -7055.698
Date:                Wed, 17 Apr 2024   AIC                          14119.395
Time:                        15:45:31   BIC                          14145.448
Sample:                             0   HQIC                         14128.528
                               - 4980
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         20.0192      0.036    554.918      0.000      19.949      20.090
ar.L1          1.3080      0.010    128.986      0.000       1.288       1.328
ar.L2         -0.7004      0.010    -68.718      0.000      -0.720      -0.680
sigma2         0.9953      0.020     50.327      0.000       0.957       1.034
================================================================================
===
Ljung-Box (L1) (Q):                   1.08   Jarque-Bera (JB):
0.36
Prob(Q):                              0.30   Prob(JB):
0.83
Heteroskedasticity (H):               0.99   Skew:
-0.01
Prob(H) (two-sided):                  0.81   Kurtosis:
3.03
================================================================================
===

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).

------------------Model AR(3)------------------

                               SARIMAX Results
==============================================================================
```

```
Dep. Variable:                      y   No. Observations:                 4980
Model:                 ARIMA(3, 0, 0)   Log Likelihood               -7054.577
Date:                Wed, 17 Apr 2024   AIC                          14119.155
Time:                        15:45:34   BIC                          14151.721
Sample:                             0   HQIC                         14130.571
                               - 4980
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         20.0191      0.037    543.212      0.000      19.947      20.091
ar.L1          1.3228      0.015     91.121      0.000       1.294       1.351
ar.L2         -0.7282      0.021    -34.101      0.000      -0.770      -0.686
ar.L3          0.0212      0.014      1.486      0.137      -0.007       0.049
sigma2         0.9948      0.020     50.340      0.000       0.956       1.034
================================================================================
===
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):
0.41
Prob(Q):                              1.00   Prob(JB):
0.82
Heteroskedasticity (H):               0.99   Skew:
-0.01
Prob(H) (two-sided):                  0.80   Kurtosis:
3.04
================================================================================
===

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).

------------------Model AR(4)------------------

                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 4980
Model:                 ARIMA(4, 0, 0)   Log Likelihood               -7054.546
Date:                Wed, 17 Apr 2024   AIC                          14121.091
Time:                        15:45:36   BIC                          14160.170
Sample:                             0   HQIC                         14134.790
                               - 4980
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         20.0192      0.037    545.025      0.000      19.947      20.091
ar.L1          1.3229      0.015     91.108      0.000       1.294       1.351
```

```
ar.L2         -0.7308      0.024    -30.806      0.000      -0.777      -0.684
ar.L3          0.0260      0.024      1.104      0.270      -0.020       0.072
ar.L4         -0.0036      0.014     -0.253      0.800      -0.031       0.024
sigma2         0.9948      0.020     50.339      0.000       0.956       1.034
================================================================================
===
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):
0.41
Prob(Q):                              0.99   Prob(JB):
0.81
Heteroskedasticity (H):               0.99   Skew:
-0.01
Prob(H) (two-sided):                  0.80   Kurtosis:
3.04
================================================================================
===

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).
```

[62]:
```python
plt.plot(p_values, aic_values)
plt.xlabel("order param p")
plt.ylabel("AIC values")
plt.title("order vs aic values for AR models")
plt.show()
```



since the aic values of ar(2), ar(3) and ar(4) are very close. we can do llr test to determine the better one

[58]:
```
from scipy.stats.distributions import chi2
```

[73]:
```
def LLR_test(m1,m2,DF=1):
 L1=m1.llf
 L2=m2.llf
 LR=2*(L2-L1)
 p=chi2.sf(LR,DF).round(3)
 return p
```

in or models list,

1. model[0] is ar(1),
2. model[1] is ar(2)
3. model[2] is ar(3) and
4. model[3] is ar(4)

[96]:
```
LLR_test(models[2], models[1])
```

[96]: 1.0

since p_values > 0.05 , we cant say that ar(2) is much better than ar(1)

[75]:
```
LLR_test(models[2], models[3])
```

[75]: 0.801

5 MA(q) Fitting

we fit MA(2) time series to our model

[69]:
```
q = 2
coefficients = [-1, 0.8]
num_samples = 5000
c = 0
ma_values = ma_model(q, coefficients, c, num_samples, burnin=100)
```

[76]:
```
from statsmodels.tsa.arima.model import ARIMA
aic_values= []
models = []
q_values = [1, 2, 3, 4]
for q in q_values:
    model = ARIMA(ma_values, order=(0, 0, q))
    model_fit = model.fit()
    models.append(model_fit)
    estimated_params = model_fit.params
    print(f"\n------------------Model MA({q})------------------\n")
    print(model_fit.summary())
    aic_values.append(model_fit.aic)
```

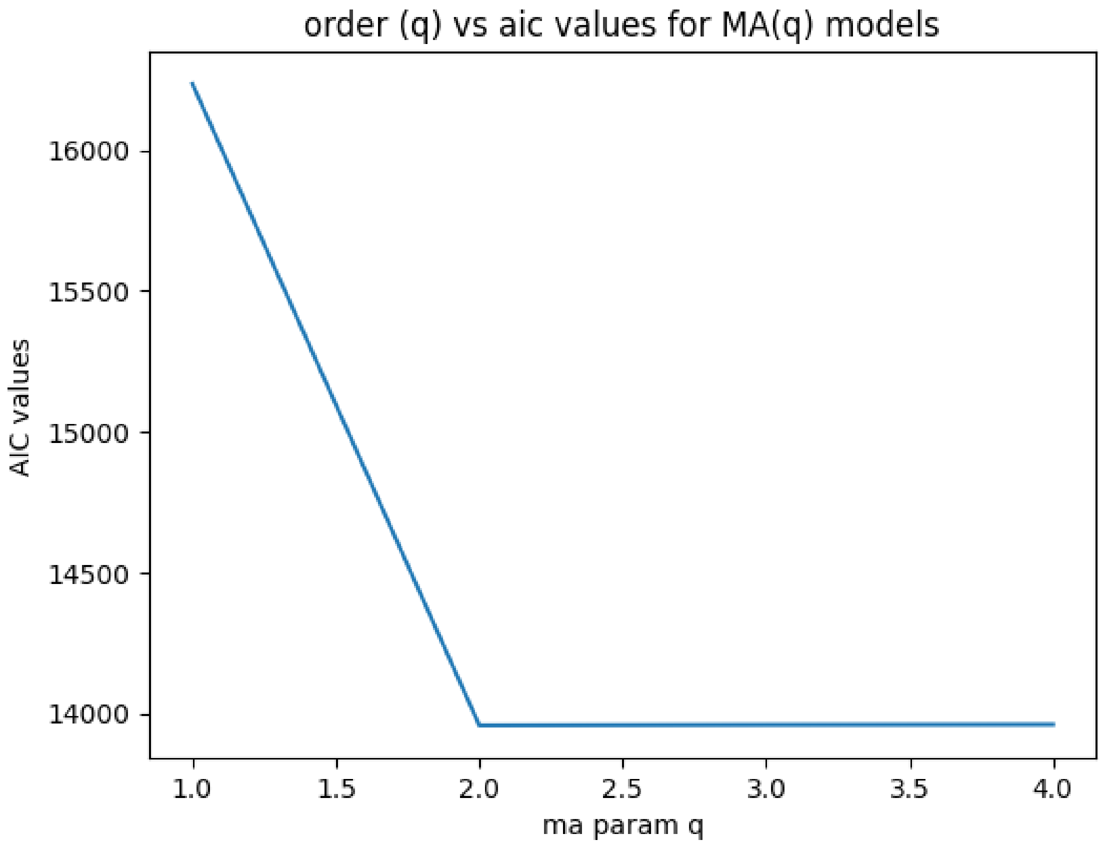
```
------------------Model MA(1)------------------

                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 4900
Model:                 ARIMA(0, 0, 1)   Log Likelihood               -8114.474
Date:                Wed, 17 Apr 2024   AIC                          16234.947
Time:                        16:05:31   BIC                          16254.438
Sample:                             0   HQIC                         16241.785
                               - 4900
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0090      0.008      1.144      0.253      -0.006       0.024
```

```
ma.L1         -0.5681      0.012    -49.322      0.000      -0.591      -0.546
sigma2         1.6065      0.033     49.403      0.000       1.543       1.670
================================================================================
===
Ljung-Box (L1) (Q):                 365.52   Jarque-Bera (JB):
1.39
Prob(Q):                              0.00   Prob(JB):
0.50
Heteroskedasticity (H):               1.09   Skew:
-0.04
Prob(H) (two-sided):                  0.09   Kurtosis:
2.99
================================================================================
===

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).

------------------Model MA(2)------------------

                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 4900
Model:                 ARIMA(0, 0, 2)   Log Likelihood               -6974.524
Date:                Wed, 17 Apr 2024   AIC                          13957.048
Time:                        16:05:33   BIC                          13983.036
Sample:                             0   HQIC                         13966.165
                               - 4900
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0091      0.011      0.813      0.416      -0.013       0.031
ma.L1         -1.0006      0.009   -110.031      0.000      -1.018      -0.983
ma.L2          0.7772      0.009     86.185      0.000       0.760       0.795
sigma2         1.0084      0.020     49.225      0.000       0.968       1.049
================================================================================
===
Ljung-Box (L1) (Q):                   0.05   Jarque-Bera (JB):
0.67
Prob(Q):                              0.83   Prob(JB):
0.72
Heteroskedasticity (H):               1.05   Skew:
-0.03
Prob(H) (two-sided):                  0.35   Kurtosis:
2.98
================================================================================
```

```
===

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).

------------------Model MA(3)------------------

                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 4900
Model:                 ARIMA(0, 0, 3)   Log Likelihood               -6974.485
Date:                Wed, 17 Apr 2024   AIC                          13958.970
Time:                        16:05:36   BIC                          13991.455
Sample:                             0   HQIC                         13970.367
                               - 4900
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0091      0.011      0.816      0.414      -0.013       0.031
ma.L1         -1.0037      0.014    -71.307      0.000      -1.031      -0.976
ma.L2          0.7812      0.017     46.708      0.000       0.748       0.814
ma.L3         -0.0040      0.014     -0.281      0.779      -0.032       0.024
sigma2         1.0084      0.020     49.222      0.000       0.968       1.049
================================================================================
===
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):
0.67
Prob(Q):                              1.00   Prob(JB):
0.72
Heteroskedasticity (H):               1.05   Skew:
-0.03
Prob(H) (two-sided):                  0.36   Kurtosis:
2.98
================================================================================
===

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).

------------------Model MA(4)------------------

                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 4900
Model:                 ARIMA(0, 0, 4)   Log Likelihood               -6973.995
```

```
Date:                  Wed, 17 Apr 2024   AIC                          13959.989
Time:                          16:05:41   BIC                          13998.971
Sample:                               0   HQIC                         13973.666
                                 - 4900
Covariance Type:                    opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0091      0.011      0.828      0.408      -0.012       0.031
ma.L1         -1.0035      0.014    -71.237      0.000      -1.031      -0.976
ma.L2          0.7691      0.020     39.140      0.000       0.731       0.808
ma.L3          0.0116      0.020      0.588      0.556      -0.027       0.050
ma.L4         -0.0148      0.014     -1.059      0.289      -0.042       0.013
sigma2         1.0082      0.021     49.108      0.000       0.968       1.048
================================================================================
===
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):
0.74
Prob(Q):                              0.99   Prob(JB):
0.69
Heteroskedasticity (H):               1.05   Skew:
-0.03
Prob(H) (two-sided):                  0.36   Kurtosis:
2.98
================================================================================
===

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).
```

[72]:
```python
plt.plot(q_values, aic_values)
plt.xlabel("ma param q")
plt.ylabel("AIC values")
plt.title("order (q) vs aic values for MA(q) models")
plt.show()
```



[82]:
```
LLR_test(models[0], models[1])
```

[82]: 0.0

MA(2) is better than MA(1) p =0

[84]:
```
LLR_test(models[1], models[2])
```

[84]: 0.78

THE AIC value of model[1] that is MA(2) is slighlty lower than MA(3) and MA(4).But we cannot certainly say that MA(2) is definitely better than MA(3)
