INDR 450/550

Spring 2022

Lecture 6: ARIMA processes

March 2, 2022

Fikri Karaesmen

Announcements

- The first lab video was uploaded on Friday. The second lab will be uploaded on Friday this week. Please follow them.
- First HW will be assigned this week
  - You can work in groups of two or three for the homeworks
- Looking forward to seeing you on campus next week.

Auto-Correlation

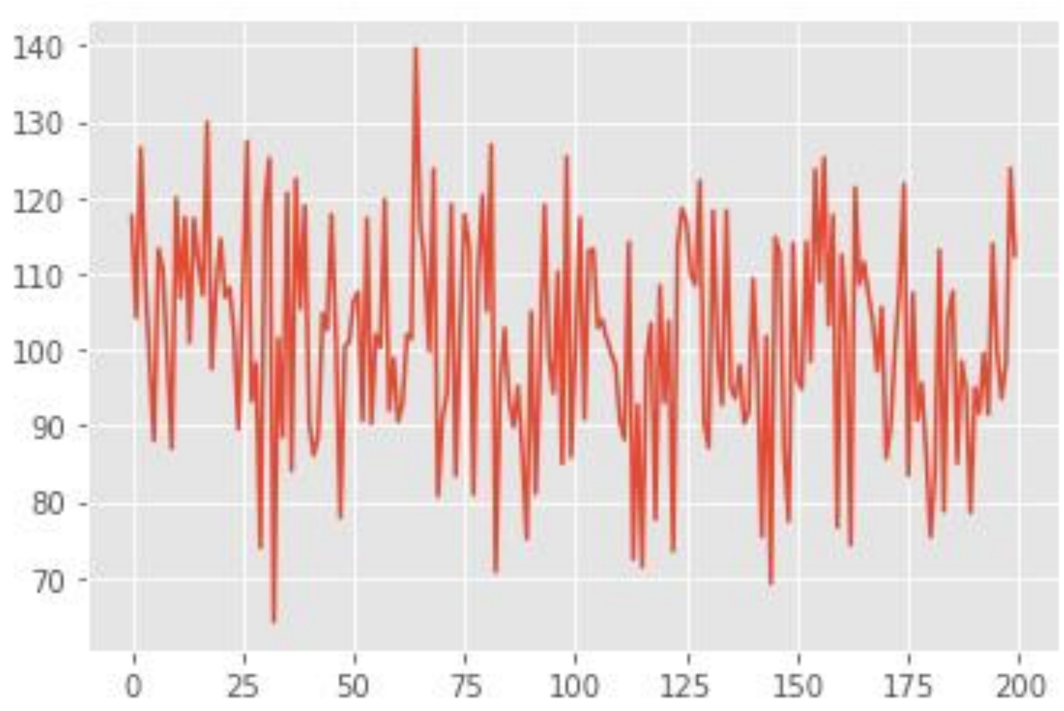
- Sample estimators:
  - For our purposes, we will be interested in the auto-correlation of the process that generates demand: for instance $Corr(Y_t, Y_{t-1})$ or $Corr(Y_t, Y_{t-k})$. This looks at the correlation between demand observation separated by k periods (how demand from k periods ago affects the demand today).
  - Note our paired observations are (y_1, y_{1+k}), (y_2, y_{2+k}),..., (y_{n-k}, y_n).
  - The k-lag autocorrelation can then be estimated by:

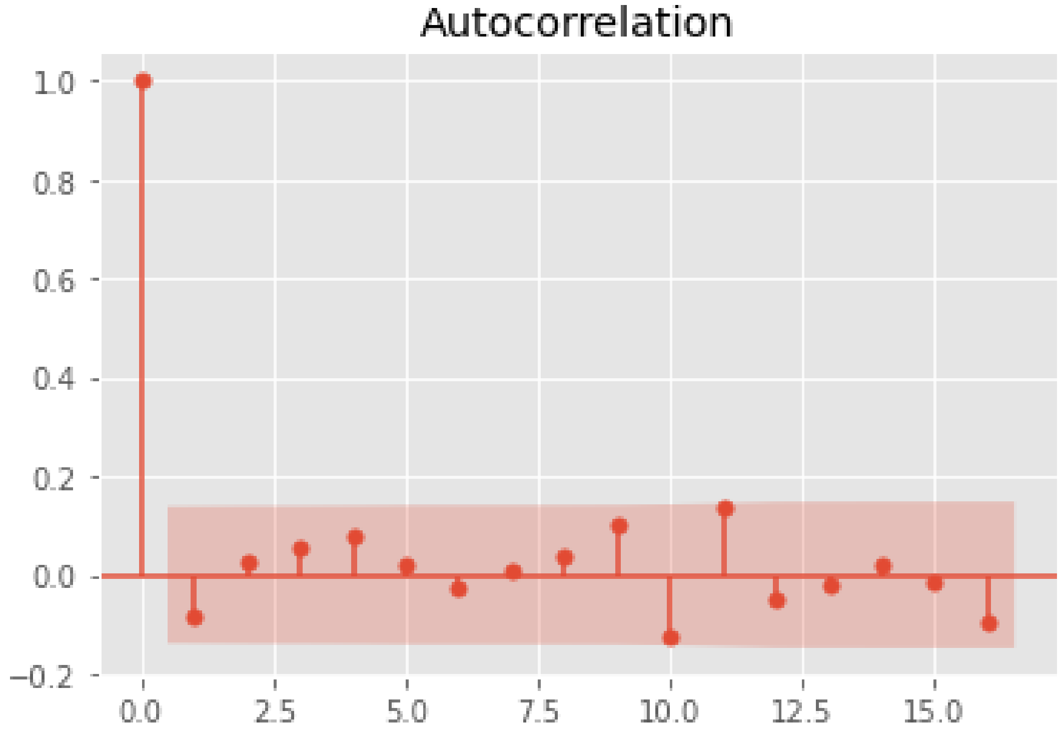
$$r_k = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

Auto-Correlation: stationary i.i.d demand

- Recall the simple model: $Y_t = c + \varepsilon_t$. Here's the autocorrelation structure:

The data (generated randomly)





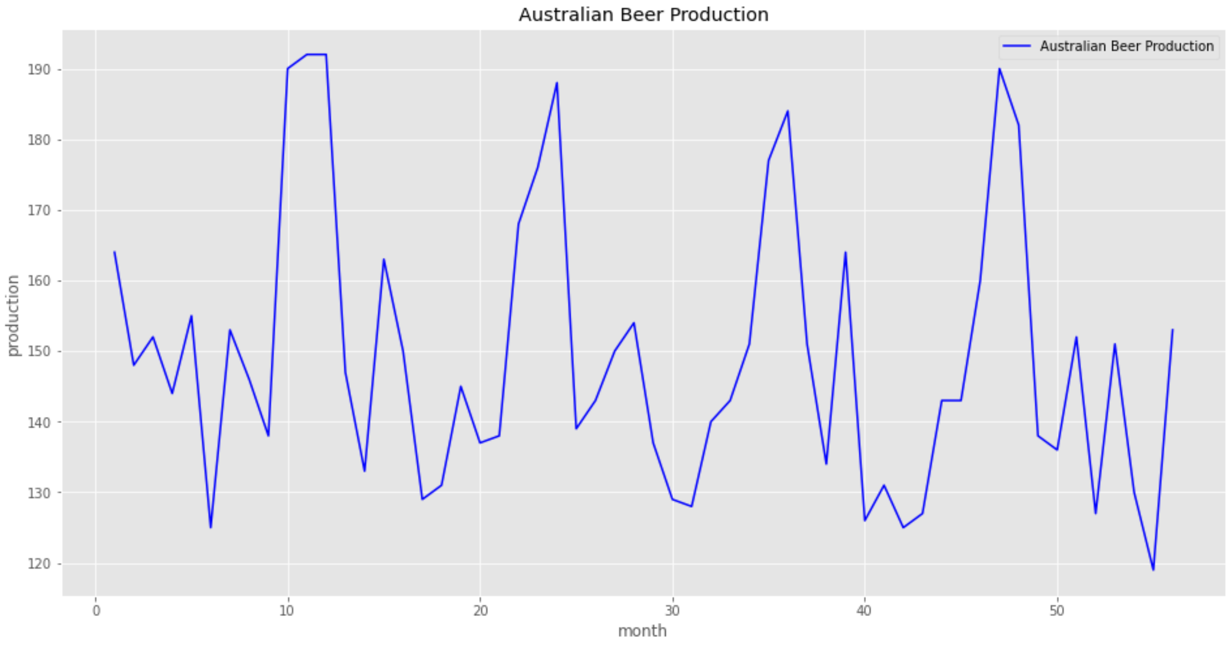
Auto-Correlation: The effects of patterns

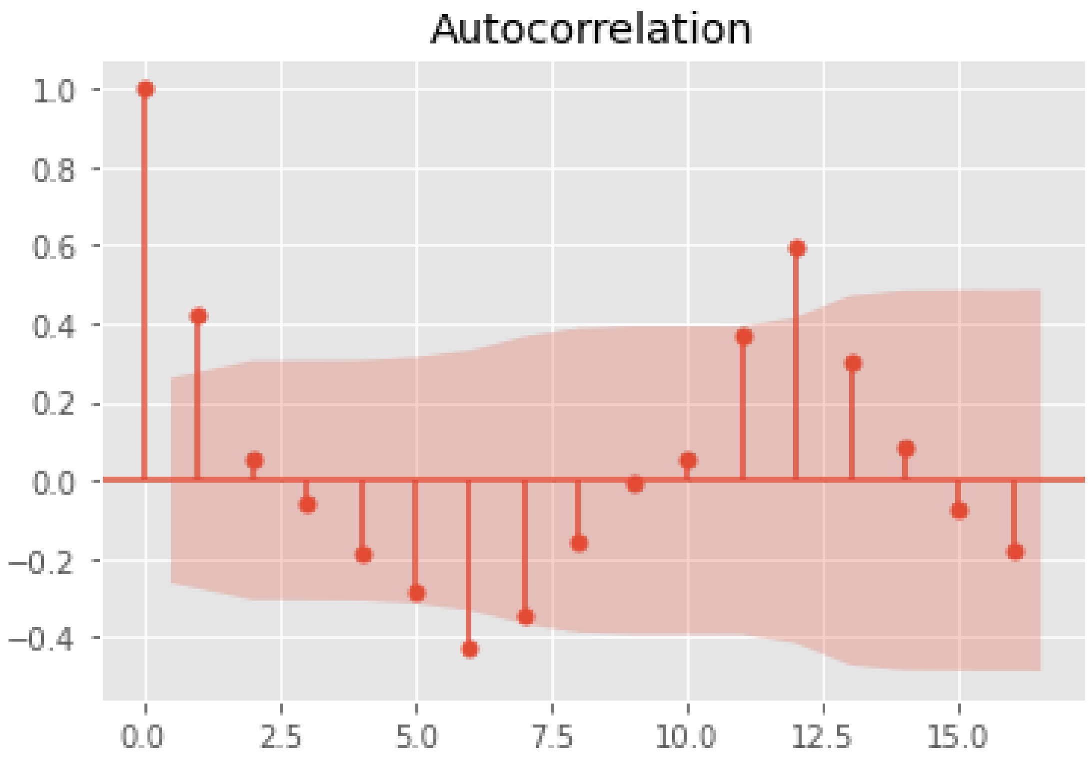
- The effect of patterns:
  - We would like to explore the auto-correlation structure of the demand time series to construct models that can take into account the dependence explicitly.
  - First, a relatively trivial observation. All basic patterns in the data (trend, seasonality) etc. reflect onto the autocorrelation structure.
  - To perform any useful autocorrelation analysis, we first have to transform the data to remove the trend, seasonality etc.

Auto-Correlation: The effects of patterns

- The effect of patterns:

Australian Monthly Beer Production

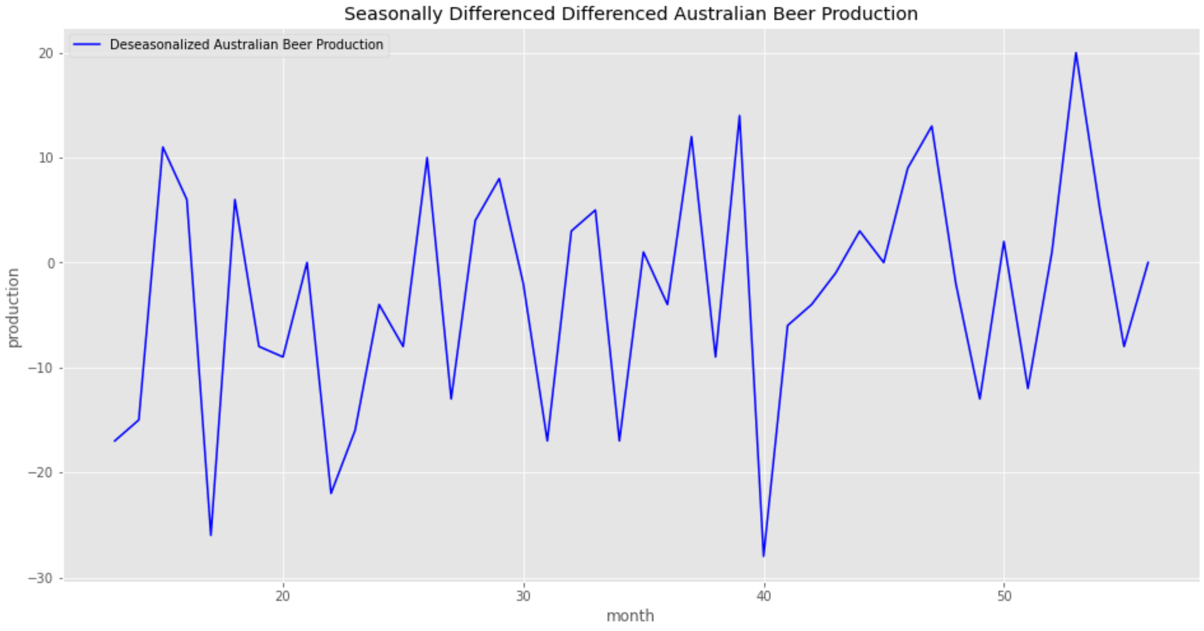


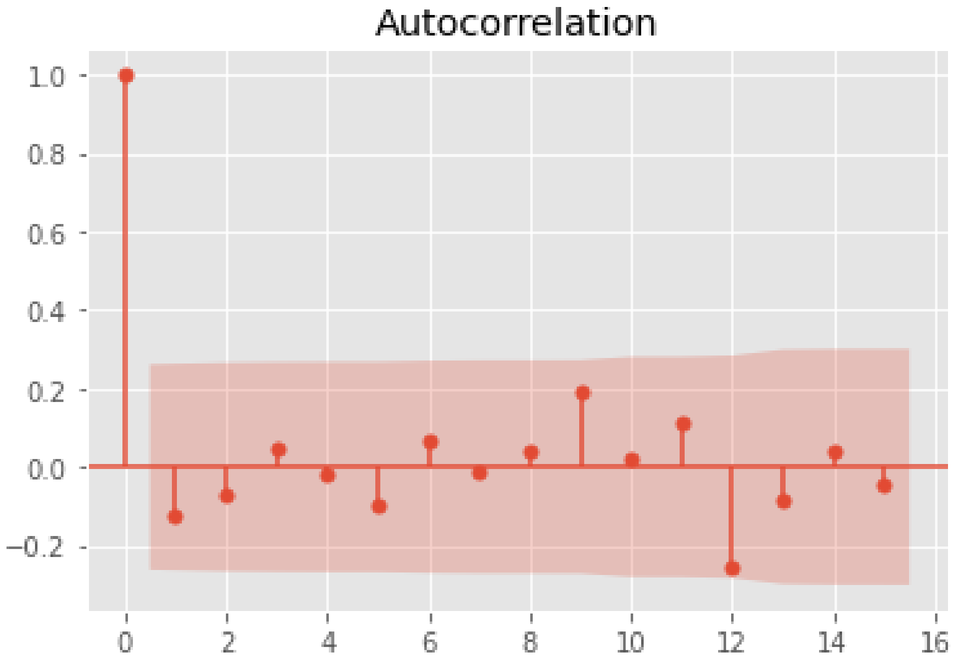


Auto-Correlation: The effects of patterns

- The effect of patterns: deseasonalized Australian Beer Production

Deseasonalized Australian Monthly Beer Production





After deseasonalizing no significant AC left. Then taking a simple seasonal difference would be a sufficient forecast.

Auto-Regressive (AR) models

- We started our modeling analysis with demand models that were in the form of $Y_t = f(t) + \epsilon_t$ (where ϵ_t are iid). Note that if we know $f(t)$ or once we figure out the functional form of its pattern from existing data, there is no remaining auto-correlation.
- We'll now consider models with a dependence structure. For instance, the AR model has the following structure:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \epsilon_t$$

  This is referred to as an AR-p model since it has p auto-regressive terms. Note that this is different than a typical regression because the right hand side involves terms from the same series (hence auto-regression).

Auto-Regressive models: AR(1)

- Let us consider the simplest model of this type, AR-1

$$Y_t = c + \phi_1 Y_{t-1} + \epsilon_t$$

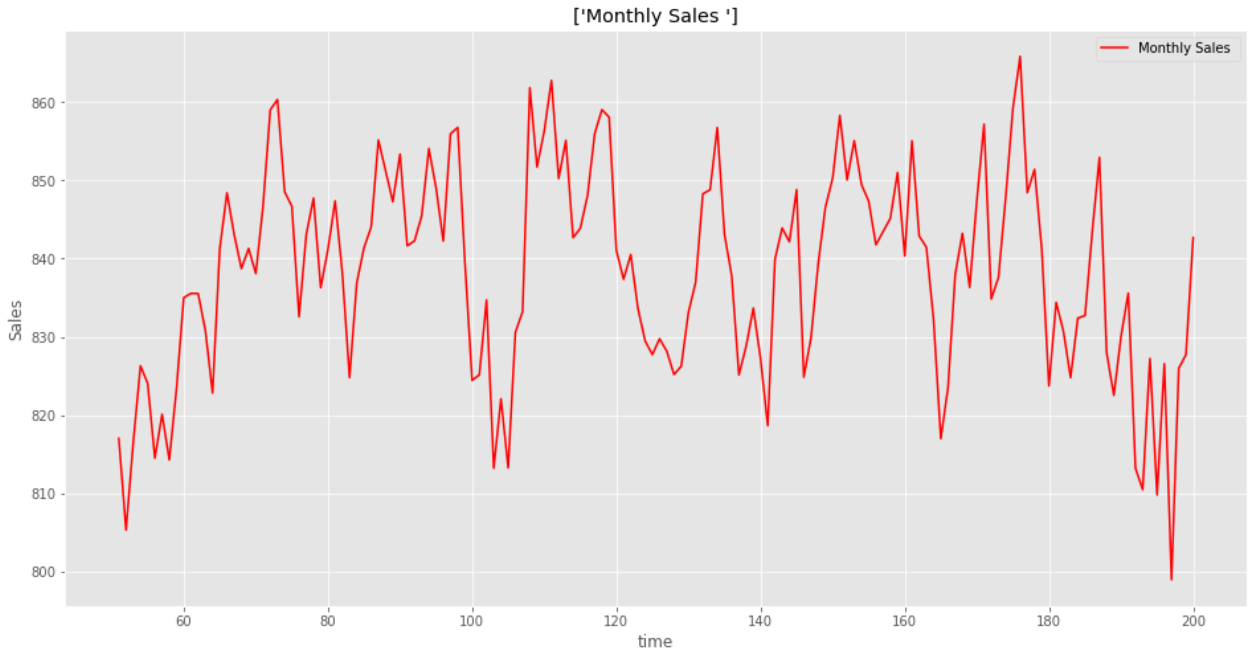
- We can already figure out some of the basic properties. First, we have to have the AR coefficient: $-1 < \phi_1 < 1$, otherwise the series would diverge (in expectation). Note that for general AR-p processes the stability conditions for the parameters are more complicated (please see Hyndman and Athansapoulos, Chapter 8).
- If we take ϕ_1 to be positive and high (i.e. close to 1), it is clear that Y_{t-1} and Y_t are highly correlated. In fact, we can verify that $Corr(Y_{t-1}, Y_t) = \phi$.
- But due to the recursive structure, Y_{t-2} and Y_t are also correlated. In fact, we can verify that $Corr(Y_{t-2}, Y_t) = \phi_1^2$ and in general $Corr(Y_{t-k}, Y_t) = \phi_1^k$.

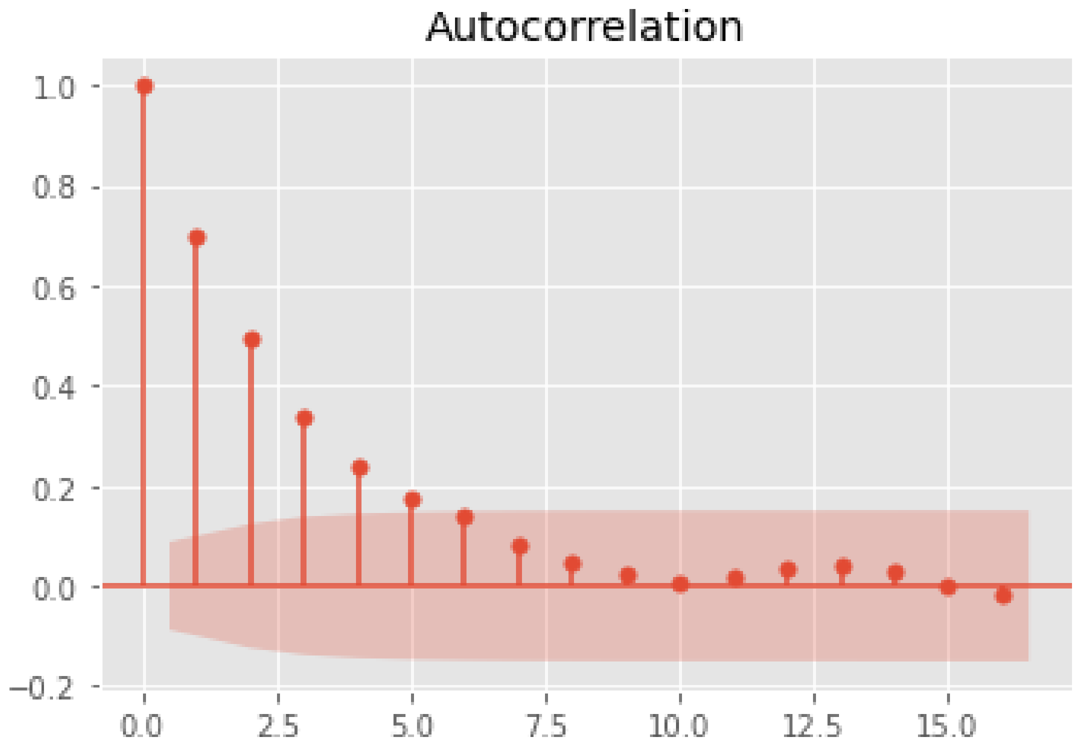
Auto-Regressive models: AR(1)

- If we take ϕ_1 to be negative and high (i.e. close to -1), it is clear that Y_{t-1} and Y_t are highly but negatively correlated (i.e. ϕ close to -1). We know that $Corr(Y_{t-k}, Y_t) = \phi^k$. Therefore, we have positive AC at even lags and negative AC at odd lags.
- When ϕ is close to $+1$, the process tends to take high (i.e. above average values) for a number of consecutive periods and then may fall due to the error component, once it falls it tends to stay low for a while.
- When ϕ is close to -1, the process tends to alternate between high and low values in consecutive periods (zigzagging).

AR(1) Examples

Randomly Generated Data: AR(1) with ϕ_1=0.7.

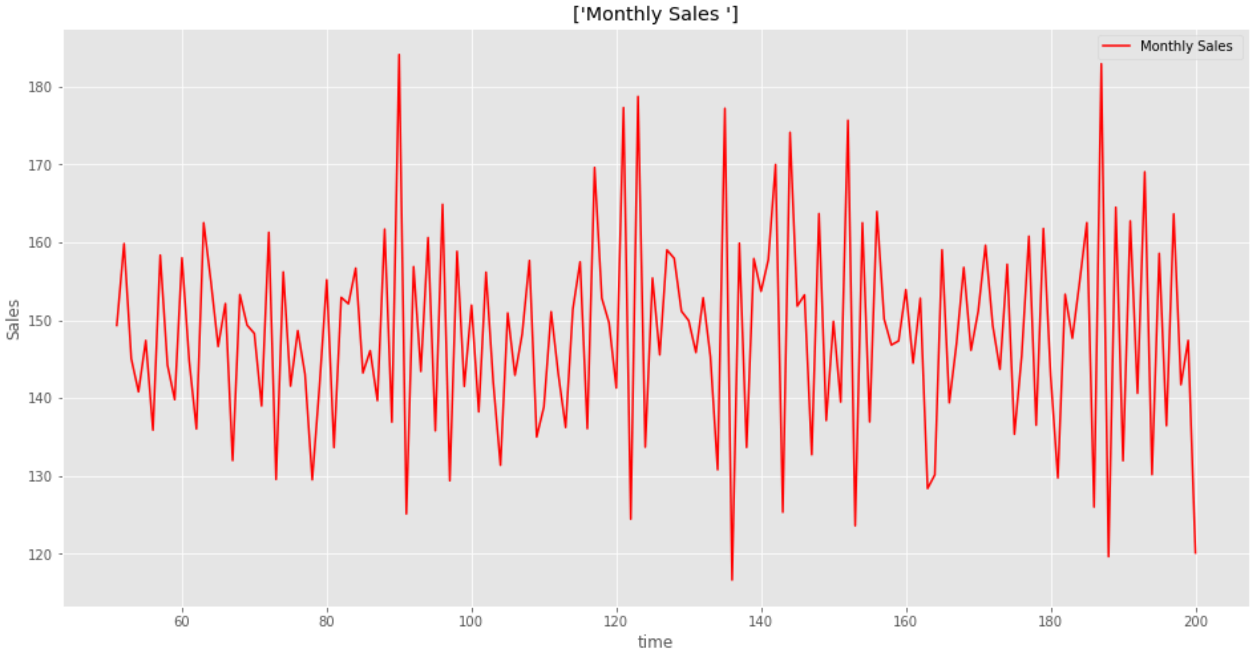


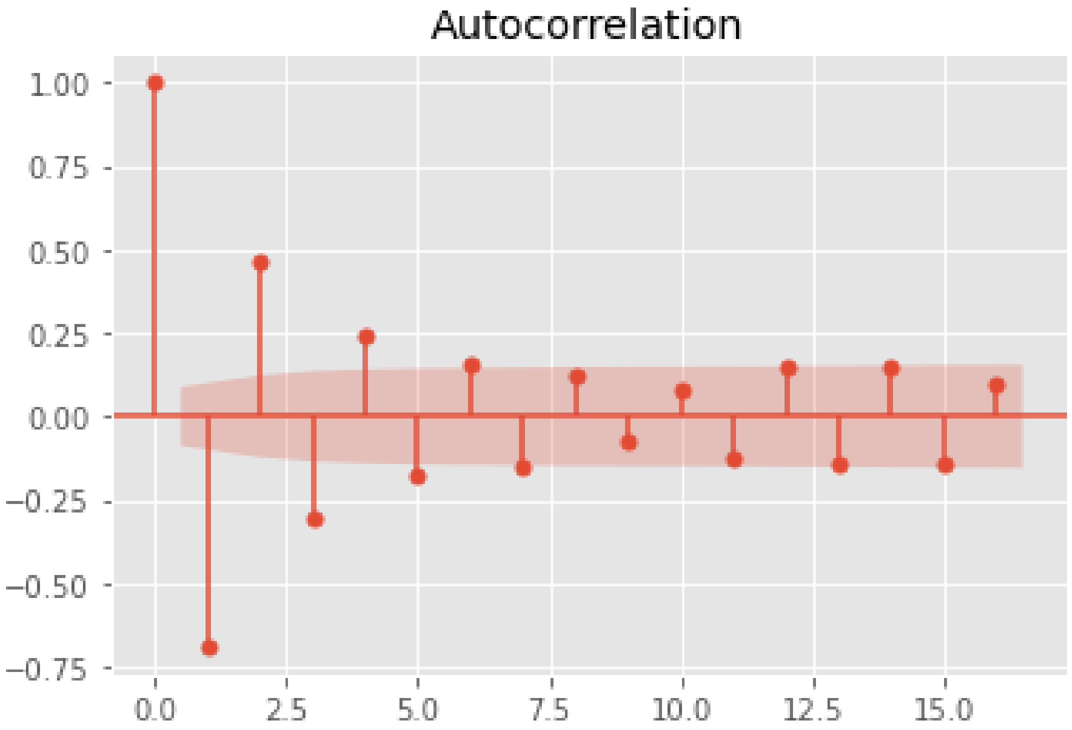


Note the geometric decrease in the AC's starting from lag 1.

AR(1) Examples

Randomly Generated Data: AR(1) with ϕ_1=-0.7.





This time the AC's geometrically decreasing in absolute value but alternating in sign -,+,- etc.

Poll: March 2, 2022

- $Y_t = 20 + (0.6)Y_{t\text{-}1} + \varepsilon_t$.

Moving Average (MA) models

- The AR-process generates dependence by making Y_t linearly dependent on Y_{t-k}. This is a particular type of dependence. An alternative to this to generate dependence through the error terms. The following process is called a Moving Average (MA)- process:

  $$Y_t = c + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} + \epsilon_t$$

  The above is referred to as an MA-q model since it has q MA terms.

- Note that this is considerably different than the AR-process. Y_t can be viewed as a weighted average of past q forecast errors. Depending on the sign of θ_j, the forecast error may have a positive or negative effect on Y_t.

Moving Average (MA) models: MA(1)

- Let us take MA-1

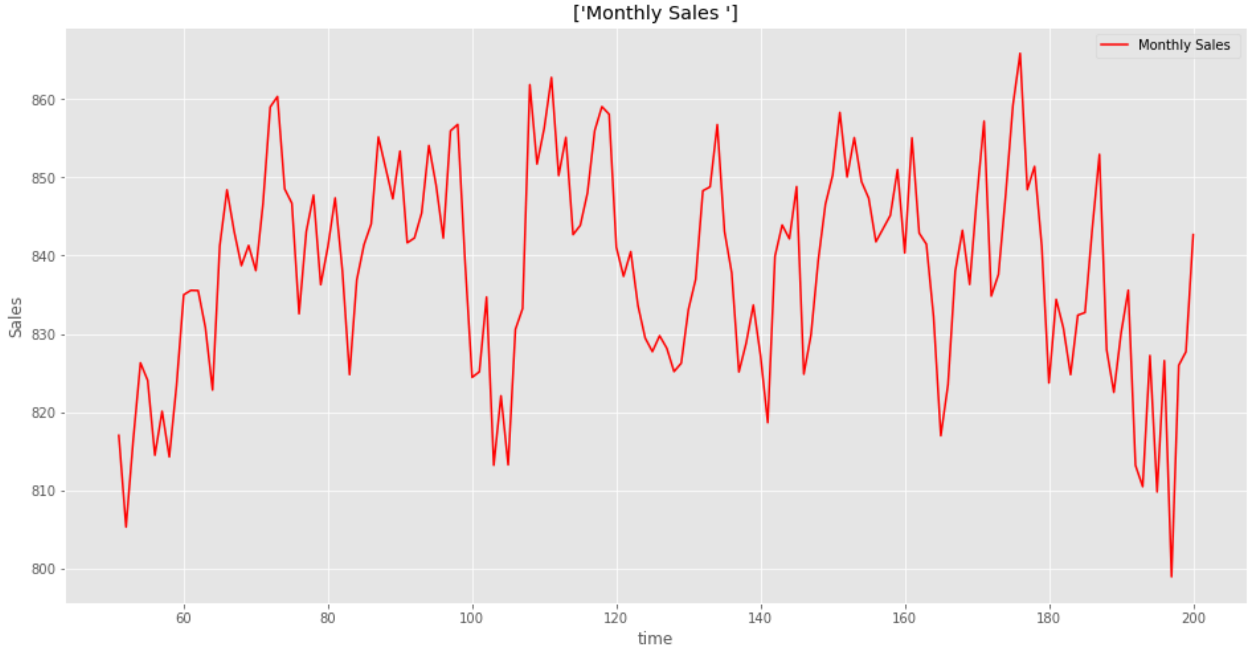
$$Y_t = c + \theta_1 \epsilon_{t-1} + \epsilon_t$$

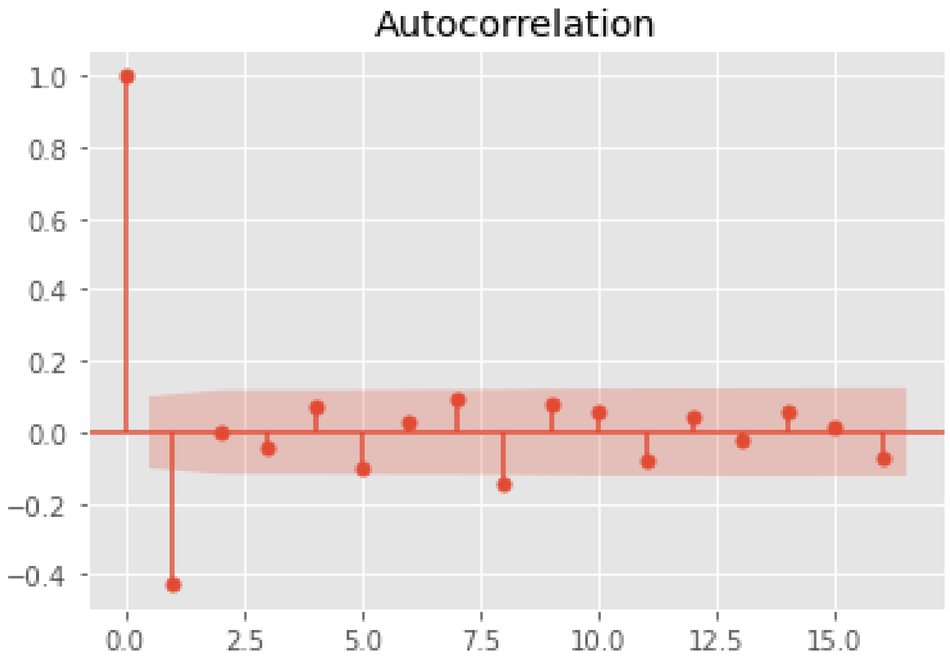
- First, for stability we need $|\theta_1| < 1$. Once again for general MA-q processes the stability conditions for the parameters are more complicated (please see Hyndman and Athansapoulos, Chapter 8).
- Next, we can verify that Y_{t-1} and Y_t are correlated but Y_{t-2} and Y_t are not. Therefore, the auto-correlation structure is very different than the AR-process.

If θ_1 is positive then AC at lag 1 is negative, if θ_1 is negative then AC at lag 1 is positive.

MA(1) Examples

Randomly Generated Data: MA(1) with θ_1=0.7.





Note that there is a single spike at lag 1 but no geometric decay (AC's at all other lags are insignificant).

ARMA Framework

- We can combine AR-terms and MA-terms. The resulting models are called ARMA and include both AR and MA components.

$$\begin{aligned} Y_t &= c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \\ &\quad + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q + \epsilon_{t-q} + \epsilon_t \end{aligned}$$

This is useful in practice because we need flexible models to fit data. Real auto-correlations rarely correspond to pure AR or MA processes.

ARMA: Mixing AR and MA

- ARMA: mixing AR and MA terms
- Ex: $Y_t = c + \phi_1 Y_{t-1} + \theta_1 \varepsilon_{t-4} + \varepsilon_t$

- A flexible model: Geometrical decrease starting from from lag 1 but there is an additional single spike at lag 4.

Model Identification

- We have a fairly rich framework but we need to have a way of guiding the model fitting process.
- As we will see, a good part of model fitting may be automated but we obtain much better results if we can do some preliminary analysis to guide the fitting process.
- How do we identify the correct model (how many AR and MA terms at which lags?)
- This cannot be done by simply plotting the data.
- The ACF plot is useful but we need more help.

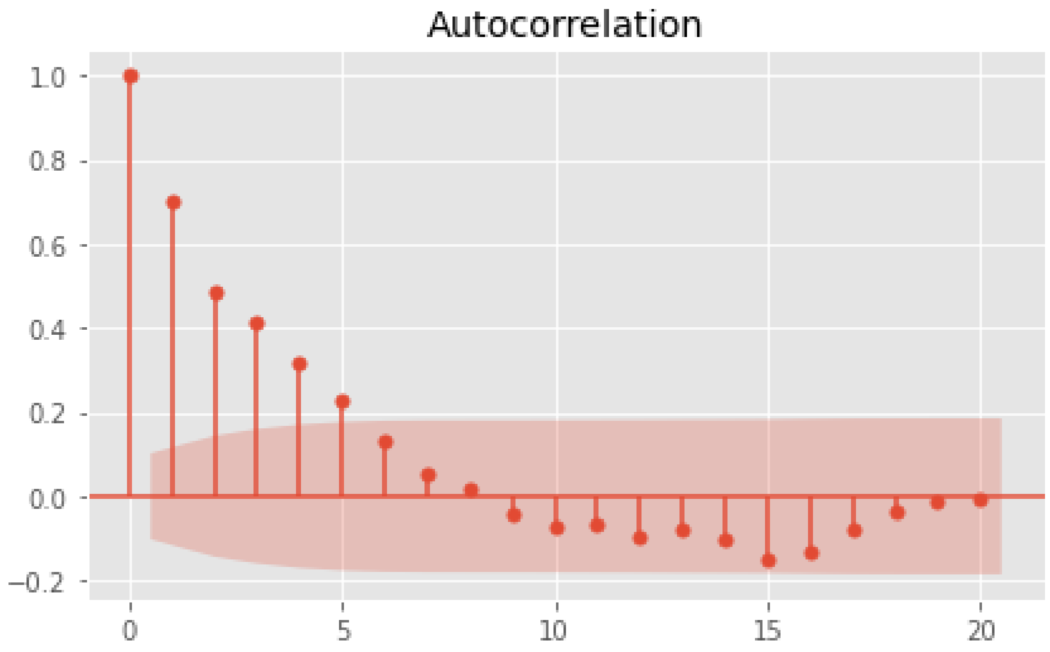
Model Identification: ACF and PACF

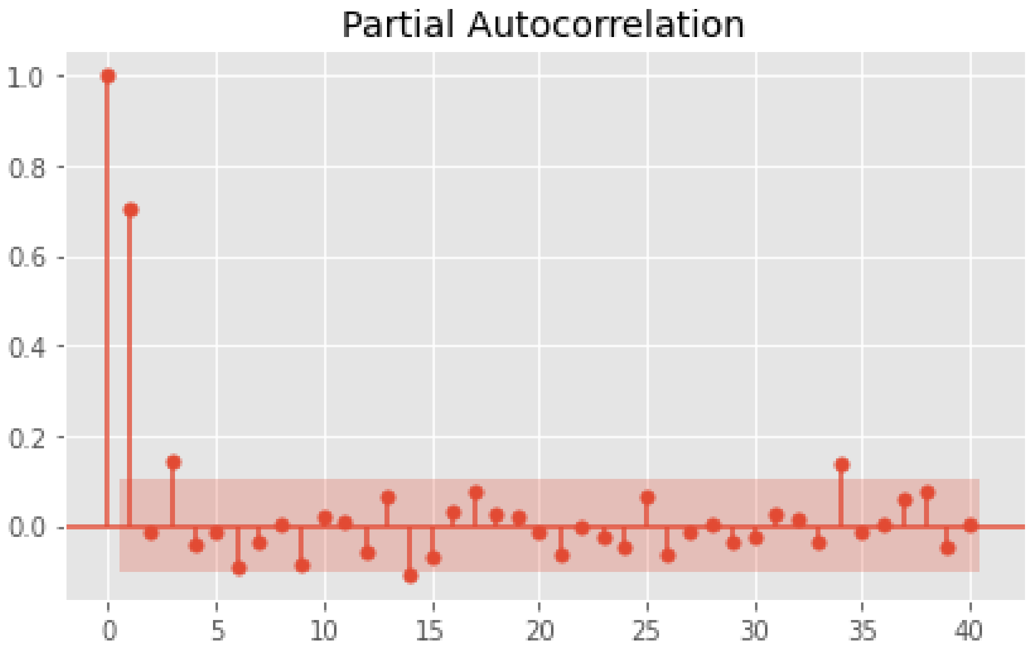
- Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF)
- The ACF gives the $Corr(Y_{t-k}, Y_t)$ as a function of the lag k.
- The PACF is the coefficient that corresponds to the coefficient lag-k when we run a linear regression with lagged observations on the right hand side.
- The ACF and PACF capture different aspects. For instance an AR(1) process has the highest AC at lag 1 but also geometrically decreasing AC's at lags 2, 3 etc. The PA coefficient just takes a non-zero value at lag 1 but there is no PA at other lags.

$$Y_t = c + \phi_1 Y_{t-1} + \epsilon_t$$

AR ACF-PACF patterns

Randomly Generated Data: AR(1) with ϕ_1=0.7.





Note that the PAC has a spike at lag 1 but vanishes at higher lags.

MA(1) ACF-PCF patterns

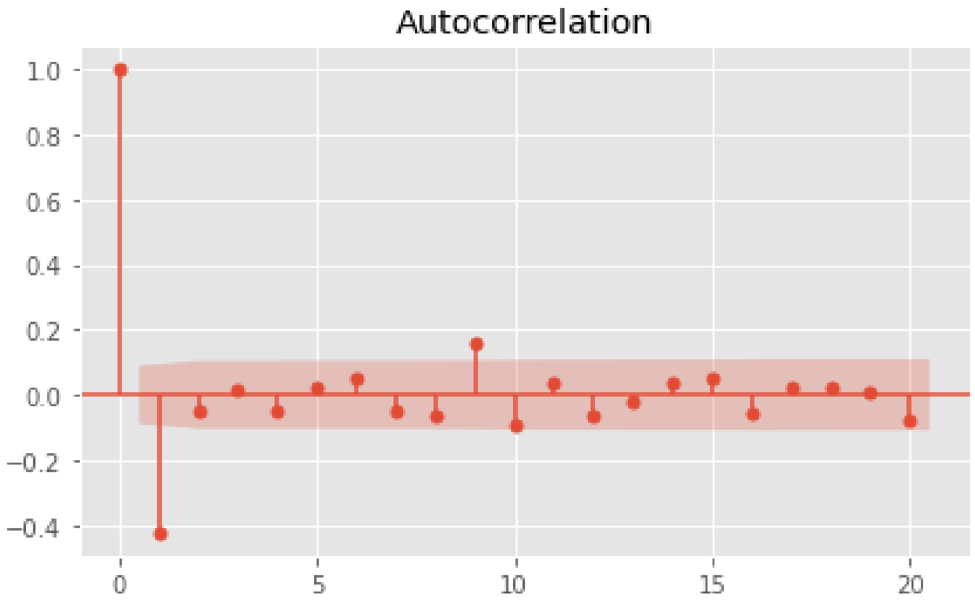
- For an MA(1) process things reverse
- The ACF plot only shows AC at lag 1 but no AC at higher lags.
- The PACF plot shows the highest PAC at lag 1 but also geometrically decreasing PAC's at higher lags.

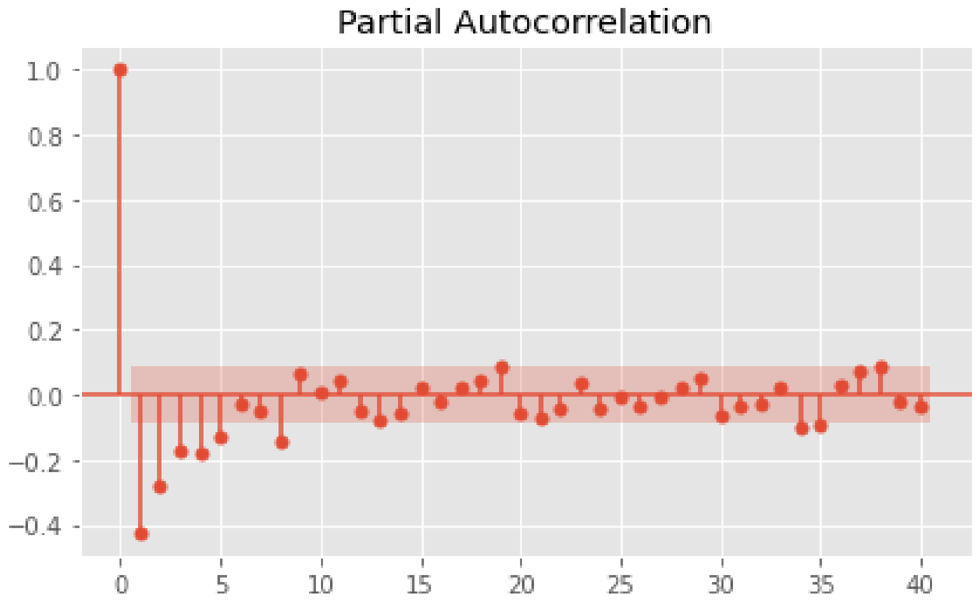
$$Y_t = c + \theta_1 \epsilon_{t-1} + \epsilon_t$$

- This enables us to distinguish MA and AR patterns from the ACF and PACF.

MA(1) ACF-PCF patterns

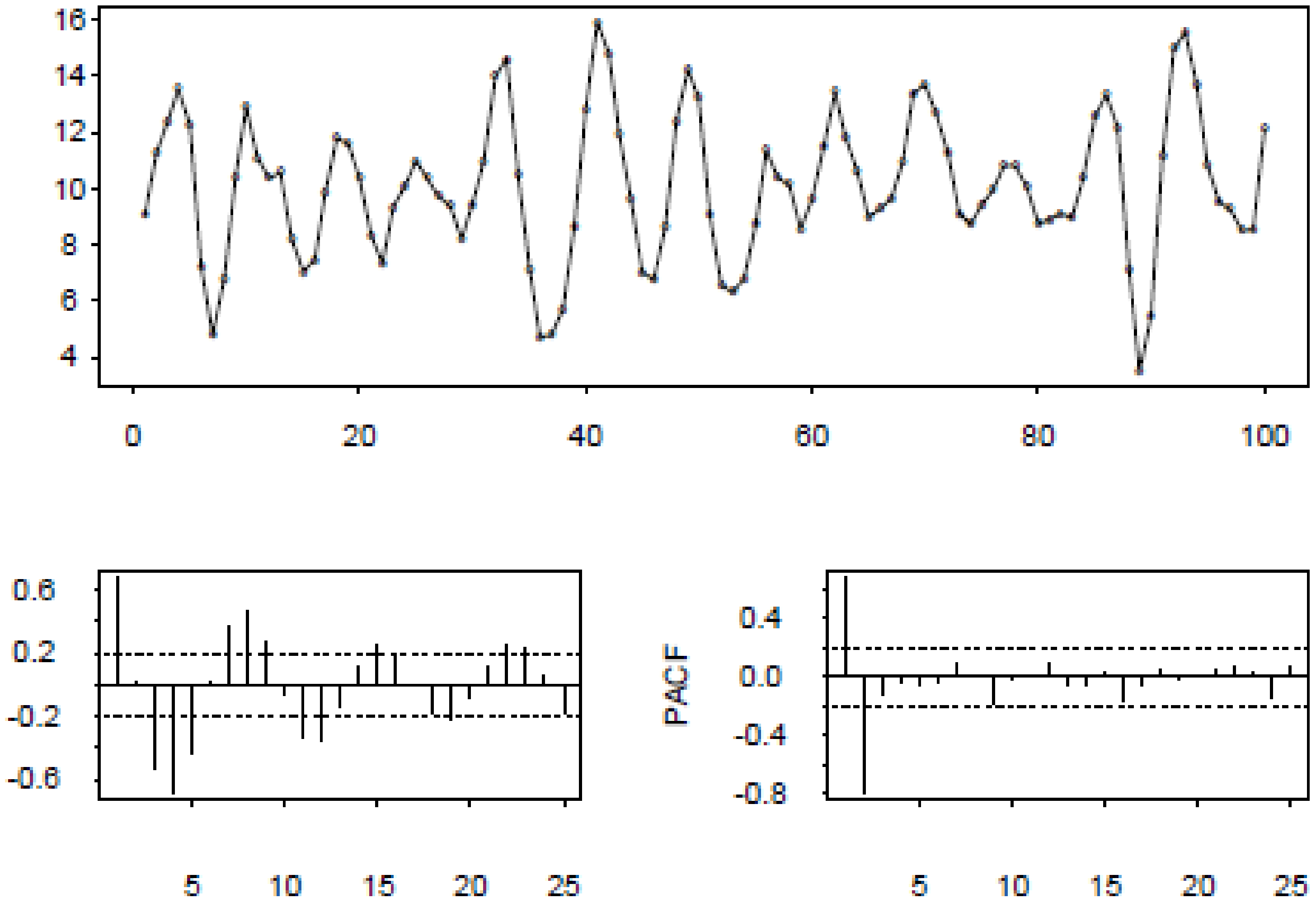
Randomly Generated Data: MA(1) with θ_1=-0.7.





Note the geometric decrease in the PAC this time.

Poll 2


16
14
12
10
8
6
4
0
20
40
60
80
100
0.6
0.2
-0.2
-0.6
5
10
15
20
25
PACF
0.4
0.0
-0.4
-0.8
5
10
15
20
25

Poll Example: AR (2)

$$Y_t = 6 + 1.2Y_{t-1} - 0.8Y_{t-2} + e_t,$$

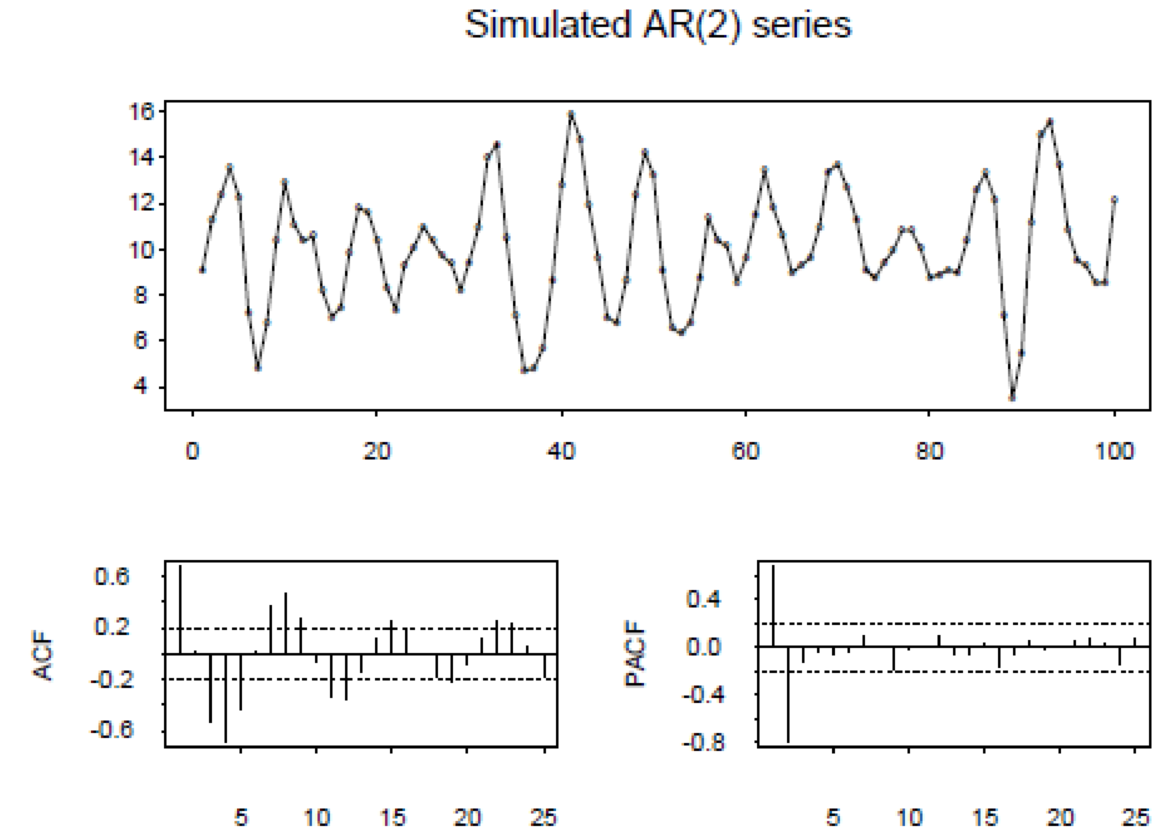


Figure 7-15: *An ARIMA(2,0,0) model:* $Y_t = 6 + 1.2Y_{t-1} - 0.8Y_{t-2} + e_t$, $n = 100$.

It's hard to tell from the ACF, this is AR(2) but with help from PACF we can conclude that it is likely to be AR(2) (two spikes at lags 1 and 2 but no decreasing behavior).

Makridakis, Wheelwright and Hyndman (1997)

Example: MA (2)

$$Y_t = e_t - e_{t-1} + 0.8e_{t-2}.$$

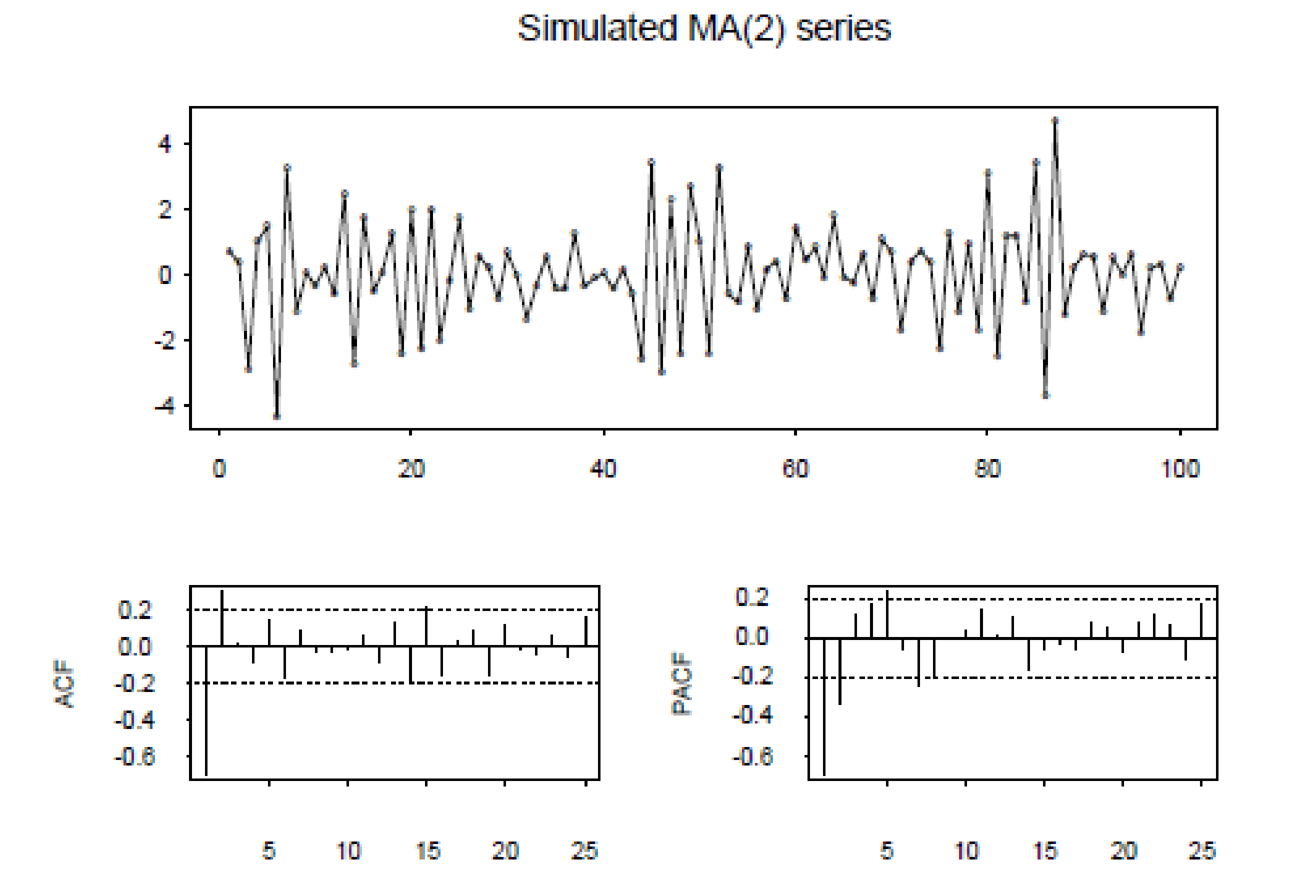


Figure 7-16: *An ARIMA(0,0,2) model:* $Y_t = e_t - e_{t-1} + 0.8e_{t-2}$, $n = 100$.

This time the ACF is more helpful.

Makridakis, Wheelwright and Hyndman (1997)

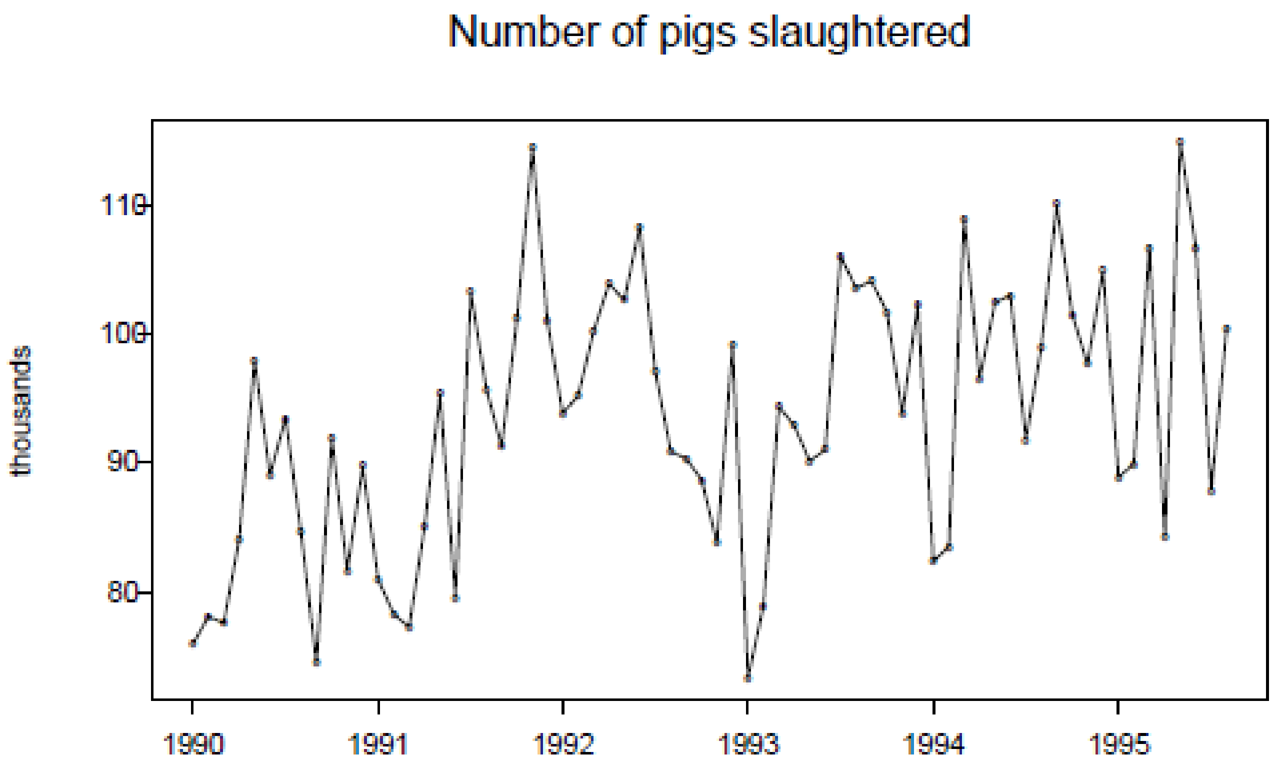
Summary of ACF and PACF patterns for simple AR and MA models

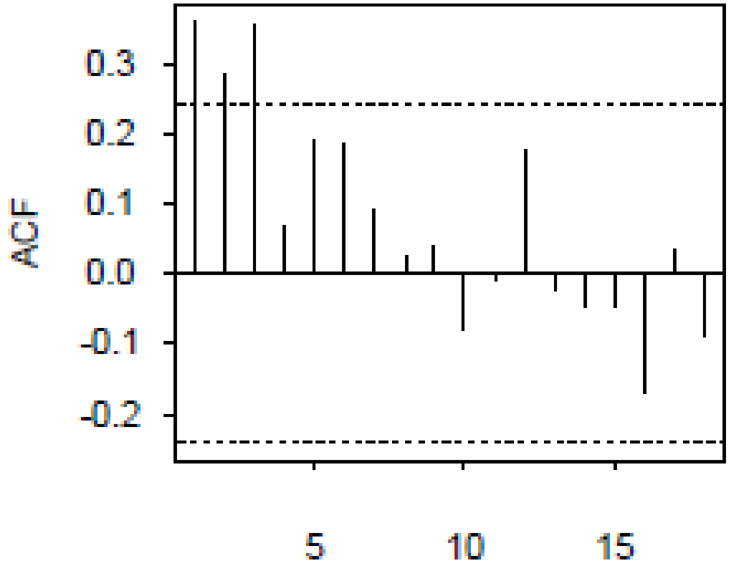
Process	ACF	PACF
AR(1)	Exponential decay: on positive side if $\phi_1 > 0$ and alternating in sign starting on negative side if $\phi_1 < 0$.	Spike at lag 1, then cuts off to zero: spike positive if $\phi_1 > 0$, negative if $\phi_1 < 0$.
AR(p)	Exponential decay or damped sine-wave. The exact pattern depends on the signs and sizes of $\phi_1, \ldots, \phi_p$.	Spikes at lags 1 to p, then cuts off to zero.
MA(1)	Spike at lag 1 then cuts off to zero: spike positive if $\theta_1 < 0$, negative if $\theta_1 > 0$.	Exponential decay: on negative side if $\theta_1 > 0$ and alternating in sign starting on positive side if $\theta_1 < 0$.
MA(q)	Spikes at lags 1 to q, then cuts off to zero.	Exponential decay or damped sine-wave. The exact pattern depends on the signs and sizes of $\theta_1, \ldots, \theta_q$.

Table 7-2: *Expected patterns in the ACF and PACF for simple AR and MA models.*

Makridakis, Wheelwright and Hyndman (1997)

Example: Number of Pigs Slaughtered





Maybe:

But the PACF would be helpful