INDR 422/522

Fikri Karaesmen

Spring 2023

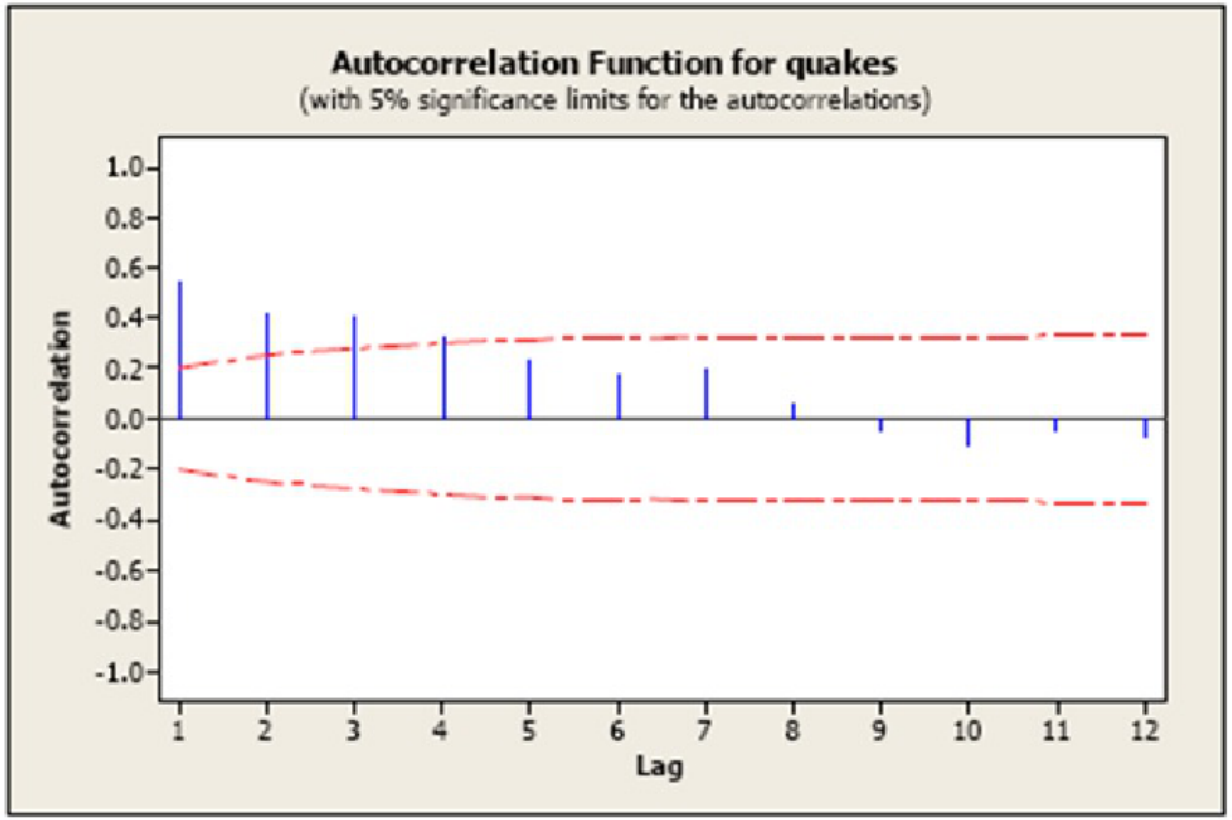
ARIMA Forecasts- 2

March 21, 2023

Reminders

- Blackboard page is becoming active
  - Last Week’s slides
  - Last year’s lecture slides
  - Class Exercise Solutions
  - Will be uploading the current slides as we proceed
- Second lab available, please take a look and work on the exercises
- Third lab will be available this Friday
- Participation taken. Please participate in polls.
- HW 1 (due-date March 31, 2023)

Class Exercise from last lecture



Class Exercise from last lecture

CLASS EXERCISE, March 16, 2023

1. Let us assume that we explore the ACF plot of some time series and see that the only significant AC is at lag 2 and is positive. Which of the following are true?

   (a) $Corr(D_t, D_{t-1})$ is close to zero

   Solution: This is true since the plot reports statistically significant AC at lag 2.

   (b) If D_t is lower than average than D_{t+1} is likely to be higher than average.

   Solution: False. Because there is no significant AC except at lag 2.

   (c) D_t should not have any trend

   Solution: True. If there had been trend, we would have seen significant and high AC at all lags.

   (d) D_t may have strong seasonality

   Solution: False. If there had been seasonality, we would have seen the seasonaly cycle with high AC at seasonal lags.

2. Consider the data series Y_t (annual number of significant earthquakes in the world, source: https://online.stat.psu.edu/stat510/lesson/1/1.2). Which staments are true?

   (a) A reasonable model for Y_t is $Y_t = c + \epsilon_t$

   Solution: False. This model does not consider AC but there is significant AC in the series.

   (b) A reasonable model for Y_t is $Y_t = a_0 + a_1 Y_{t-1} + \epsilon_t$

   Solution: This might be plausible. There appears to be a geometric decay in the AC starting from lag 1.

   (c) Number of earthquakes in consecutive years are positively correlated

   Solution: True.

   (d) There might be a strong trend for the series Y_t *Solution:* False. With strong trend we would have seen higher AC at all lags.

   (e) There might be a strong seasonality in the series Y_t

   Solution: False. With strong trend we would have seen the seasonal cycle.

   (f) Y_t does not have statistically significant correlation with Y_{t-6}

   Solution: True.

Summary last lecture

- Auto-regressive processes generate auto-correlation
- Our goal is to understand how AC is generated by particular models so that we can do predictions which require the reverse direction:
  - Given data with AC properties, what is the model that is likely to generate this data
- Patterns (trend, seasonality etc.) reflect onto the AC function
  - But this is not a major problem, we have ways of detrending and desasonalizing the data.

Auto-Regressive (AR) models

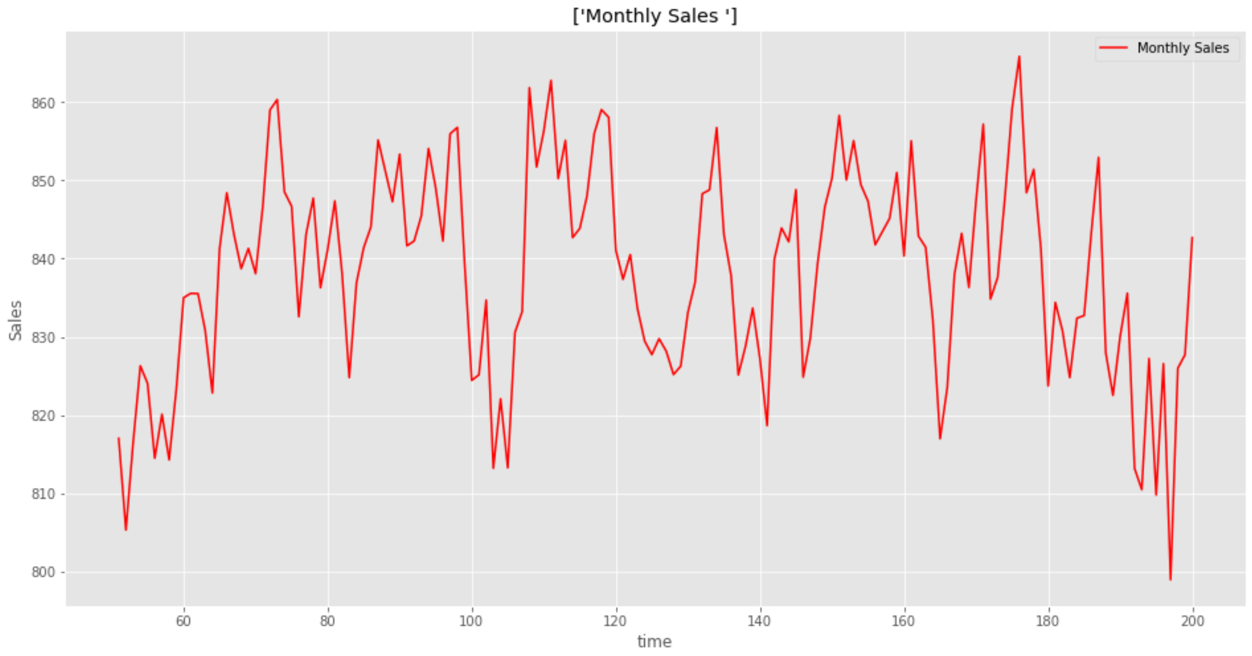
- We started our modeling analysis with demand models that were in the form of $Y_t = f(t) + \epsilon_t$ (where ϵ_t are iid). Note that if we know $f(t)$ or once we figure out the functional form of its pattern from existing data, there is no remaining auto-correlation.
- We'll now consider models with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

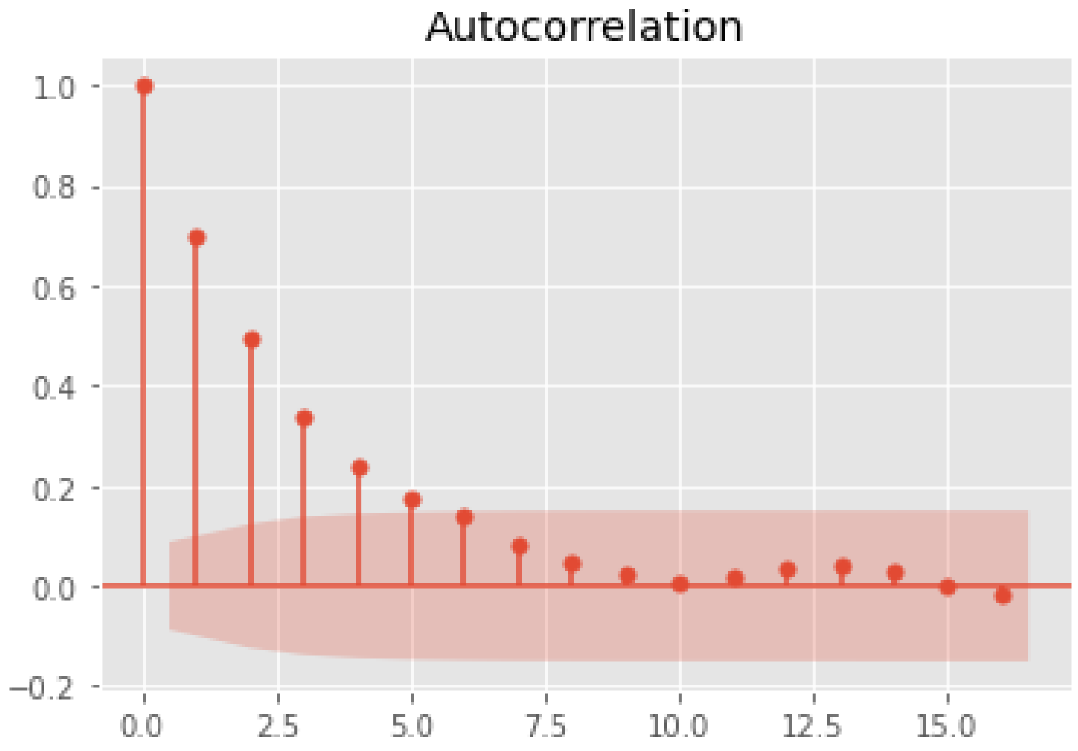
$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \epsilon_t$$

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).

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.

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

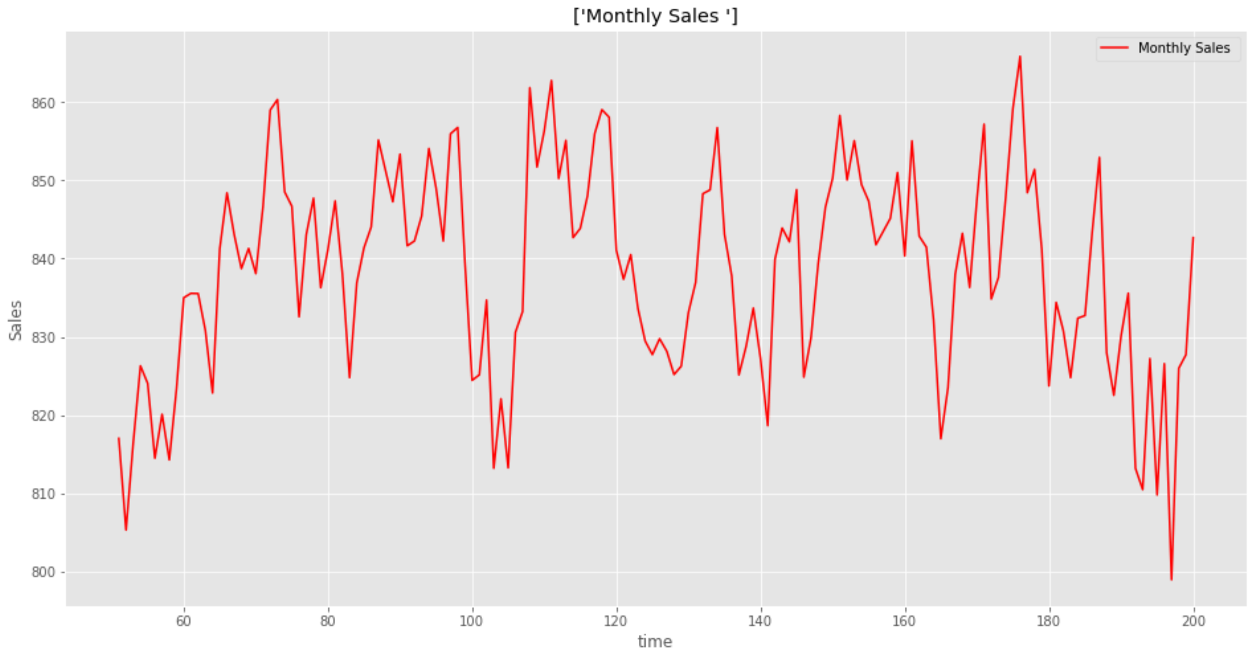
$$Y_t = c + \theta_1 \epsilon_{t-1} + \epsilon_t$$

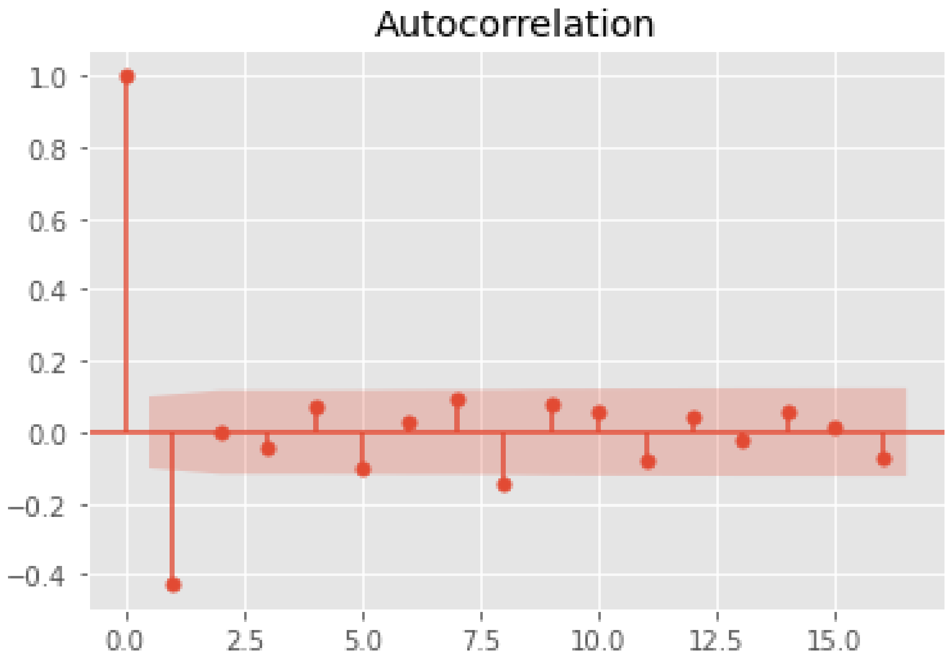
- First, for invertibility (please see Hyndman and Athansapoulos, Chapter 8). we need $|\theta_1| < 1$. Once again for general MA-q processes the invertibility conditions for the parameters are more complicated
- Next, we can verify that Y_{t-1} and Y_t are correlated but Y_{t-2} and Y_t are not. Therefore,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





$$\rho_1 = \frac{\theta_1}{1+\theta_1^2} = \frac{-0.7}{1.49} = -0.47,\ \ \rho_2 = \rho_3 = \ldots = 0.$$

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.

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

Partial auto-correlation functions

- The Auto-Correlation function computes the correlations of pairs of observations separated by time lags one pair at a time i.e D_{t-1} and D_t or D_{t-3} and D_t etc.
- This is similar to running a single variable linear regression: $D_t = a + b\, D_{t-\tau} + \varepsilon_t$ and computing its R^2 value (for each lag).
- Consider the AR(1) process:

$$D_t = a_0 + a_1 D_{t-1} + \varepsilon_t$$

- We can see that the true relationship is between the consecutive observations but the auto-correlation at other lags is also non-zero since observations at time t-2 affect observations at time t-1 and therefore indirectly those at time t.
- The partial auto-correlation function (PACF) overcomes this drawback.

Model Identification: Partial autocorrelations

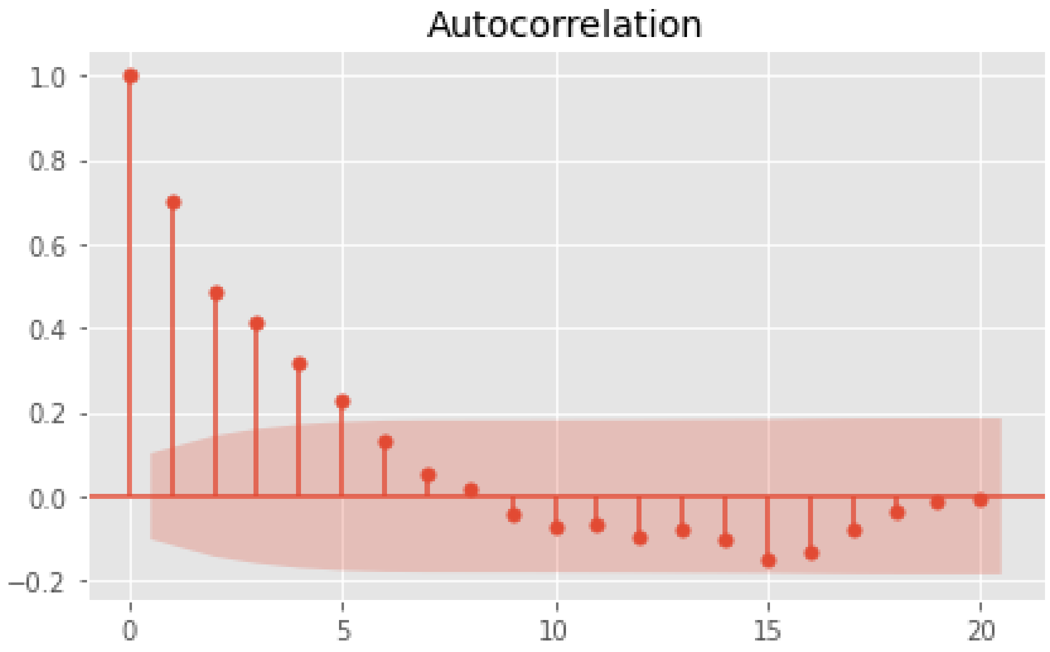
- Consider the following regression for D_t as a function of the past q demand realizations:

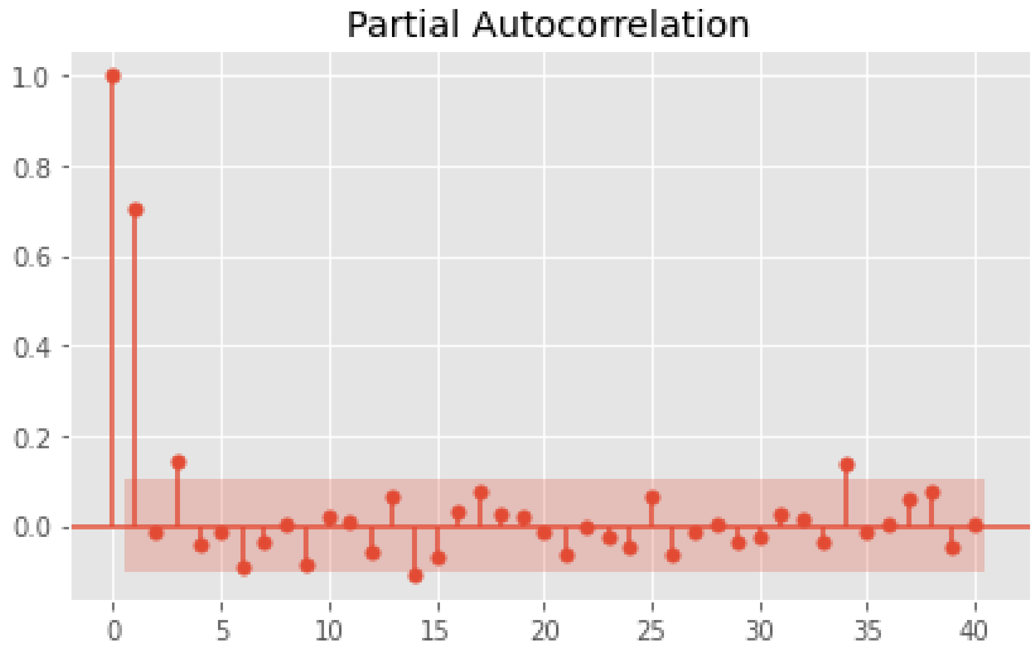
$$D_t = b_0 + b_1 D_{t-1} + b_2 D_{t-2} + \ldots + b_q D_{t-q} + \varepsilon_t.$$

- Note that this is not a typical regression (the explanatory variables are past realizations of demand) but an autoregression.
- The partial autocorrelation α_k is the estimated coefficient for b_k in the above autoregression. The first partial autocorrelation equals the first autocorrelation.
- We then plot the Partial Auto-Correlation Function (PACF) along with the ACF.

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

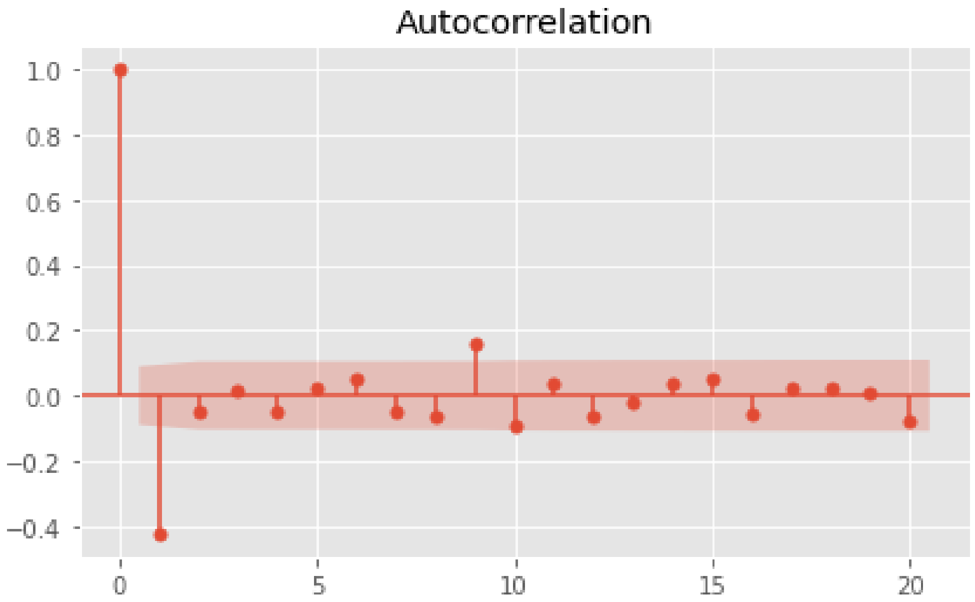
- For an MA(1) process things reverse
- The ACF plot only shows AC at lag 1 but no AC at higher lags.
- The PACF plot shows the highest PAC at lag 1 but also geometrically decreasing PAC's at higher lags.

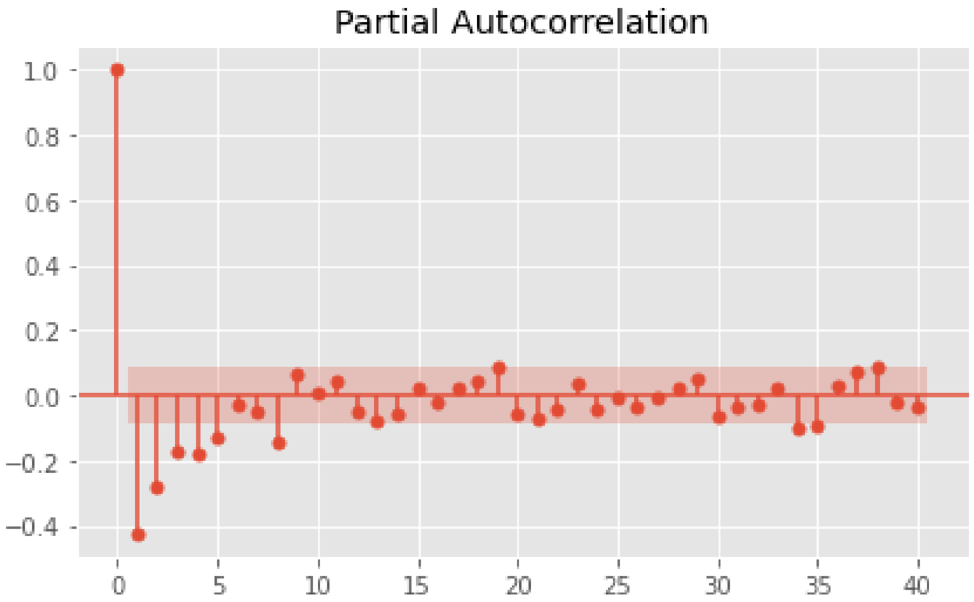
$$Y_t = c + \theta_1 \epsilon_{t-1} + \epsilon_t$$

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

MA(1) ACF-PCF patterns

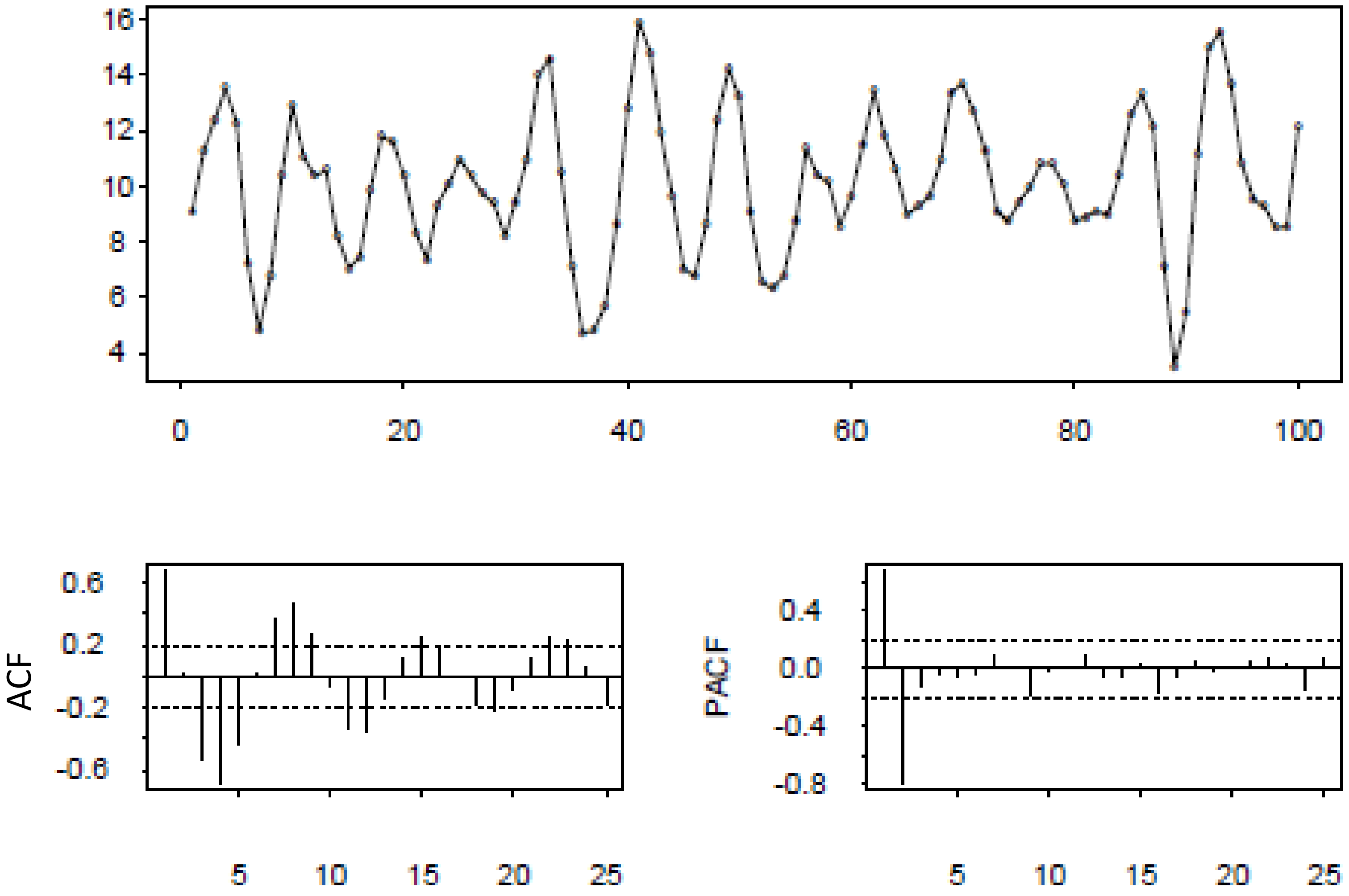
Randomly Generated Data: MA(1) with θ_1=0.7.





Note the geometric decrease in the PAC this time.

Poll 2



16
14
12
10
8
6
4
0
20
40
60
80
100
ACF
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)

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

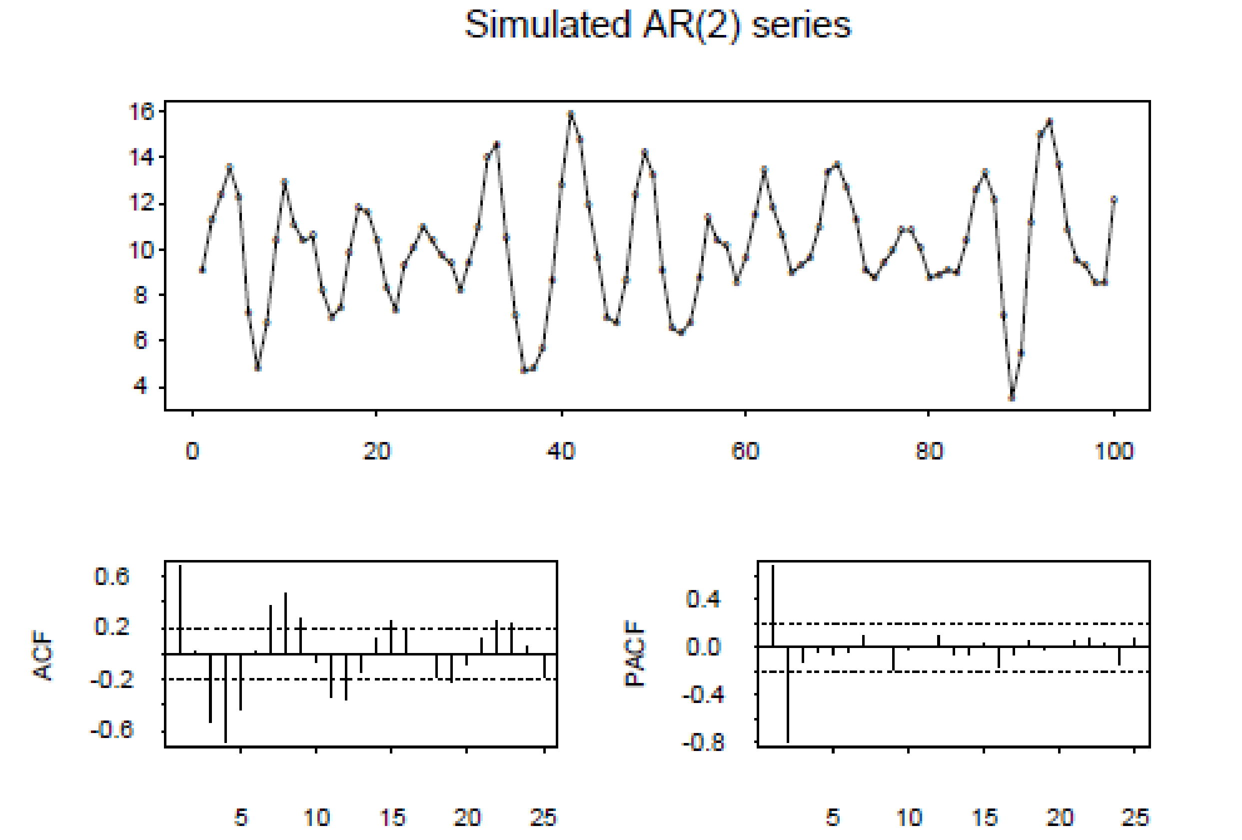


Figure 7-15: *An ARIMA(2,0,0) model:* $Y_t = 6 + 1.2Y_{t-1} - 0.8Y_{t-2} + e_t$, $n = 100$.

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)

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

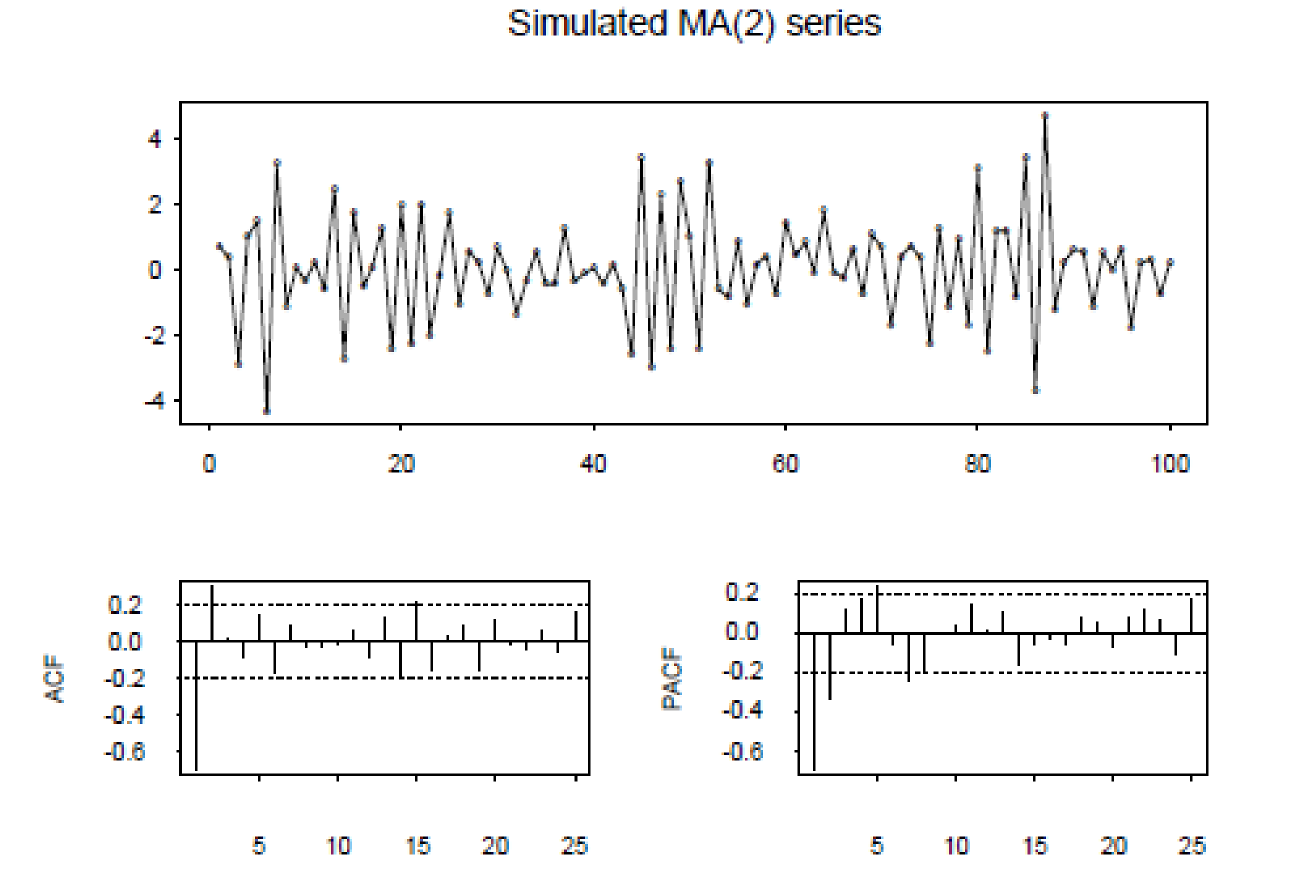


Figure 7-16: *An ARIMA(0,0,2) model:* $Y_t = e_t - e_{t-1} + 0.8e_{t-2}$, $n = 100$.

This time the ACF is more helpful.

Makridakis, Wheelwright and Hyndman (1997)

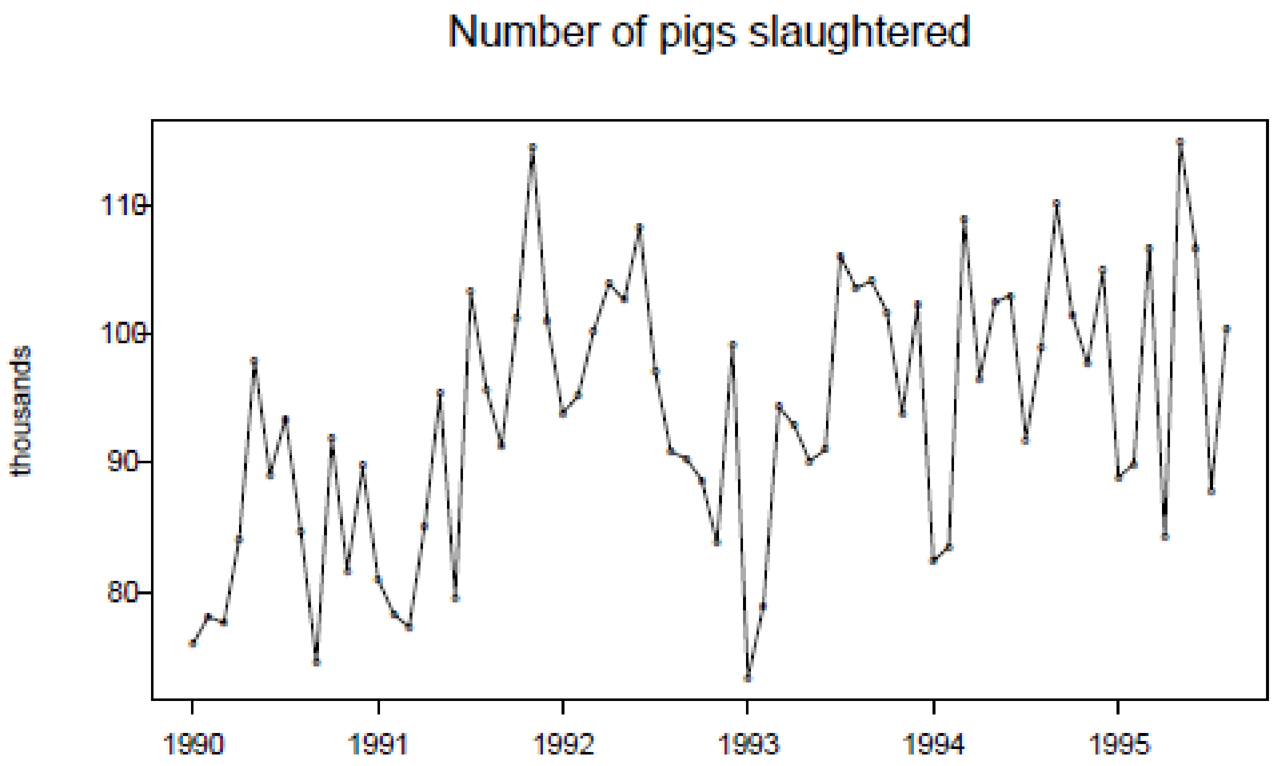
Summary of ACF and PACF patterns for simple AR and MA models

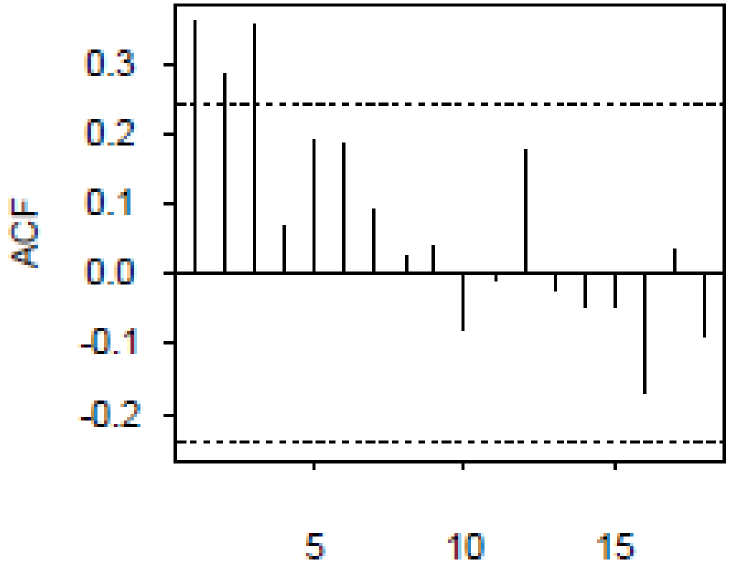
Process	ACF	PACF
AR(1)	Exponential decay: on positive side if $\phi_1 > 0$ and alternating in sign starting on negative side if $\phi_1 < 0$.	Spike at lag 1, then cuts off to zero: spike positive if $\phi_1 > 0$, negative if $\phi_1 < 0$.
AR(p)	Exponential decay or damped sine-wave. The exact pattern 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

ARIMA Framework

- Finally, we incorporate the basic transformations that are needed to convert the original series to a stationary series. One basic operation is differencing (multiple times if necessary).
- ARMA processes that require differencing are called ARIMA (Auto Regressive Integrated Moving Averages). Integration in this context is viewed as undoing the differencing (i.e. summation).
- We use the convention ARIMA(p, d, q) to denote that the original series was differenced d times, and then p AR and q MA terms were used on the differenced series.
- The ARIMA class is a broad and useful class.

ARIMA Framework: example

- ARIMA(1,1,0) refers to a process which was differenced once, and has an AR term on the difference:

$$
\begin{aligned}
i)\ W_t &= Y_t - Y_{t-1} \\
ii)\ W_t &= c + \phi_1 W_{t-1} + \epsilon_t
\end{aligned}
$$

We can revert the transformations to recover the original process:

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

Finally replacing W_{t-1} by $Y_{t-1} - Y_{t-2}$, we have:

$$Y_t = Y_{t-1} + W_t = c + Y_{t-1} + \phi_1 (Y_{t-1} - Y_{t-2}) + \epsilon_t$$

ARIMA Framework: example

- ARIMA(1,2,1) refers to a process which was differenced twice, and has an AR term and MA term on the second difference:

$$\begin{aligned} i)\ W_t &= Y_t - Y_{t-1} \\ ii)\ Z_t &= W_t - W_{t-1} \\ iii)\ Z_t &= c + \phi_1 Z_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t \end{aligned} \tag{8}$$

SARIMA Framework: taking into account seasonality

- In addition to differencing (to remove trend), another common transformation is seasonal differencing.
- This leads to the bigger framework of Seasonal ARIMA (SARIMA).
- The convention is SARIMA$(p, d, q)(P, D, Q, m)$. The second parenthesis refers to the seasonal terms: P is the number of seasonal AR terms, D refers to the degree of seasonal differencing, Q to the number of seasonal MA terms and m the length of the season.

SARIMA Framework: Example

- SARIMA(1,0,1)(1,1,0,12) refers to a process which has an one regular AR term and was seasonally differenced once and has an AR term on the seasonal difference. The length of the season is 12.

This gets quite messy to write in terms of the original series and is very difficult to do without the backshift notation. We'll take a look at next time.

Forecasting in the ARIMA Framework

- Once we pick a model such as ARIMA(1,0,1), the data is estimated from the parameters by Maximum Likelihood Estimation (i.e. find the parameters ϕ_1 and ϵ_1 that would make the observed series most probable).
- This is the hard part of the task but is done by numerical optimization and software has become reliable. item Let us assume that the MLE estimators of the parameters are $\phi_1 = 0.2$ and $\theta_1 = -0.5$.
- The model is then:

$$Y_t = 0.2Y_{t-1} - 0.5\epsilon_{t-1} + \epsilon_t$$

- To 'forecast' from the above process we simply plug in the observed values in the above evolution equation. Assume that $y_{t-1} = 20$

$$\hat{y}_t = 0.2(20) - 0.5(20 - \hat{y}_{t-1})$$

Forecasting in the ARIMA Framework

- Note that $y_1, y_2, ...y_{t-1}$ are observable to us but $\epsilon_1, \epsilon_2, ...\epsilon_{t-1}$ are not observable, we therefore estimate

$$\hat{\epsilon}_{t-1} = y_{t-1} - \hat{y}_{t-1}$$

- If we are forecasting for time $t+1$ using data up to time $t-1$. We proceed with:

where we replaced the observation y_t with its estimator from the model $\hat{y}_t$. Since we have not observed y_t, our best estimator for $\hat{\epsilon}_t = 0$. The multi-step look ahead forecast simply reduces to:

$$\hat{y}_{t+h|t-1} = 0.2\hat{y}_{t+h-1} \text{ for } h = 1, 2, 3...$$

Forecasting in the ARIMA Framework

- Things get messier if we take more complicated models,but the principles are the same. Let us take ARIMA(1,2,1).
- Let us assume that this time the MLE estimators of the parameters are $c = 100$, $\phi_1 = 0.4$ and $\theta_1 = 0.6$.
- The model is then:

$$Z_t = 100 + 0.4Z_{t-1} + 0.6\epsilon_{t-1} + \epsilon_t$$

  where $Z_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$.
- For this model, we need the last three observations $y_{t-1}, y_{t-2}, y_{t-3}$ to forecast for period t.

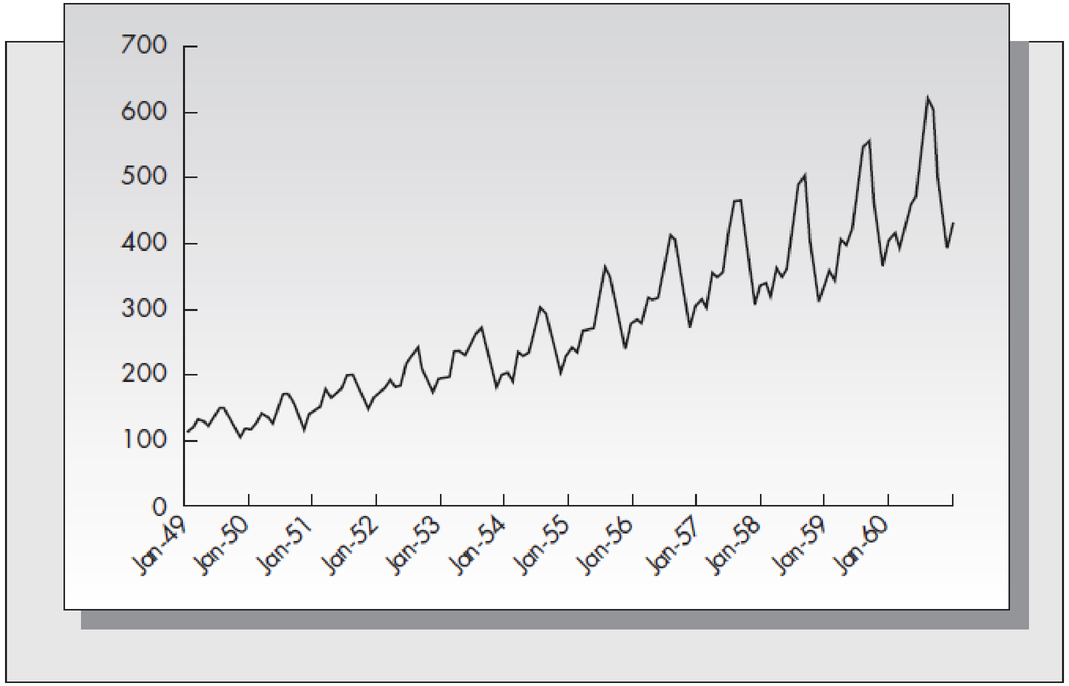
Forecasting in the ARIMA Framework

- With some guidance from our part, software enables us to fit models and perform comparisons across models.
- We can also perform exhaustive searches of a large class of models.
- It is then critical to validate and interpret the results.

Case: Forecasting Airline Passenger Demand

- Airline passenger demand from 1949 to 1960.

FIGURE 2–17
International airline passengers (thousands)

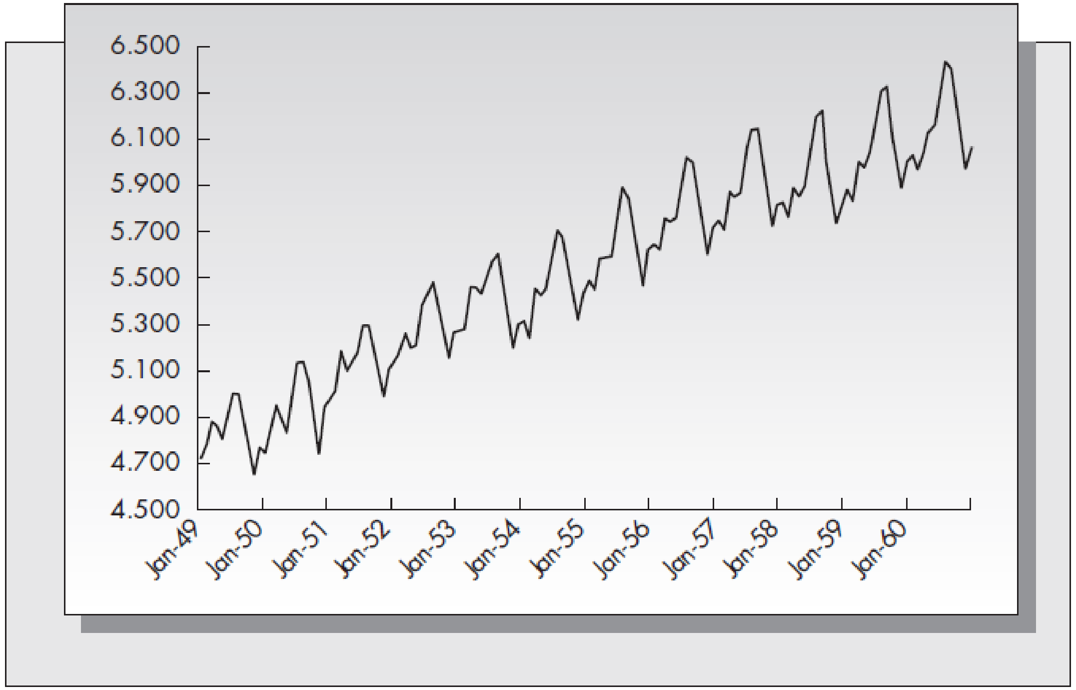


- There is trend and seasonality and increasing variance (fluctuations are increasing).

Step 1: Logarithmic transformation

- To stabilize variance take $Z_t = \ln(Y_t)$

FIGURE 2–18
Natural log of international airline passengers



- The fluctuations now appear stable.
- Trend and seasonality still remain.

Step 2: Detrend and deasonalize by differencing

- First order differencing to remove linear trend.
- 12 month differencing to remove annual seasonality.
- Check autocorrelations after differencing:

TABLE 2–4
Autocorrelations for the Transformed Airline Data Pictured in Figure 2–19 (after taking logarithms and two levels of differencing)

Lag	Autocorrelation	Lag	Autocorrelation	Lag	Autocorrelation
1	−0.34	13	0.15	25	−0.10
2	0.11	14	−0.06	26	0.05
3	−0.20	15	0.15	27	−0.03
4	0.02	16	−0.14	28	0.05
5	0.06	17	0.07	29	−0.02
6	0.03	18	0.02	30	−0.05
7	−0.06	19	−0.01	31	−0.05
8	0.00	20	−0.12	32	0.20
9	0.18	21	0.04	33	−0.12
10	−0.08	22	−0.09	34	0.08
11	0.06	23	0.22	35	−0.15
12	−0.39	24	−0.02	36	−0.01

- Still significant auto-correlation at 1 lag and 12 lag (additional MA terms are needed) but no other significant AC left.

Step 3: Final model and parameter estimation

- Final model (parameters optimized in Statistical software):

$$z_t = z_{t-1} + z_{t-12} - z_{t-13} + \epsilon_t - 0.333\epsilon_{t-1} - 0.544\epsilon_{t-12} + 0.181\epsilon_{t-13}.$$

- And don't forget:

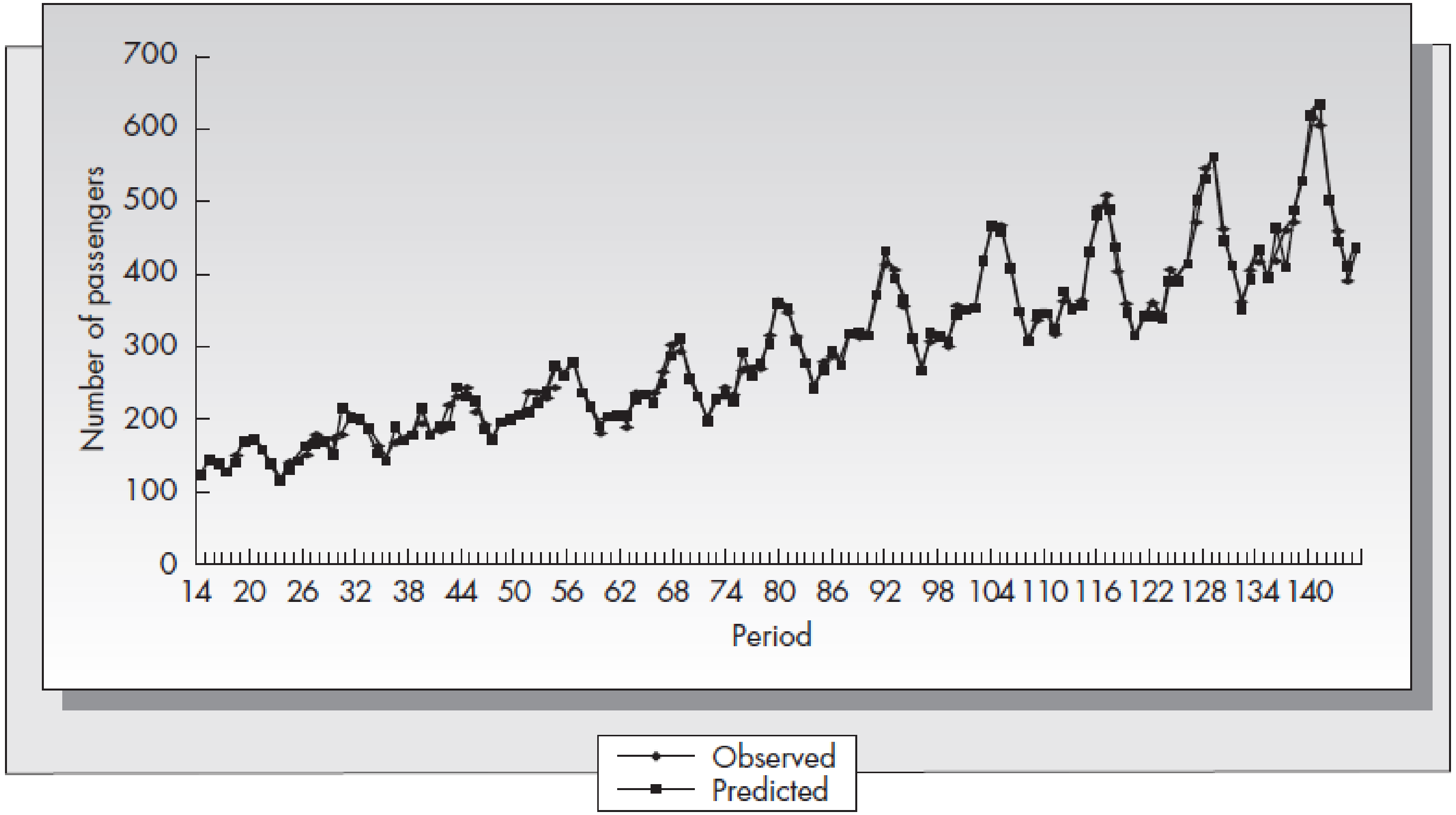
$$Z_t = \ln(Y_t), \quad Y_t = e^{Z_t}$$

- Because of the first order differencing and the yearly differencing, the first forecast can be made for period 14.

Results

- From period 14 on:

FIGURE 2–20
Observed versus predicted number of airline sales



- The resulting forecasts are excellent!