PREDICTIVE ANALYSIS FOR DECISION MAKING

WEEKS 7 AND 8

MODELLING TIME SERIES MODELS: ARMA MODELS I

Dr Issam Malki
School of Finance and Accounting
Westminster Business School

November 2021

ARMA MODELS: ASSUMPTIONS

- We employ two types of stationary processes
  - Strictly Stationary Process

$$P\{y_{t_1} \le b_1, \ldots, y_{t_n} \le b_n\} = P\{y_{t_1+m} \le b_1, \ldots, y_{t_n+m} \le b_n\}$$

  - Weakly (covariance) stationary process

1. $E(y_t) = \mu\,, \qquad t = 1,2,\ldots,\infty$
2. $E(y_t - \mu)(y_t - \mu) = \sigma^2 < \infty$
3. $E(y_{t_1} - \mu)(y_{t_2} - \mu) = \gamma_{t_2 - t_1}$ for any t_1 , t_2

MODELLING STATIONARY DATA

- Dynamics of a single time series
  - Dependents on the data generating process
  - General properties include
    - Autoregressive behaviour
    - Moving average behaviour
    - Combination of the two
    - Others (not considered in this course)

- Autoregressive Process
  - The model can be explained by its past values
  - The general form is called AR(p) model
  - The error term, u, is assumed to be iid with zero mean and constant variance

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + u_t$$

AR(1) MODEL

- Remarks
  - The issue is choosing the lag length *p*
  - Many methods
    - Using information criteria
    - Using t statistics and significance of the lag
    - Using Box-Jenkins Method
- AR(1)

$$y_t = \mu + \phi_1 y_{t-1} + u_t$$

  - The slope captures the effect of the past shock
  - For stationary data: $0<|\phi_1|<1$
  - The closer to 1: the longer the shock remains
  - The closer to 0: the faster the shock disappears

Properties of Stationary AR(1)

- The statistical properties

$$E(y_t) = \frac{\mu}{(1-\phi_1 L)}$$

$$var(y_t) = \frac{\sigma^2}{(1-\phi_1^2)}$$

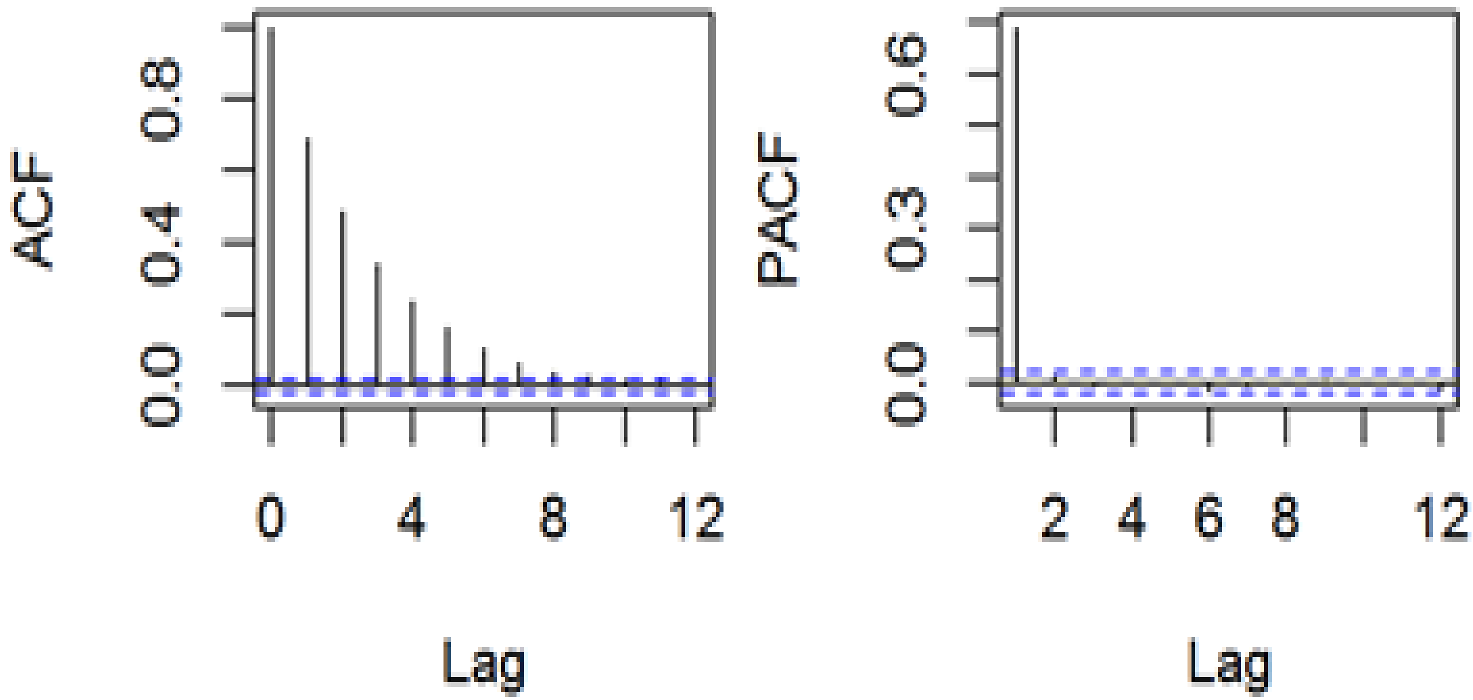
$$cov(y_t, y_{t-j}) = \phi_1^j \gamma_0$$

$$\rho_j = corr(y_t, y_{t-j}) = \frac{\gamma_j}{\gamma_0} = \frac{\phi_1^j \gamma_0}{\gamma_0} = \phi_1^j$$

The last part implies that the process converges back to the long run:

$$\lim_{j\to\infty} \rho_j = \lim_{j\to\infty} \phi_1^j = 0 \text{ since } |\phi_1| < 1$$

Properties of Stationary AR(1)



Properties of Stationary AR(2)

- The AR(2) Model

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$$

where $|\phi_1 + \phi_2| < 1$

The statistical properties

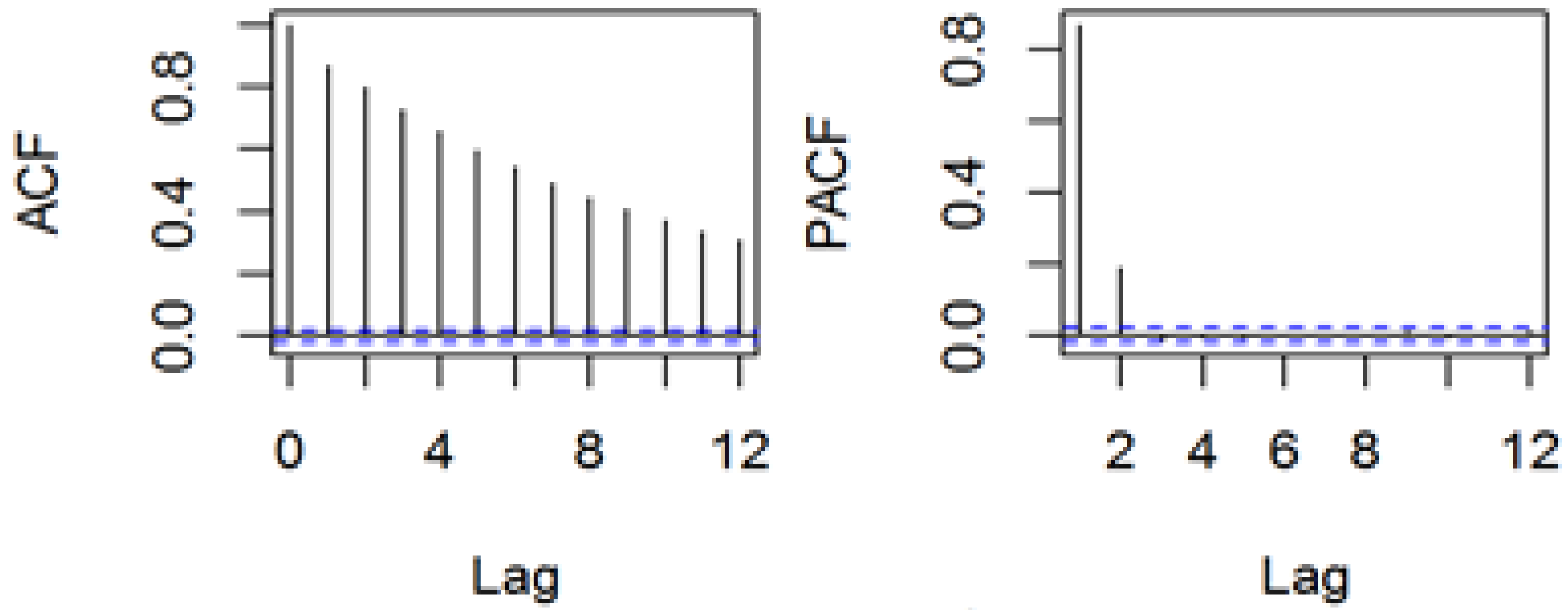
$$E(y_t) = \frac{\mu}{(1-\phi_1-\phi_2)}$$

$$var(y_t) = \frac{\sigma^2}{(1-\phi_1^2-\phi_2^2)}$$

$$cov(y_t, y_{t-j}) = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2}$$

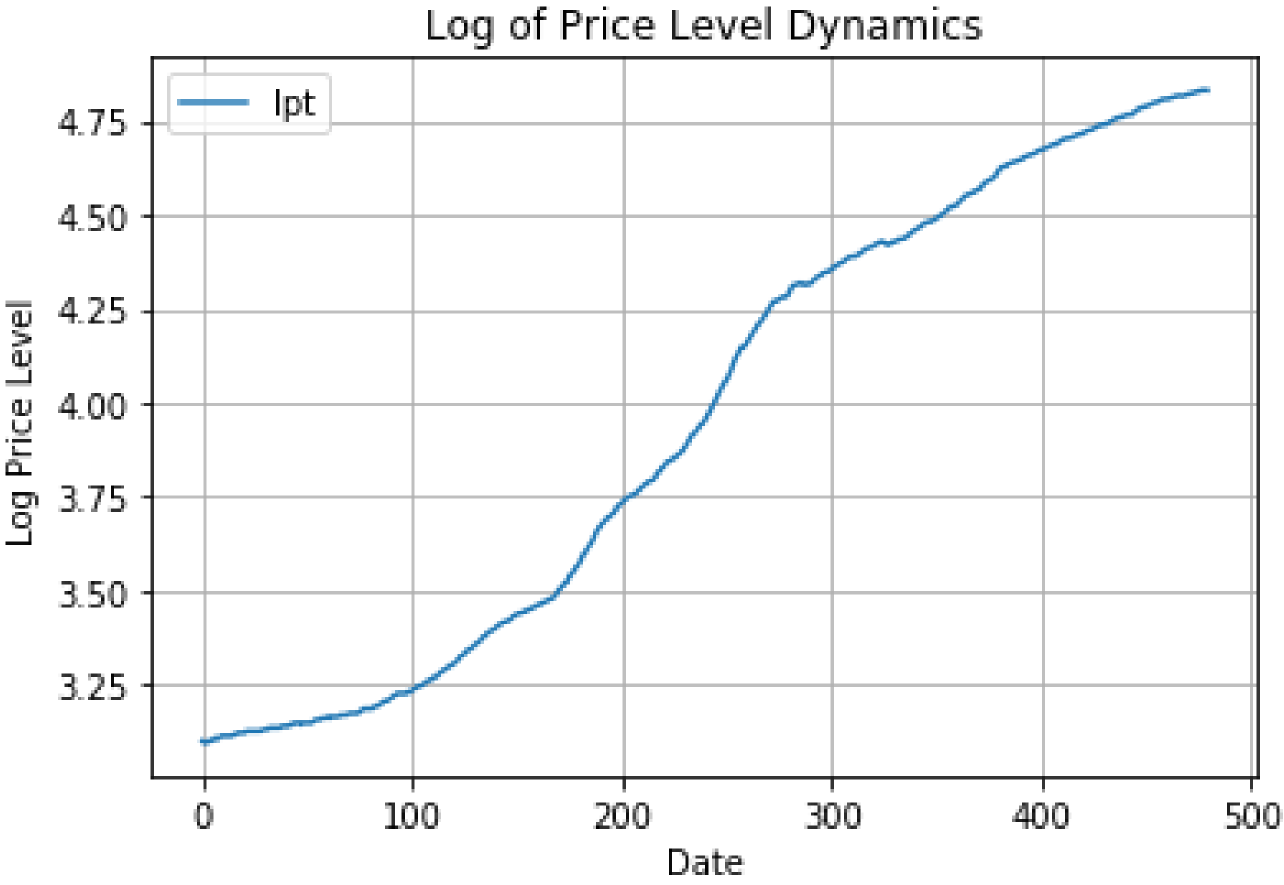
$$\rho_j = corr(y_t, y_{t-j}) = \frac{\gamma_j}{\gamma_0}$$

PROPERTIES OF STATIONARY AR(2)



EXAMPLE (1)

- Consider the price level

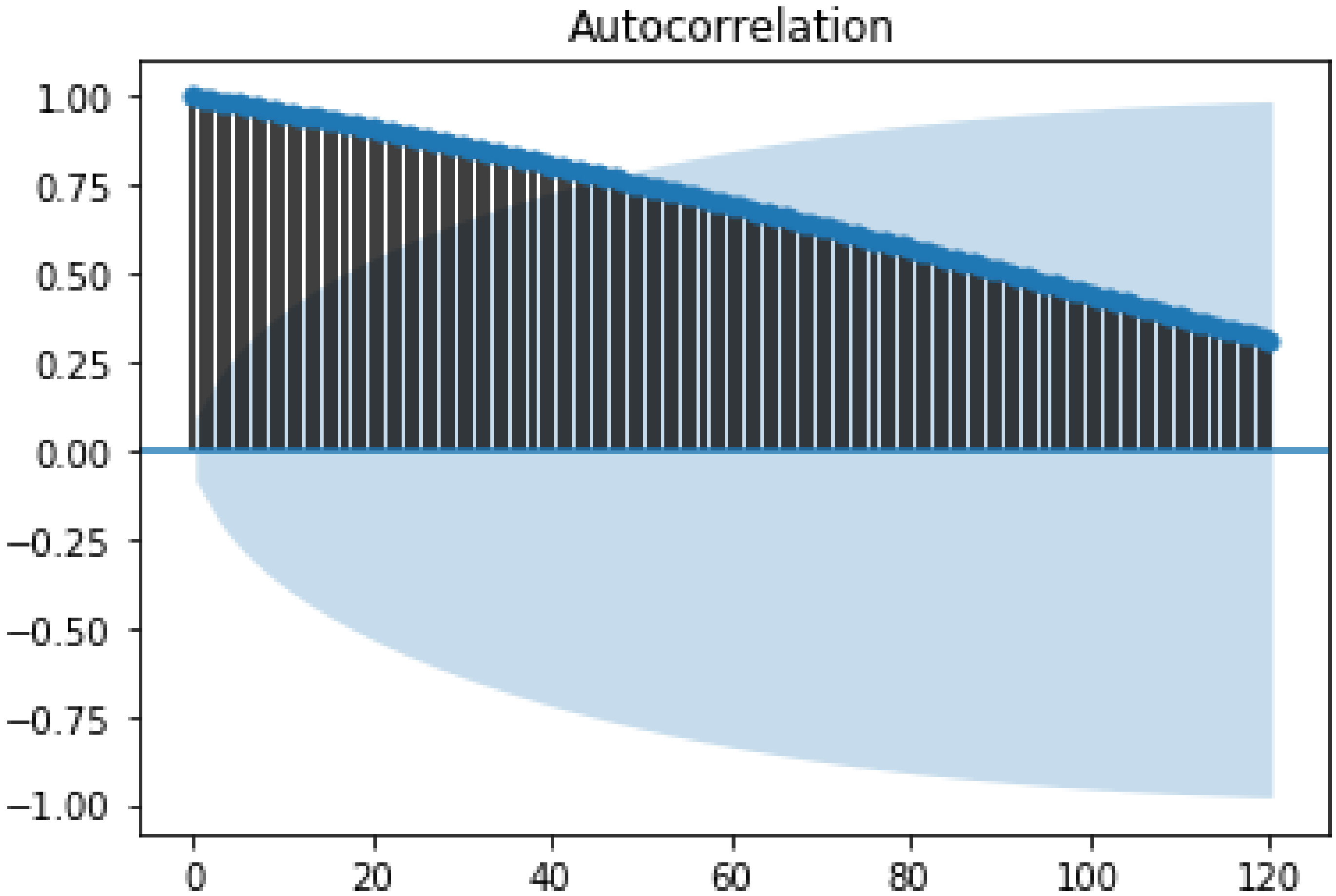


Example (1)

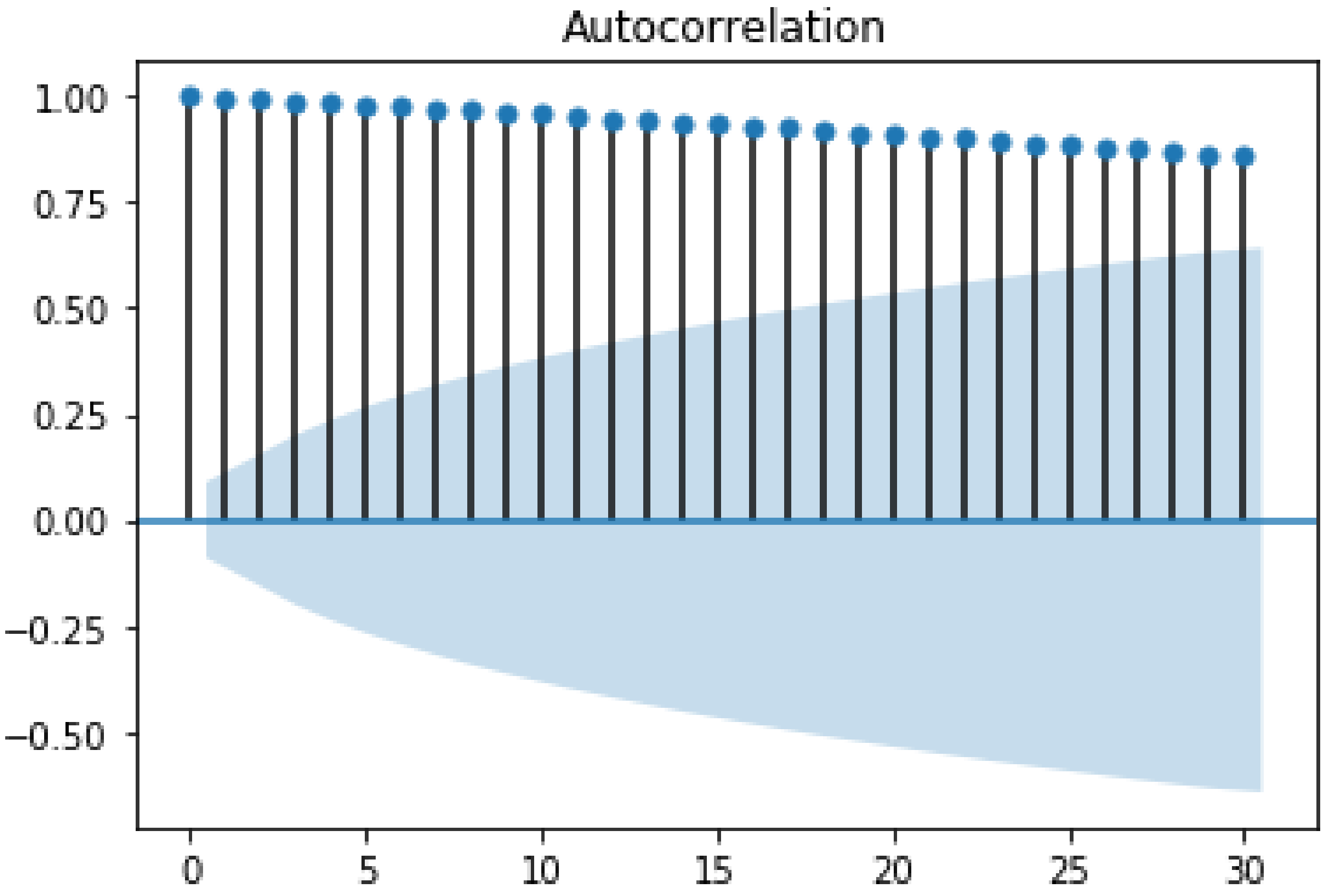
- Testing for the presence of stationarity:
  - Inspect the ACF.
  - Test whether the process is white noise using the Q statistic.
  - Confirm after estimation by inspecting the coefficients of the autoregressive process.

- Inspecting the ACF
  - Choose the lag length *(k-th):* rule of thumb is to use quarter of the sample size. You can use less or more. The outcome is very sensitive to the lag length.
  - If the ACF dies off before the *k*-th lag, we conclude stationarity.
  - This test is informal. So you have to be careful.
- See below

EXAMPLE (1): *k=120*



EXAMPLE (1): *k=30*



Example (1)

- Testing the White Noise Hypothesis
  - Use the Q statistic, a non parametric test for autocorrelation.
  - The test statistics is called Ljung – Box

$$Q = T(T+2)\sum_{j=1}^{k}\frac{\hat{\rho}_j^2}{T-k} \sim \chi_k^2$$

  - The null hypothesis states white noise, the alternative is otherwise.
  - Again, the test is sensitive to the number of lags, *k*. I report here one example using 120 observations using Python.

EXAMPLE (1)

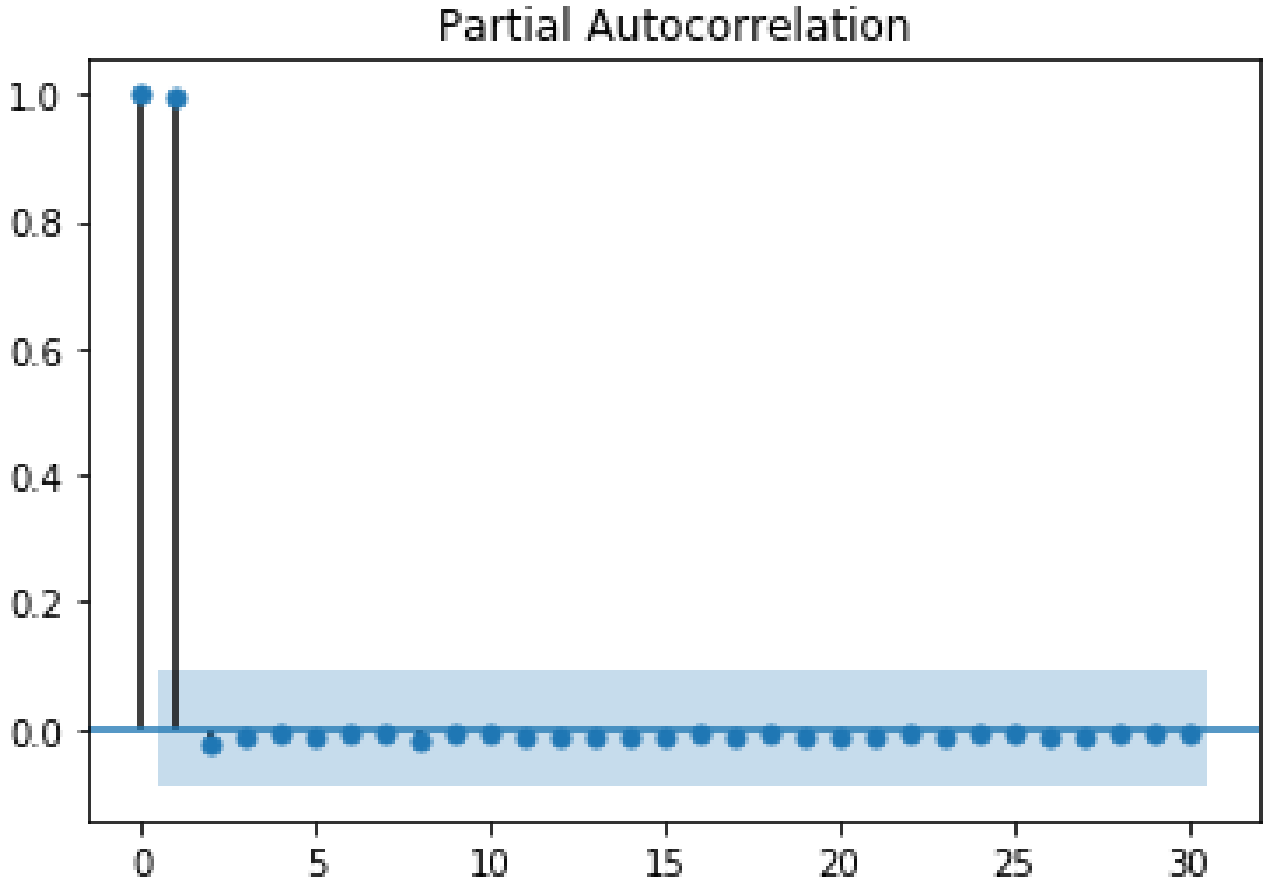
	acf	pacf	Q	p-val
0	0.995826	0.997905	478.982738	3.557116e-106
1	0.991576	-0.021451	954.879207	4.473163e-208
2	0.987276	-0.014817	1427.646393	2.951746e-309
3	0.982961	-0.006640	1897.273943	0.000000e+00
4	0.978601	-0.013615	2363.724870	0.000000e+00
..	...	...	...	...
115	0.338511	-0.012698	31551.567174	0.000000e+00
116	0.332064	-0.005880	31621.846035	0.000000e+00
117	0.325633	-0.006220	31689.615859	0.000000e+00
118	0.319213	-0.008097	31754.920323	0.000000e+00
119	0.312799	-0.010132	31817.800995	0.000000e+00

EXAMPLE (1)

- Determine the lag length:
  - Visual inspection: use PACF. The order of the AR is the same as the number of significant lags in the PACF.
  - It is not an easy to do so in practice.
  - Using Information Criteria: you may specify different models with different lag lengths and choose the model with the smallest information criterion.
  - Testing Down: use t statistic to choose the model with the highest significant lag length.

- Determine the lag length:
  - Visual inspection: use PACF. The order of the AR is the same as the number of significant lags in the PACF.

EXAMPLE (1)



EXAMPLE (1) – ESTIMATING AN AR(1) MODEL

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                    lpt   No. Observations:                  480
Model:                     ARMA(1, 0)   Log Likelihood                1860.957
Method:                       css-mle   S.D. of innovations              0.005
Date:                Wed, 10 Mar 2021   AIC                          -3715.914
Time:                        11:42:04   BIC                          -3703.393
Sample:                             0   HQIC                         -3710.992

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.9480        nan        nan        nan        nan         nan
ar.L1.lpt      1.0000        nan        nan        nan        nan         nan
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.0000           +0.0000j            1.0000            0.0000
-----------------------------------------------------------------------------
```

Example (1) – Estimating an AR(2) Model

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                    lpt   No. Observations:                  480
Model:                     ARMA(2, 0)   Log Likelihood                1360.826
Method:                       css-mle   S.D. of innovations              0.014
Date:                Wed, 10 Mar 2021   AIC                          -2713.653
Time:                        11:48:16   BIC                          -2696.958
Sample:                             0   HQIC                         -2707.090

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.9969        nan        nan        nan         nan         nan
ar.L1.lpt           0   6.87e-07          0      1.000   -1.35e-06    1.35e-06
ar.L2.lpt      1.0000   4.28e-09   2.34e+08      0.000       1.000       1.000
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.0000           +0.0000j            1.0000            0.0000
AR.2           -1.0000           +0.0000j            1.0000            0.5000
-----------------------------------------------------------------------------
```

EXAMPLE (1) – ESTIMATING AN AR(3) MODEL

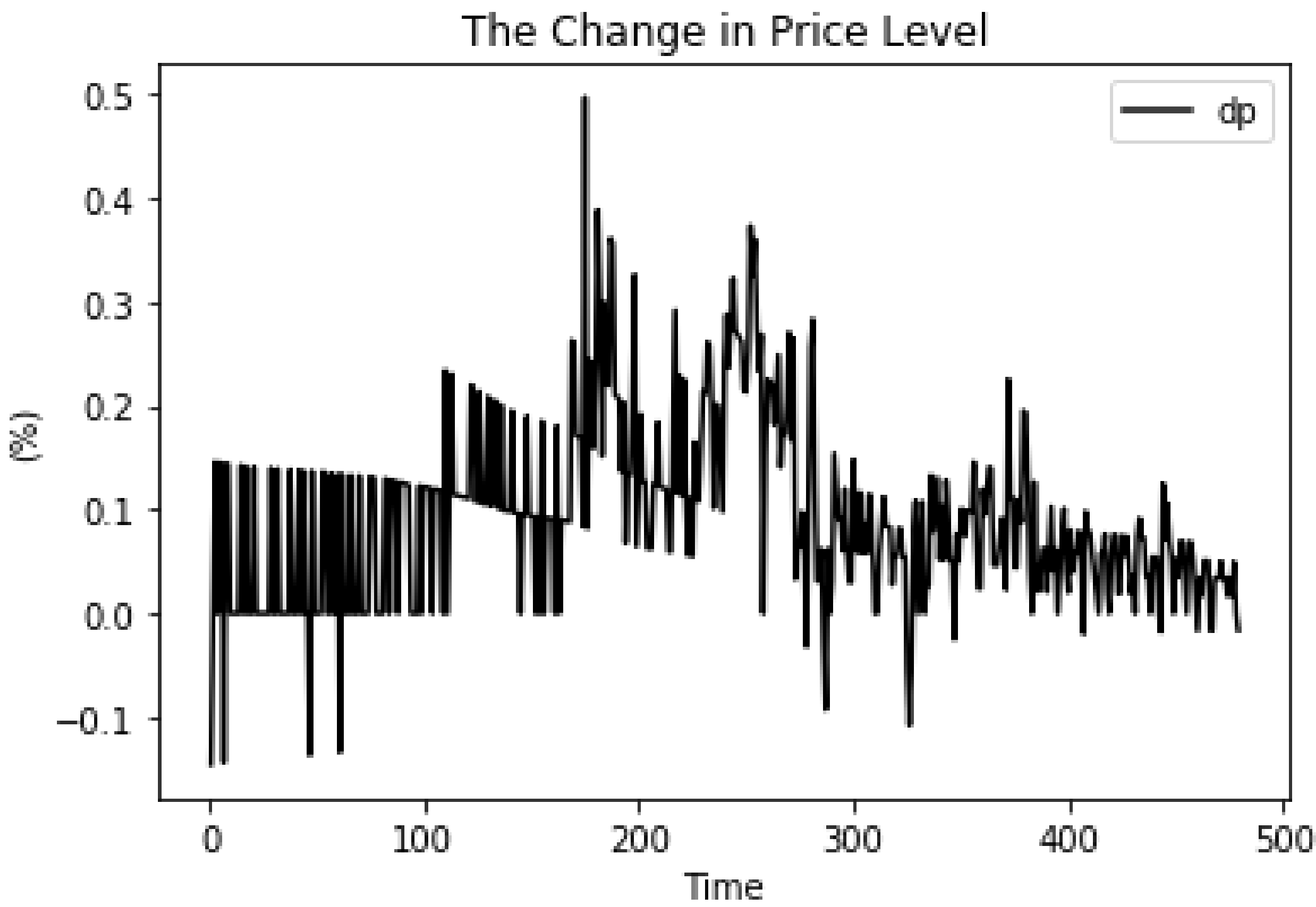
```
        ARMA Model Results
==============================================================================
Dep. Variable:                    lpt   No. Observations:                  480
Model:                     ARMA(3, 0)   Log Likelihood                2135.787
Method:                       css-mle   S.D. of innovations              0.003
Date:                Wed, 10 Mar 2021   AIC                          -4261.573
Time:                        11:48:17   BIC                          -4240.704
Sample:                             0   HQIC                         -4253.370

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         24.3715        nan        nan        nan         nan         nan
ar.L1.lpt      1.5600        nan        nan        nan         nan         nan
ar.L2.lpt     -0.1630        nan        nan        nan         nan         nan
ar.L3.lpt     -0.3970        nan        nan        nan         nan         nan
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.0000           +0.0000j            1.0000            0.0000
AR.2            1.0315           +0.0000j            1.0315            0.0000
AR.3           -2.4423           +0.0000j            2.4423            0.5000
-----------------------------------------------------------------------------
```

EXAMPLE (1)

- Checking the unit circle
  - The coefficients of all the AR models. They are all near or equal to 1.
  - This means the process is not stationary.
  - The AR roots are outside the unit circle.

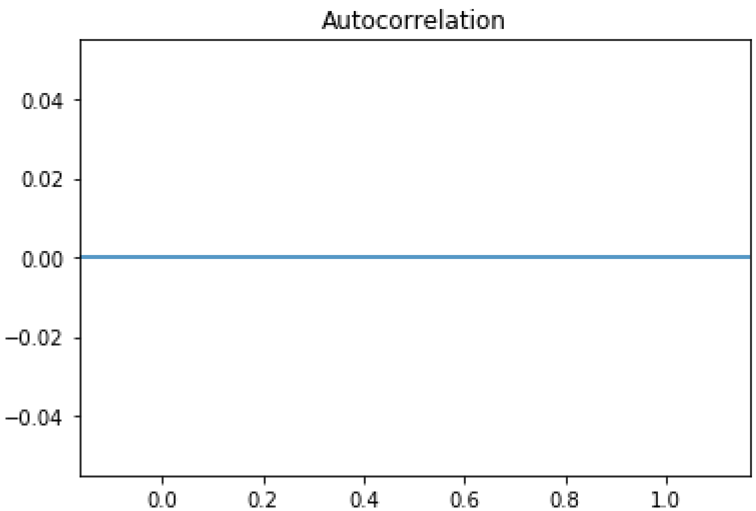
- The solution
  - Take the first difference of the data to make it stationary.
  - The process becomes an ‘integrated process’.
  - If the process is stationary after the first difference, then it is called ‘integrated of order 1’.
  - The order of integration is equal to the number of times the first difference has been taken to make it stationary.
  - ARIMA: The ‘I’ in the acronym refers to the Integration, the order of integration.

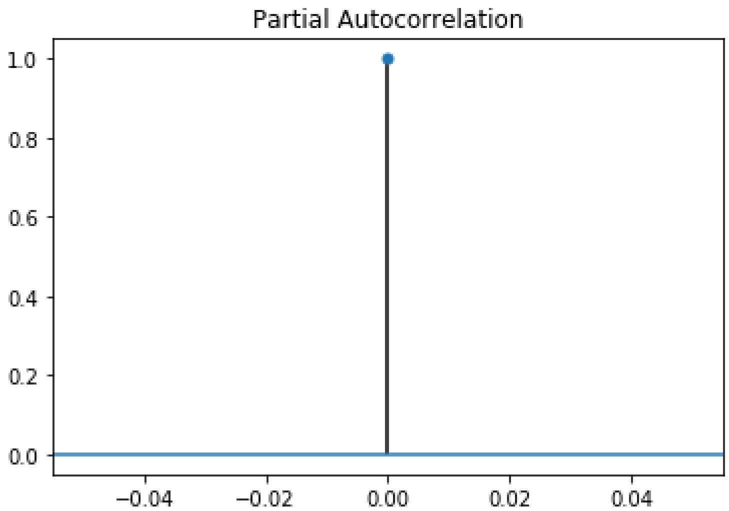
EXAMPLE (1)



The Change in Price Level
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
(%)
dp
0
100
200
300
400
500
Time

EXAMPLE (1)


Autocorrelation
0.04
0.02
0.00
−0.02
−0.04
0.0
0.2
0.4
0.6
0.8
1.0


Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.04
−0.02
0.00
0.02
0.04

Example (1)

```
   ARIMA Model Results
==============================================================================
Dep. Variable:                  D.lpt   No. Observations:                  479
Model:                 ARIMA(1, 1, 0)   Log Likelihood                2142.250
Method:                       css-mle   S.D. of innovations              0.003
Date:                Wed, 10 Mar 2021   AIC                          -4278.501
Time:                        12:17:37   BIC                          -4265.986
Sample:                             1   HQIC                         -4273.581

===============================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------
const           0.0036      0.000     12.379      0.000       0.003       0.004
ar.L1.D.lpt     0.5648      0.038     14.847      0.000       0.490       0.639
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.7706           +0.0000j            1.7706            0.0000
-----------------------------------------------------------------------------
```

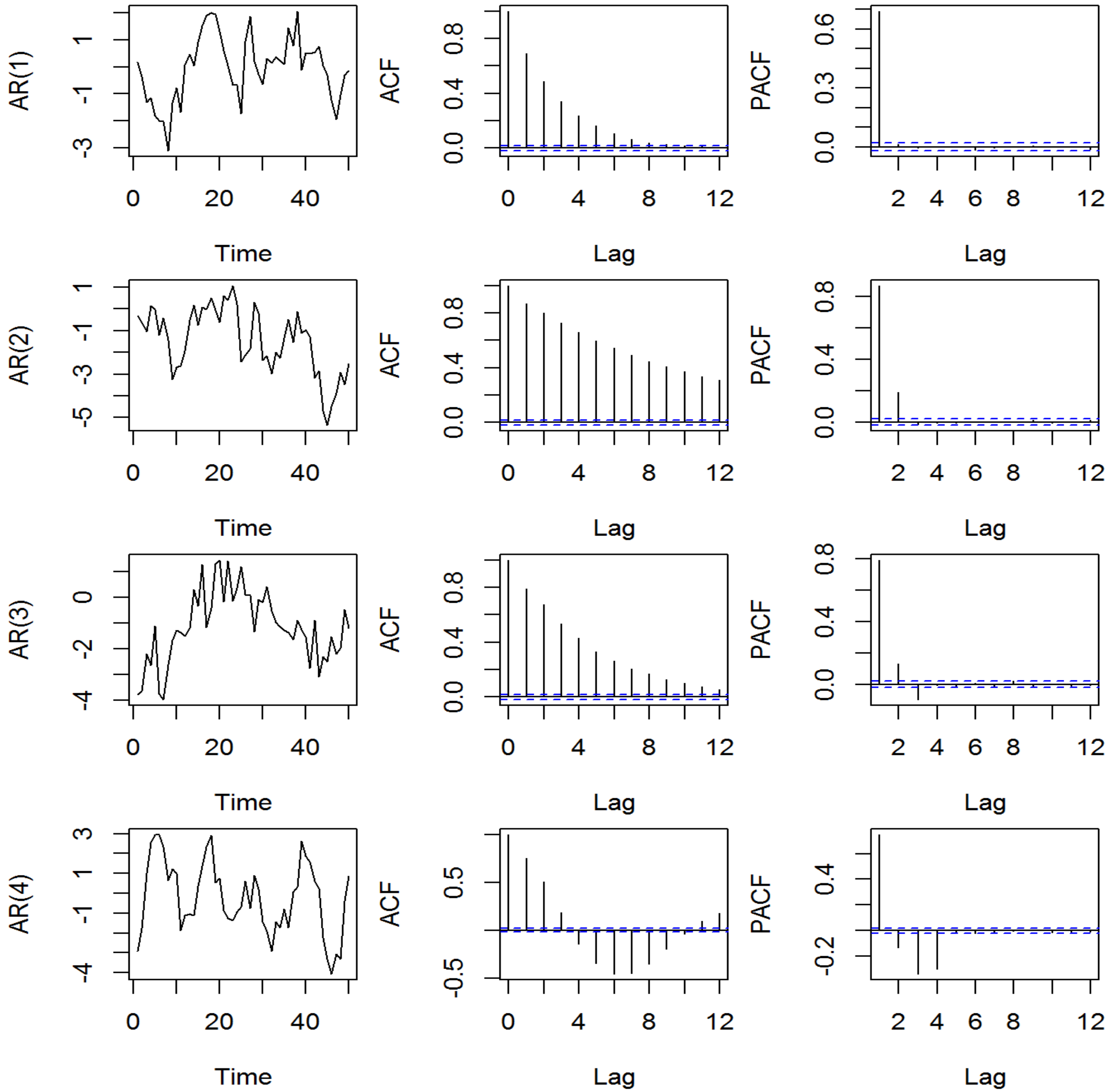
Example (1)

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                  D.lpt   No. Observations:                  479
Model:                 ARIMA(2, 1, 0)   Log Likelihood                2162.523
Method:                       css-mle   S.D. of innovations              0.003
Date:                Wed, 10 Mar 2021   AIC                          -4317.045
Time:                        12:17:38   BIC                          -4300.359
Sample:                             1   HQIC                         -4310.486

===============================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------
const           0.0035      0.000      9.095      0.000       0.003       0.004
ar.L1.D.lpt     0.4047      0.044      9.207      0.000       0.319       0.491
ar.L2.D.lpt     0.2867      0.044      6.508      0.000       0.200       0.373
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.2908           +0.0000j            1.2908            0.0000
AR.2           -2.7026           +0.0000j            2.7026            0.5000
-----------------------------------------------------------------------------
```

EXAMPLE (1)

```
                                    ARIMA Model Results
==============================================================================
Dep. Variable:                  D.lpt   No. Observations:                  479
Model:                 ARIMA(3, 1, 0)   Log Likelihood                2170.448
Method:                       css-mle   S.D. of innovations              0.003
Date:                Wed, 10 Mar 2021   AIC                          -4330.896
Time:                        12:17:38   BIC                          -4310.037
Sample:                             1   HQIC                         -4322.696

================================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
const           0.0035      0.000      7.528      0.000       0.003       0.004
ar.L1.D.lpt     0.3513      0.045      7.768      0.000       0.263       0.440
ar.L2.D.lpt     0.2143      0.047      4.568      0.000       0.122       0.306
ar.L3.D.lpt     0.1820      0.045      4.015      0.000       0.093       0.271
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.1725           -0.0000j            1.1725           -0.0000
AR.2           -1.1751           -1.8180j            2.1647           -0.3413
AR.3           -1.1751           +1.8180j            2.1647            0.3413
-----------------------------------------------------------------------------
```

Properties of AR Process: ACF



Modelling Stationary Data

- Moving Average, MA
  - The model can be explained by past and current unobserved shocks
  - The general form is called MA(q) model
  - The error term, *u*, is assumed to be iid with zero mean and constant variance.

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q}$$

  - MA(1)

$$y_t = \mu + u_t + \theta_1 u_{t-1}$$

  - The choice of lags follows the same pattern as with AR process

Properties of Stationary MA(1)

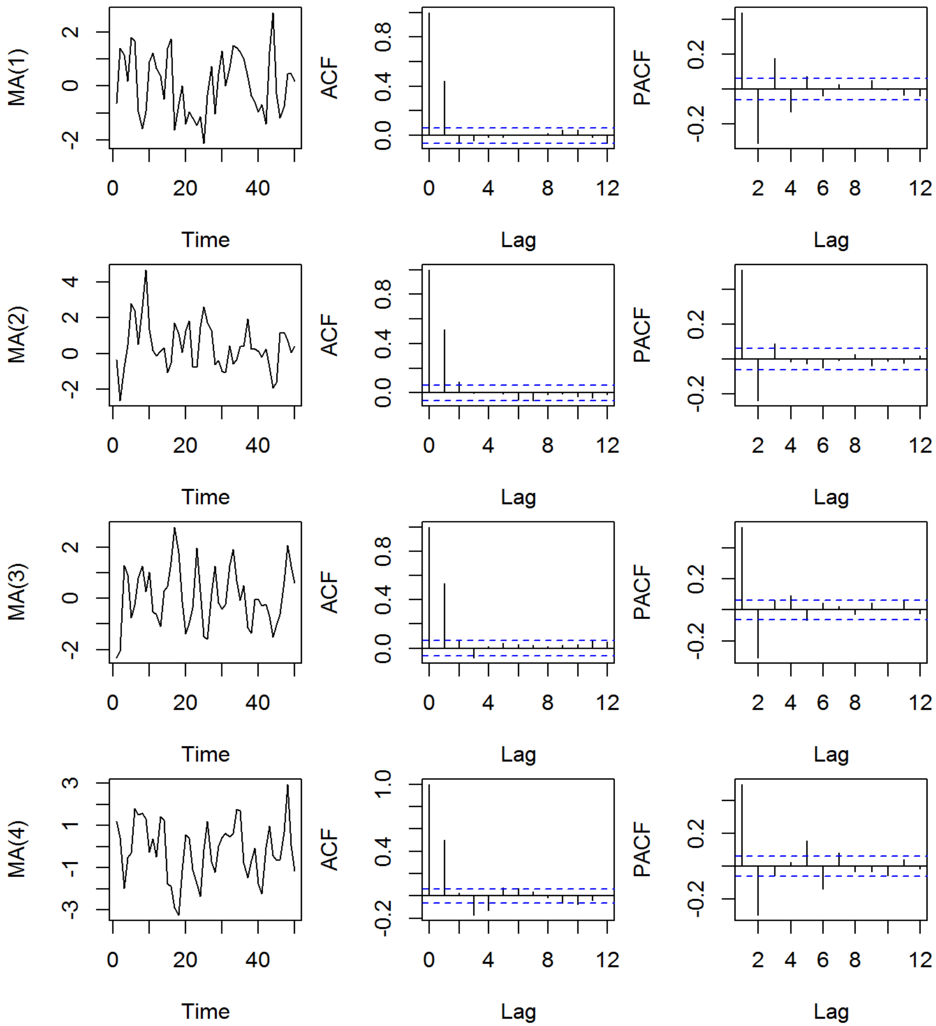
- The statistical properties

$$E(y_t) = \mu$$
$$var(y_t) = (1 - \theta_1^2)\sigma^2$$
$$cov(y_t, y_{t-1}) = \theta_1\sigma^2$$
$$cov\left(y_t, y_{t-j}\right) = 0, \text{ for j>1.}$$

The last part implies that the process converges back to the long run as soon as the lag goes beyond 1.

In general, MA*(q)* has an ACF that dies off beyond the *q*-th lag.

ACF of an MA Process



Example: MA(1)

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                  D.lpt   No. Observations:                  479
Model:                 ARIMA(0, 1, 1)   Log Likelihood                2106.084
Method:                       css-mle   S.D. of innovations              0.003
Date:                Wed, 10 Mar 2021   AIC                          -4206.168
Time:                        12:34:15   BIC                          -4193.653
Sample:                             1   HQIC                         -4201.248

===============================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------
const           0.0036      0.000     19.031      0.000       0.003       0.004
ma.L1.D.lpt     0.3929      0.035     11.220      0.000       0.324       0.462
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
MA.1           -2.5449           +0.0000j            2.5449            0.5000
-----------------------------------------------------------------------------
```

EXAMPLE: MA(2)

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                  D.lpt   No. Observations:                  479
Model:                 ARIMA(0, 1, 2)   Log Likelihood                2128.755
Method:                       css-mle   S.D. of innovations              0.003
Date:                Wed, 10 Mar 2021   AIC                          -4249.510
Time:                        12:34:15   BIC                          -4232.823
Sample:                             1   HQIC                         -4242.950

===============================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------
const           0.0036      0.000     16.509      0.000       0.003       0.004
ma.L1.D.lpt     0.4011      0.046      8.652      0.000       0.310       0.492
ma.L2.D.lpt     0.2784      0.039      7.077      0.000       0.201       0.356
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
MA.1           -0.7202           -1.7530j            1.8952           -0.3120
MA.2           -0.7202           +1.7530j            1.8952            0.3120
-----------------------------------------------------------------------------
```

EXAMPLE: MA(3)

```
                              ARIMA Model Results
==============================================================================
Dep. Variable:                  D.lpt   No. Observations:                  479
Model:                 ARIMA(0, 1, 3)   Log Likelihood                2142.834
Method:                       css-mle   S.D. of innovations              0.003
Date:                Wed, 10 Mar 2021   AIC                          -4275.668
Time:                        12:34:15   BIC                          -4254.809
Sample:                             1   HQIC                         -4267.468

===============================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------
const           0.0036      0.000     14.726      0.000       0.003       0.004
ma.L1.D.lpt     0.3965      0.046      8.565      0.000       0.306       0.487
ma.L2.D.lpt     0.3059      0.041      7.455      0.000       0.225       0.386
ma.L3.D.lpt     0.2351      0.044      5.345      0.000       0.149       0.321
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
MA.1            0.2187           -1.5488j            1.5642           -0.2277
MA.2            0.2187           +1.5488j            1.5642            0.2277
MA.3           -1.7385           -0.0000j            1.7385           -0.5000
-----------------------------------------------------------------------------
```

ARIMA MODELS

- Autoregressive Integrated Moving Average, ARIMA
  - The model can be explained by both AR and MA
  - The general form is called ARMA(p, q) model
  - The error term, *u*, is assumed to be iid with zero mean and constant variance.

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots + \theta_q u_{t-q} + u_t$$

  - ARMA(1,1)

$$y_t = \mu + \phi_1 y_{t\text{-}1} + u_t + \theta_1 u_{t\text{-}1}$$

  - The choice of lags follows the same pattern as with AR and MA process

ACF OF AN ARMA PROCESS

ACF and PACF of an ARMA(1,1)

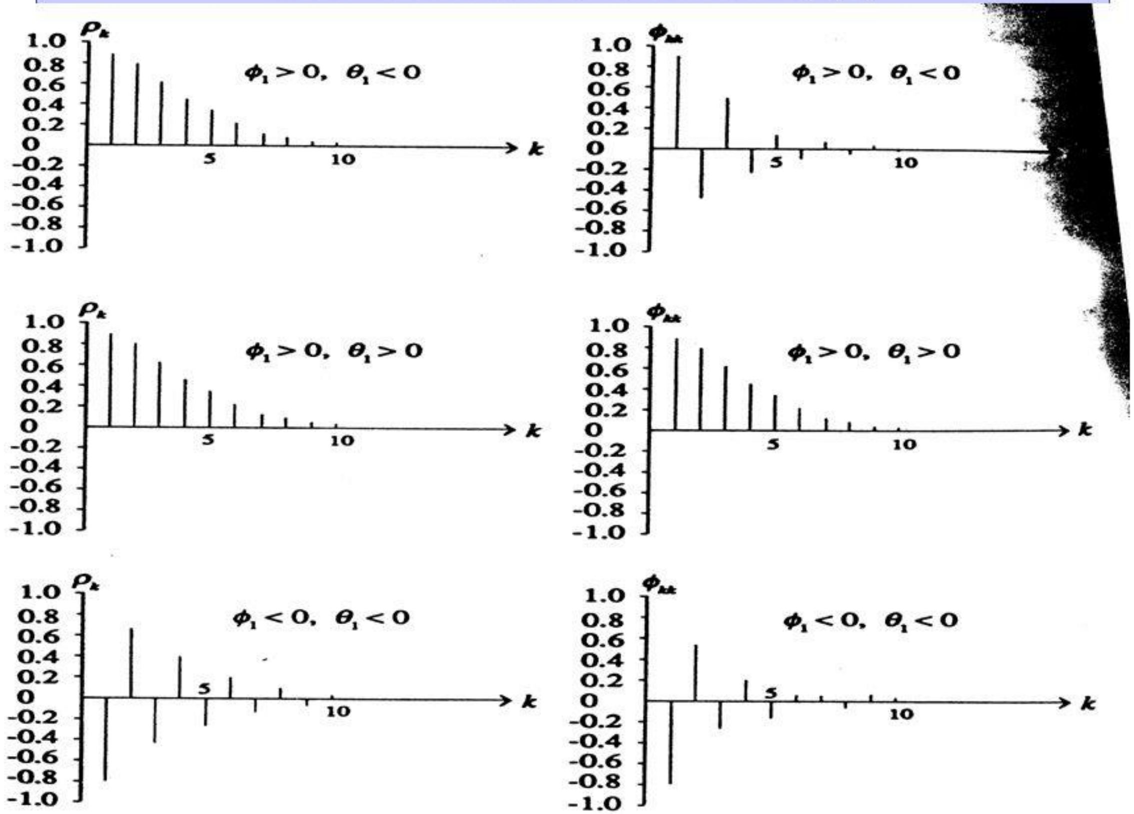


Fig. 3.14 ACF and PACF of ARMA(1, 1) model $(1 - \phi_1 B)\dot{Z}_t = (1 - \theta_1 B)a_t$.

ARIMA(1,1,1) Regression Output

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                  D.lpt   No. Observations:                  479
Model:                 ARIMA(1, 1, 1)   Log Likelihood                2186.521
Method:                       css-mle   S.D. of innovations              0.003
Date:                Wed, 10 Mar 2021   AIC                          -4365.043
Time:                        12:45:16   BIC                          -4348.356
Sample:                             1   HQIC                         -4358.483

===============================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------
const           0.0032      0.001      3.144      0.002       0.001       0.005
ar.L1.D.lpt     0.9775      0.012     83.377      0.000       0.955       1.000
ma.L1.D.lpt    -0.7898      0.035    -22.433      0.000      -0.859      -0.721
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.0230           +0.0000j            1.0230            0.0000
MA.1            1.2661           +0.0000j            1.2661            0.0000
-----------------------------------------------------------------------------
```

THE BEST MODEL?

- So far, we estimated three models using differenced data.
- This shows that the model should be defined as ARIMA and not as ARMA since the data in levels cannot be used (i.e. the time series is not stationary).
- The natural questions now is which model of the seven ARIMA models is best to use?
- We need a set of criteria to discriminate between all variables:

THE SET OF CRITERIA: PRINCIPLES

- **Information Criteria**: choose the model that produces the smallest information criteria.
  - The chosen model does not necessarily produce serially uncorrelated residuals.

- **White noise residuals:** Once the model is chosen, residuals must be uncorrelated. If the model fails to produce white noise residuals, we choose the second best from the information criteria stage.

- **Parsimony Principle:** In case two or models satisfy the second criterion, we choose the model with least number of parameters.

EXAMPLE (3)

Model	AIC	BIC	Q stat
ARIMA(1,1,0)	-4278.501	-4265.986	112.79***
ARIMA(2,1,0)	-4317.045	-4300.359	46.88***
ARIMA(3,1,0)	-4330.896	-4310.037	34.18***
ARIMA(0,1,1)	-4206.168	-4193.653	520.09***
ARIMA(0,1,2)	-4249.510	-4232.823	291.67***
ARIMA(0,1,3)	-4275.668	-4254.809	182.25***
ARIMA(1,1,1)	-4365.043	-4348.356	28.45***

Box – Jenkins Methodology

- It is an algorithm to aid analysing ARIMA models.
- Now we bring together all the elements we learned so far to combine them into a set of steps to build the model that is best fit the data.
- The purpose is forecasting. For forecasting, we require:
  - A parsimonious model
  - White noise residuals.

Box – Jenkins Methodology

- Step 1: (**Check for stationarity**).
  - Use formal tests such as unit root tests (note: these will be covered in Week 11).
  - If data are not stationary, take the first difference till the data is stationary.
  - Determine the degree of integration, *I*.
- Step 2: (**Identification**)
  - Use ACF and PACF to identify the ARIMA process.
  - Identify several models.
- Step 3: **(Estimation)**
  - Save key statistics and verify significance of parameters.
- Step 4: **(Model Selection)**
  - Use information criteria to select the lag length and the most suitable model.
- Step 5: **(Diagnostic Checking)**
  - Check for white noise residuals. If the model does not produce a white noise residuals, go back to Step 2.
- Step 6: **(Forecasting)**

THANK YOU