Foundations of Data Science
Lecture 9, Module 2
Fall 2022

Rumi Chunara, PhD

CS6053

<u>Fine Print</u>: these slides are, and always will be a work in progress. The material presented herein is original, inspired, or borrowed from others' work. Where possible, attribution and acknowledgement will be made to content's original source. Do not distribute without the instructor's permission.

ARIMA: Putting the pieces together

- Autoregressive model of order p: AR(p)
- Moving average model of order q: MA(q)
- ARMA(p,q)

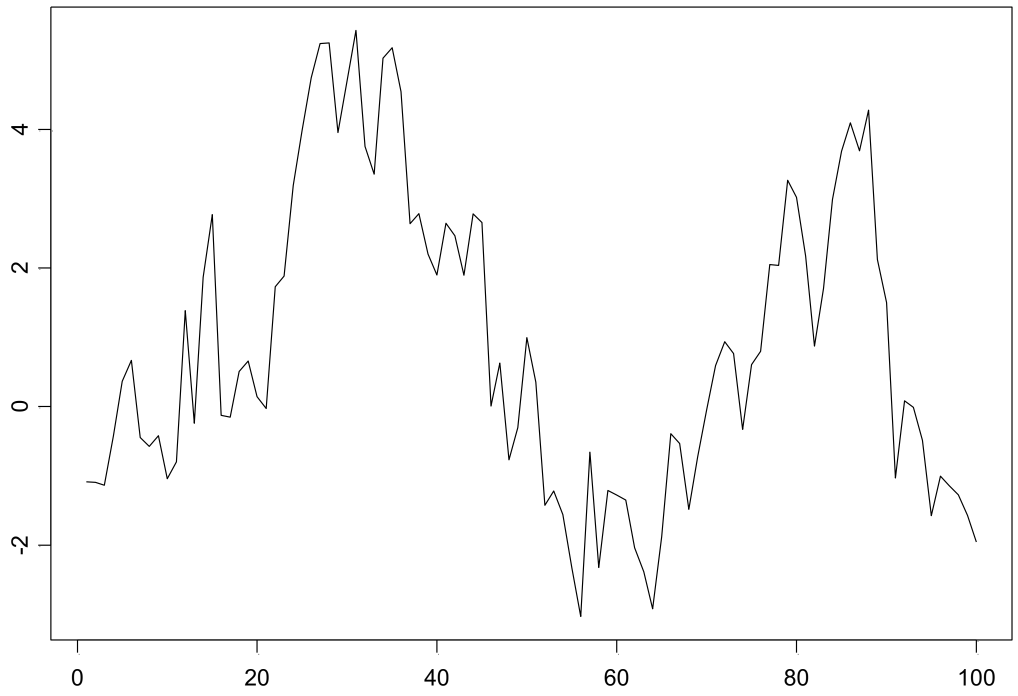
ARIMA: Putting the pieces together

- Autoregressive model of order p: AR(p)

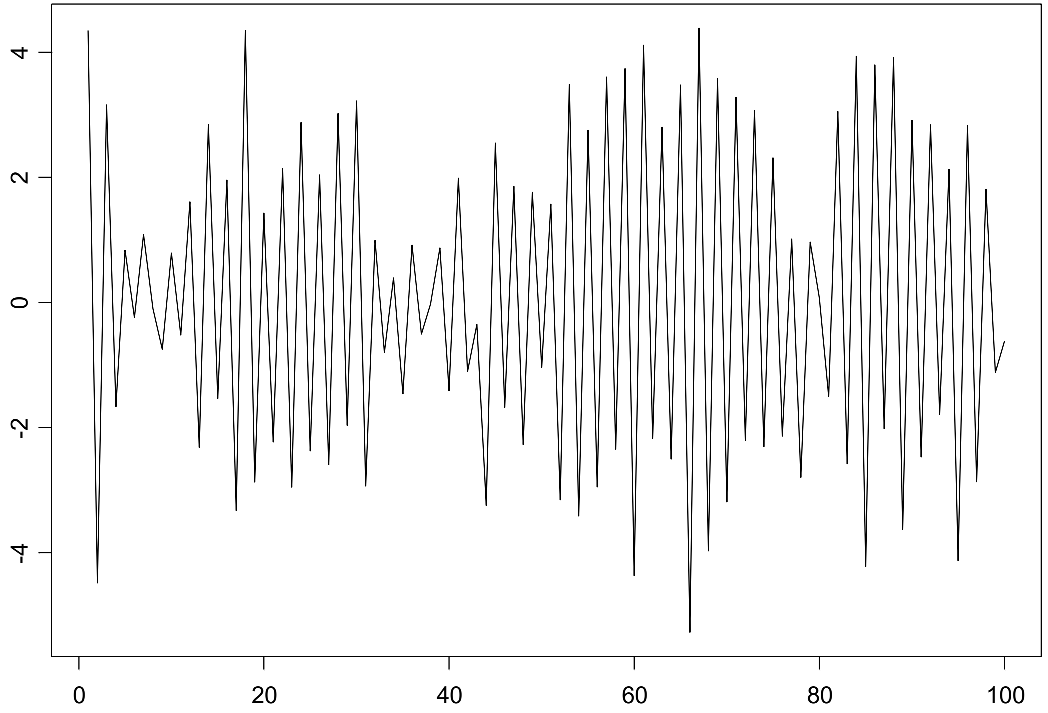
$$\boxed{x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + .. + \phi_p x_{t-p} + w_t}$$

- Moving average model of order q: MA(q)
- ARMA(p,q)

AR(1), $\phi = 0.9$



AR(1), $\phi = -0.9$



ARIMA: Putting the pieces together

- Autoregressive model of order p: AR(p)

$$\boxed{x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + .. + \phi_p x_{t-p} + w_t}$$

- Moving average model of order q: MA(q)

$$\boxed{x_t = \theta_1 w_{t-1} + \theta_2 w_{t-2} + .. + \theta_q w_{t-q} + w_t}$$

- ARMA(p,q)

ARIMA: Putting the pieces together

- Autoregressive model of order p: AR(p)

$$\boxed{x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + .. + \phi_p x_{t-p} + w_t}$$

- Moving average model of order q: MA(q)

$$\boxed{x_t = \theta_1 w_{t-1} + \theta_2 w_{t-2} + .. + \theta_q w_{t-q} + w_t}$$

- ARMA(p,q)
  - A time series is ARMA(p,q) if it is stationary and

$$\boxed{\begin{gathered} x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + .. + \phi_p x_{t-p} + w_t + \\ \theta_1 w_{t-1} + \theta_2 w_{t-2} + .. + \theta_q w_{t-q} \end{gathered}}$$

AR Process

- Start with AR(1) sequence

$$\rho(h) - \phi\rho(h-1) = 0, \quad h = 1,2,\ldots$$

- This means

$$\begin{aligned}
&\rho(1) = \phi\rho(0) \\
&\rho(2) = \phi\rho(1) = \phi^2\rho(0) \\
&\ldots \\
&\rho(n) = \phi\rho(n-1) = \phi^n\rho(0)
\end{aligned}$$

- Which we can solve given roots z_i

$$\rho_n = \phi^n\rho(0) = (z_0^{-1})^n\rho(0)$$

ARIMA (AutoRegressive Integrated Moving Average)

- ARMA only applies to stationary process
- Apply differencing to obtain stationarity
  - Replace its value by incremental change from last value

Differenced	x1	x2	x3	x4
1 time		x2-x1’	x3’-x2’	x4’-x3’
2 times			x3’-2x2’+x1’	x4’-2x3’+x2’

- A process x_t is ARIMA(p,d,q) if
  - AR(p)
  - MA(q)
  - Differenced d times
- The Box Jenkins method applies ARMA/ARIMA models to find the best fit of a time series model to past values of a time series

Model Building For ARIMA time series

Consists of three steps

1. Identification
2. Estimation
3. Diagnostic checking

Checking Stationarity

- An important step before fitting an ARIMA function is to make sure the timeseries is stationary.
- To do this, there are many ways to remove trend and seasonality. One easy approach is to use the stats package to remove seasonality, and use diff to remove a trend.

Identification

- The first step is to identify the model. Identification consists of specifying the appropriate structure (AR, MA or ARMA) and order of model.
- Identification is sometimes done by looking at plots of the acf and partial autocorrelation function (pacf).
- Sometimes identification is done by an automated iterative procedure; fitting many different possible model structures and orders and using a goodness-of fit statistic to select the best model.

Identification by visual inspection

- The classical method of model identification as described by Box and Jenkins is judge the appropriate model structure and order from the appearance of the plotted acf and pacf.
- Reminder: the acf at lag k measures the correlation of the series with itself lagged k time-points.
- The pacf at lag k is the autocorrelation at lag k after first removing autocorrelation with a AR(k-1) model
- If the AR (k-1) model effectively whitens the time series, the pacf at lag k will be zero.

Diagnostic Pattern Summary

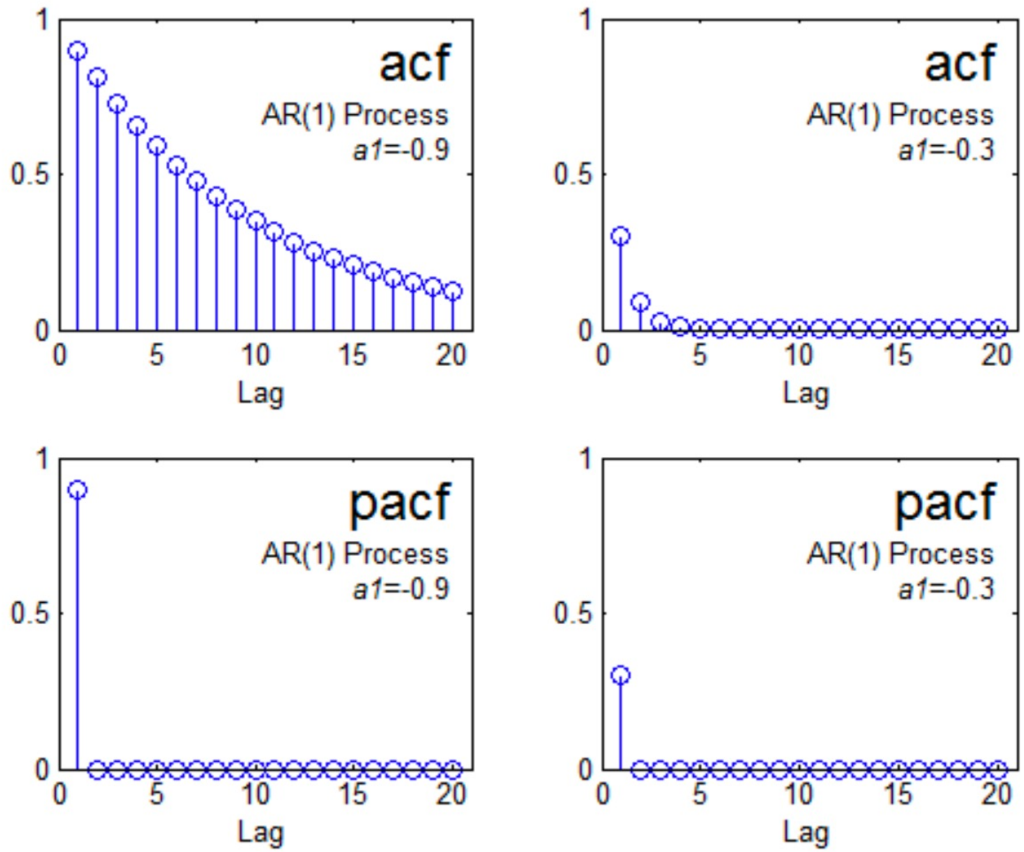
- The diagnostic patterns of acf and pacf for an AR(1) model are:
  - Acf: declines in geometric progression from its highest value at lag 1
  - Pacf: cuts off abruptly after lag 1
- The opposite types of patterns apply to an MA(1) process:
  - Acf: cuts off abruptly after lag 1
  - Pacf: declines in geometric progression from its highest value at lag 1

Hyndman, Rob J; Athanasopoulos, George. ["Forecasting: principles and practice"]

Identification by visual inspection (cont.)

- The identification of ARMA models from the acf and pacf plots is difficult and requires much experience for all but the simplest models.
- Let’s look at the diagnostic patterns for the two simplest models: AR(1) and MA(1).

- Theoretical acf and pacf for two AR(1) processes with large and small autoregressive coefficients are shown below.



Rules of Thumb to determine q and p

- **Rule 1:** If the ACF shows exponential decay, the PACF has a spike at lag 1, and no correlation for other lags, then use one autoregressive (p) parameter
- **Rule 2:** If the ACF shows a sine-wave shape pattern or a set of exponential decays, the PACF has spikes at lags 1 and 2, and no correlation for other lags, the use two autoregressive (p) parameters.
- **Rule 3:** If the ACF has a spike at lag 1, no correlation for other lags, and the PACF damps out exponentially, then use one moving average (q) parameter.
- **Rule 4:** If the ACF has spikes at lags 1 and 2, no correlation for other lags, and the PACF has a sine-wave shape pattern or a set of exponential decays, then use two moving average (q) parameter.
- **Rule 5:** If the ACF shows exponential decay starting at lag 1, and the PACF shows exponential decay starting at lag 1, then use one autoregressive (p) and one moving average (q) parameter.

ACF For Model Identification

Shape	Indicated Model
Exponential, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to identify the order of the autoregressive model.
Alternating positive and negative, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to help identify the order.
One or more spikes, rest are essentially zero	Moving average model, order identified by where plot becomes zero.
Decay, starting after a few lags	Mixed autoregressive and moving average (ARMA) model.
All zero or close to zero	Data are essentially random.
High values at fixed intervals	Include seasonal autoregressive term.
No decay to zero	Series is not stationary.

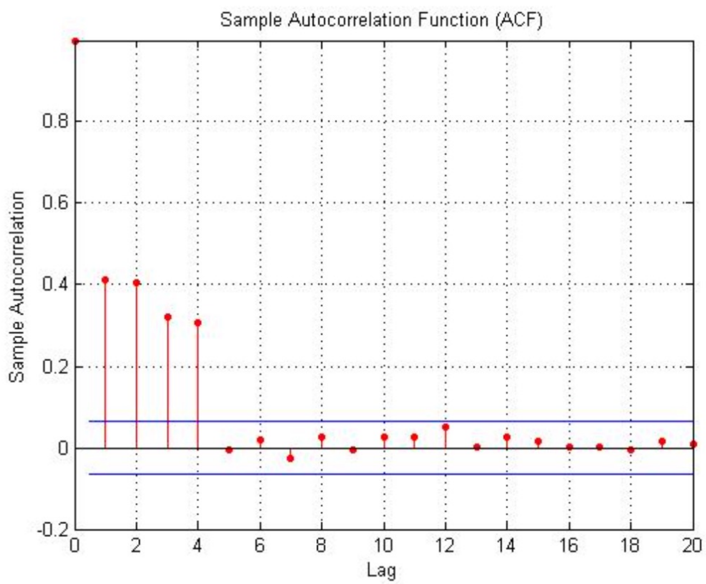
(From Wikipedia)

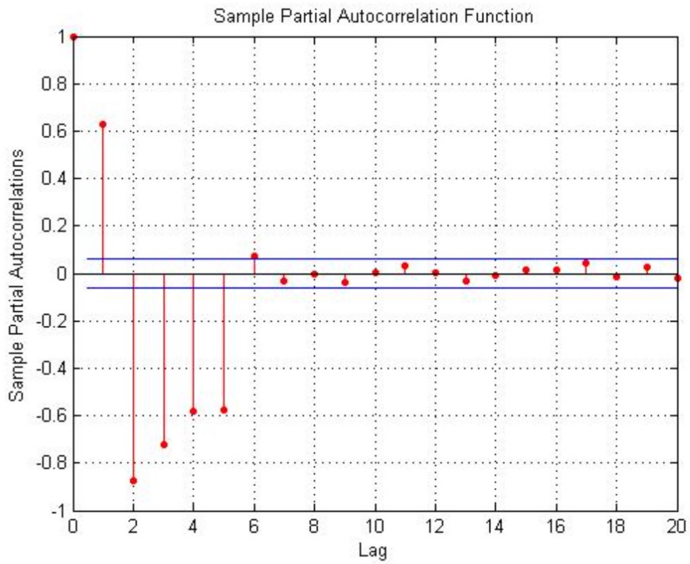
Summary of ACF/PACF

Process	ACF	PACF
AR(p)	Tails off as exponential decay or damped sine wave	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off as exponential decay or damped sine wave
ARMA(p,q)	Tails off after lag (q-p)	Tails off after lag (p-q)

Estimating AR and MA order

- Example ACF plots suggesting AR(5) and MA(4) values, left and right respectively, are shown below.





Estimation

- The second step is to estimate the coefficients of the model. Coefficients of AR models can be estimated by least-squares regression.
- Estimation of parameters of MA and ARMA models usually requires a more complicated iteration procedure.
- In practice, estimation is fairly transparent to the user, as it accomplished automatically by a computer program with little or no user interaction

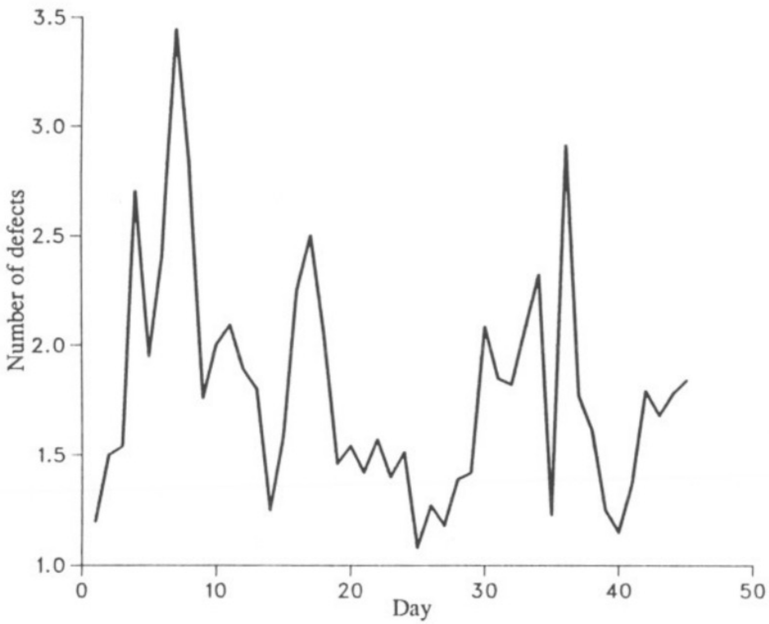
Diagnostic Checking

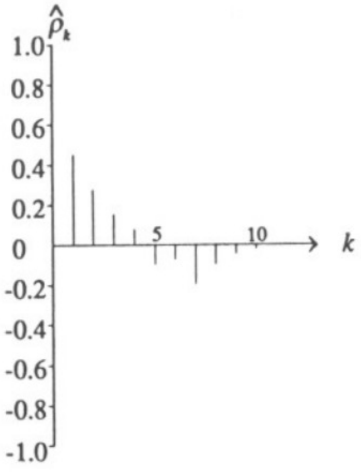
- The third step is to check the model. This step is also called diagnostic checking, or verification.
- A key question in ARMA modeling is does the model effectively describe the persistence? If so, the model residuals should be random –or uncorrelated in time – and the autocorrelation function (acf) of residuals should be zero at all lags except lag zero.
- Of course, for sample series, the acf will not be exactly zero, but should fluctuate close to zero

Diagnostic Checking II

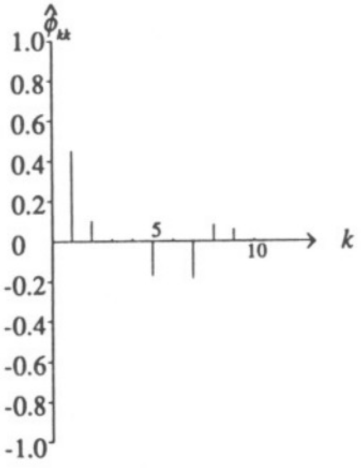
- Usually the fitting process is guided by the principal of parsimony, by which the best model is the simplest possible model – the model with the fewest parameters -- that adequately describes the data.

Example

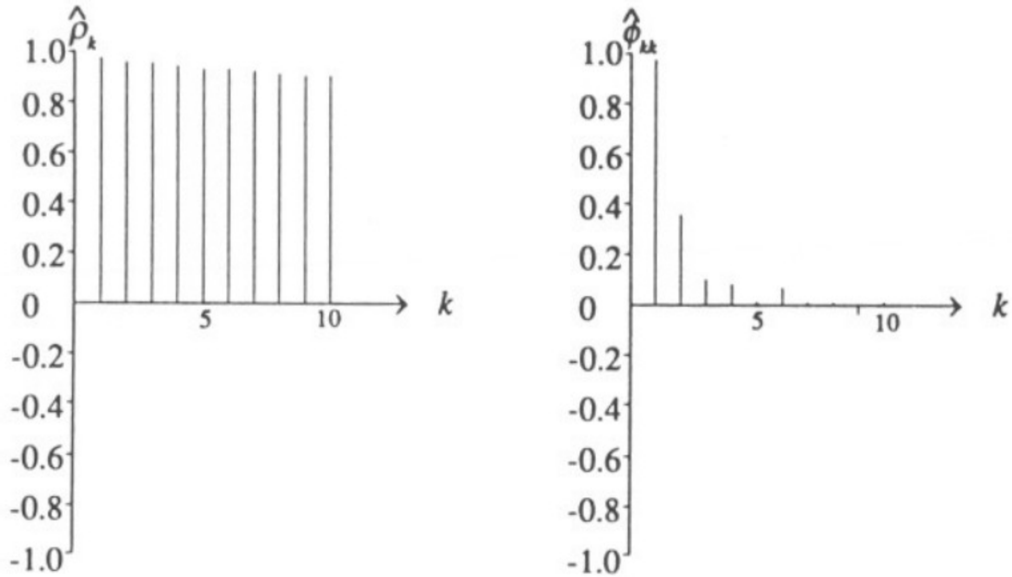




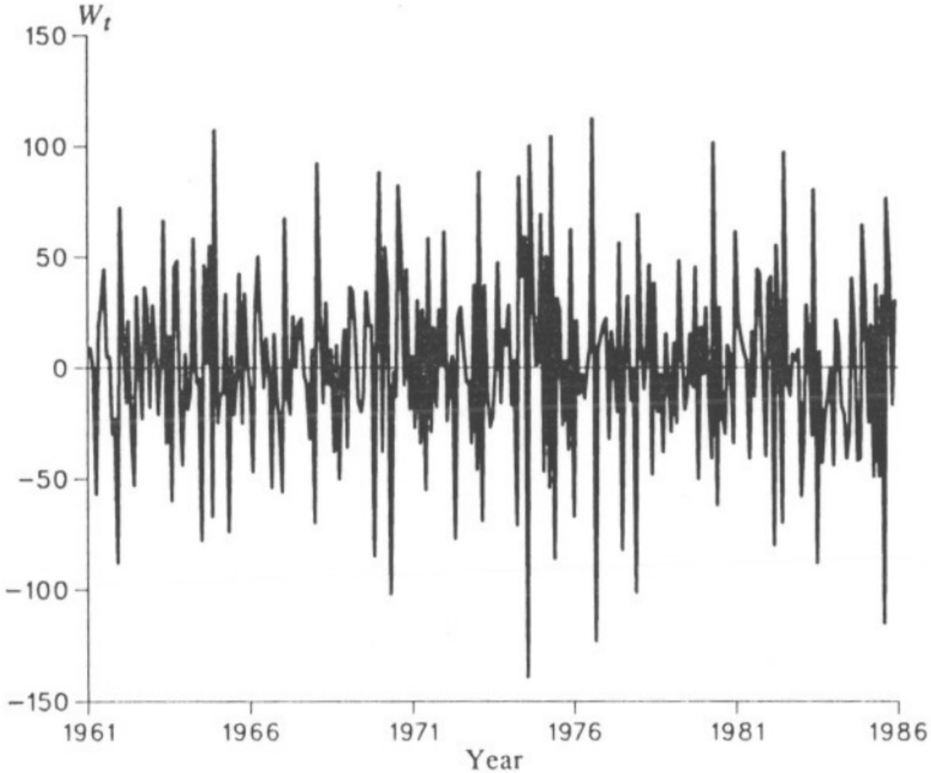
ACF



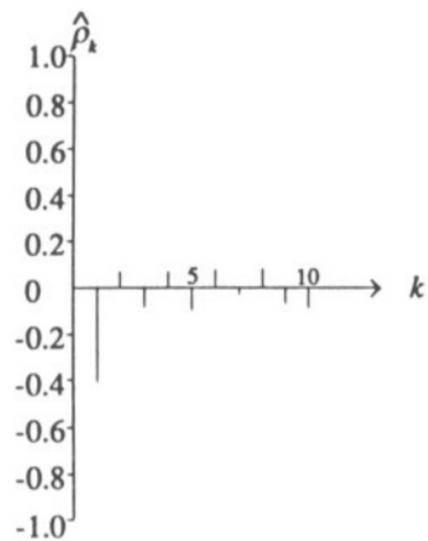
PACF



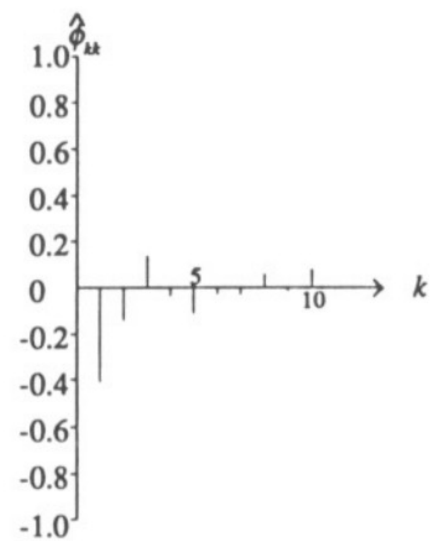
$\hat{\rho}_k$
1.0
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1.0
5
10
k
$\hat{\phi}_{kk}$
1.0
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1.0
5
10
k



W_t
150
100
50
0
−50
−100
−150
1961
1966
1971
1976
1981
1986
Year



$\hat{\rho}_k$
1.0
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1.0
5
10
k



$\hat{\phi}_{kk}$
1.0
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1.0
5
10
k