INDR 422/522

Fikri Karaesmen

Spring 2023

ARIMA forecasts - 1

March 16, 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
- First lab available, please take a look and work on the exercises
- Second lab will be available this Friday
- Participation taken. Please participate in polls.
- Please follow announcements

Class Exercise from last lecture

CLASS EXERCISE, March 14, 2023

1. Let X and Y be independent random variables with equal variances $\sigma^2 > 0$.

   (a) $Var(X+Y) =$
   Solution: Since X and Y are independent:
   $$Var(X+Y) = Var(X) + Var(Y) = 2\sigma^2.$$

   (b) $Var(X-Y) =$
   Solution: Since X and Y are independent:
   $$Var(X-Y) = Var(X) + (-1)^2 Var(Y) = 2\sigma^2.$$

   (c) $Var(X+X) =$
   Solution: We can have two different arguments. The first one is direct:
   $$Var(X+X) = Var(2X) = (2)^2 Var(X) = 4\sigma^2.$$
   The other one is more general. We note that X and X are not independent (in fact $Corr(X,X) = 1$). Then:
   $$Var(X+X) = Var(X) + Var(X) + 2Cov(X,X)$$
   and
   $$Cov(X,X) = E[XX] - E[X][X] = Var(X) = \sigma^2$$
   Therefore, $Var(X+X) = 2\sigma^2 + 2\sigma^2$

   (d) $Var(X-X)$
   Solution: Once again, let us review two different arguments. The first one:
   $$Var(X-X) = Var(0) = 0.$$
   For the second one, we note that X and $-X$ are not independent (in fact $Corr(X,X) = -1$). Then:
   $$Var(X-X) = Var(X)+Var(X)+2Cov(X,-X) = 2\sigma^2-2\sigma^2 = 0.$$

Class Exercise from last lecture

2. Assume that $Y_t = f(t) + \epsilon_t$ where ϵ_t are i.i.d random variables with mean zero and variance σ^2. Let $Z_t = Y_t - Y_{t-1}$. Which of the following are unbiased estimators for $f'(t)$? (*Note: This is a badly posed question because $f'(t)$ is not random but we can give an approximate answer.*)

   Solution: We note that

   $$E[Z_t] = f(t) - f(t-1)$$

   In general, $f(t+1) - f(t) \neq f'(t)$

   For instance let us take:

   $$f(t) = a_2 t^2$$

   $$E[Z_t] = a_2 t^2 - a_2(t-1)^2 = 2a_2 t - a_2 \neq f'(t)$$

3. Let $f(t) = a_2 t^2$. Which one is an unbiased forecast for y_{t+1}? Recall that $f(x+h) = f(x) + f'(x)h + f''(x)h^2/2 + \ldots.$

   Solution: We can follow the Taylor expansion and guess:

   $$\hat{Y}_{t+1} = y_t + (y_t - y_{t-1}) + ((y_t - y_{t-1})) - ((y_{t-1} - y_{t-2}))/2.$$

   We can then check:

   $$\begin{aligned} E[\hat{Y}_{t+1}] &= a_2 t^2 + (a_2 t^2 - a_2(t-1)^2) \\ &\quad +((a_2 t^2 - a_2(t-1)^2) - ((a_2(t-1)^2 - a_2(t-2)^2)/2 \\ &= a_2 t^2 + 2(2a_2 t - a_2) - (2a_2(t-1) - a_2) \\ &= a_2 t^2 + 2a_2 t + a_2 \\ &= a_2(t+1)^2 \end{aligned}$$

   Therefore, $E[\hat{Y}_{t+1}] = a_2(t+1)^2 = E[Y_{t+1}]$.

Summary last lecture

- Data transformations enable simplification of complicated patterns
- Some examples:
  - $z_t = d_t - d_{t-1}$
  - $z_t = d_t - d_{t-12}$
  - $w_t = (d_t - d_{t-1}) - (d_{t-1} - d_{t-2})$
  - $z_t = \sqrt{d_t}$
  - $z_t = \log(d_t)$
- Most series can be transformed to a stationary series after such simple transformations

Next: correlation structure

- The models for time-series we have considered so far were of the type: $D_t = \mathrm{f(t)} + \epsilon_t$
- We assumed that ϵ_t are independent. Therefore, if the functional form of $\mathrm{f(t)}$ is known, there is no correlation in the model and D_t are independent across periods.
- But real data frequently has significant short-term correlation.
- We can potentially improve forecasts if we can model the correlation structure.

Correlation Across Time: ARIMA forecasts

- Consider two jointly distributed random variables X and Y with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2 . Recall that:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$$

and the corresponding normalized measure:

$$Corr(X, Y) = \rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

- The sign indicates the direction of the relationship and the absolute value corresponds the strength of the relationship.

Covariance and Correlation

- Sample estimators:

  - Given two samples $x_1, x_2, \ldots x_n$ and $y_1, y_2, \ldots y_n$ we can obtain an estimator for the correlation:

$$r_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

  - Note that the observations are paired: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

Auto-Correlation

- Sample estimators:

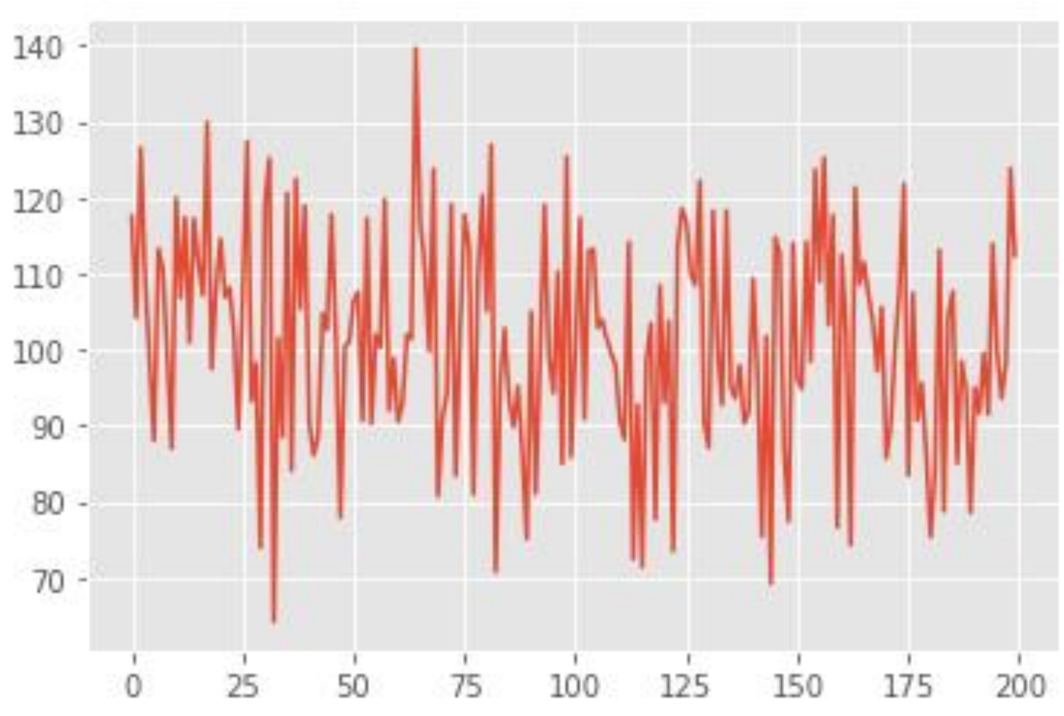
  - For our purposes, we will be interested in the auto-correlation of the process that generates demand: for instance $Corr(Y_t, Y_{t-1})$ or $Corr(Y_t, Y_{t-k})$. This looks at the correlation between demand observation separated by k periods (how demand from k periods ago affects the demand today).
  - 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:

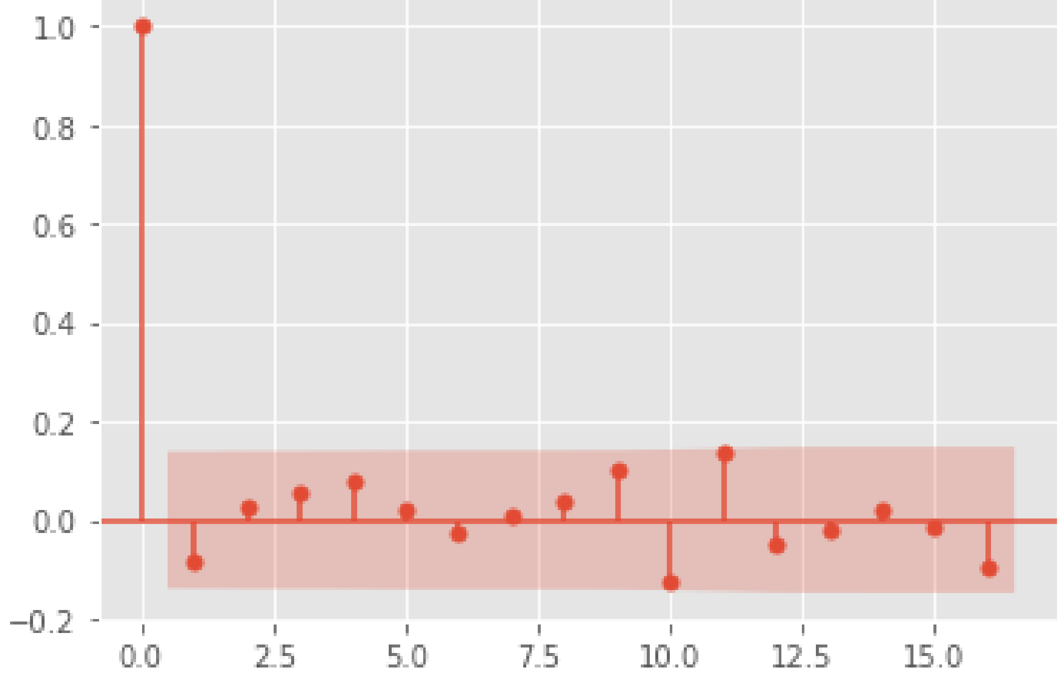
$$r_k = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

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

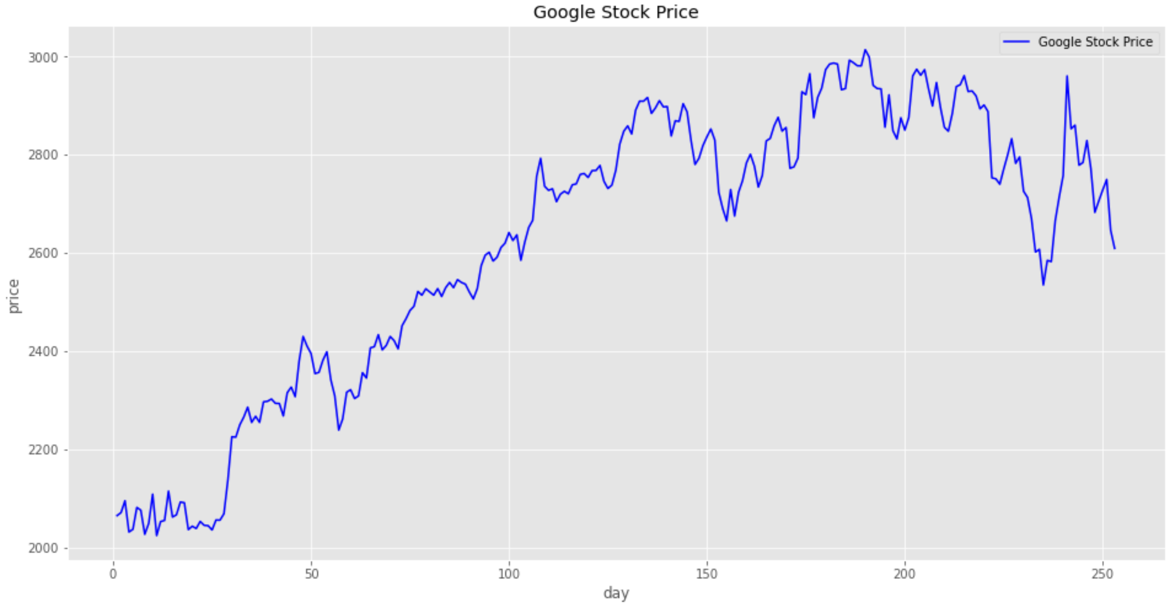
- The effect of patterns:

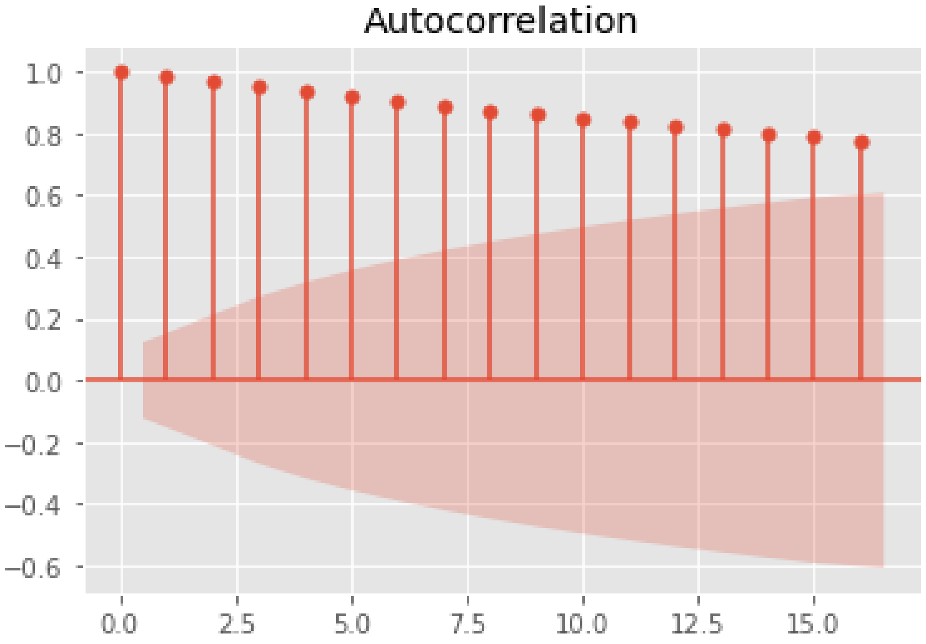
  - We would like to explore the auto-correlation structure of the demand time series to construct models that can take into account the dependence explicitly.
  - First, a relatively trivial observation. All basic patterns in the data (trend, seasonality) etc. reflect onto the autocorrelation structure.
  - 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:

Daily Closing Price of Google – Alphabet Stock

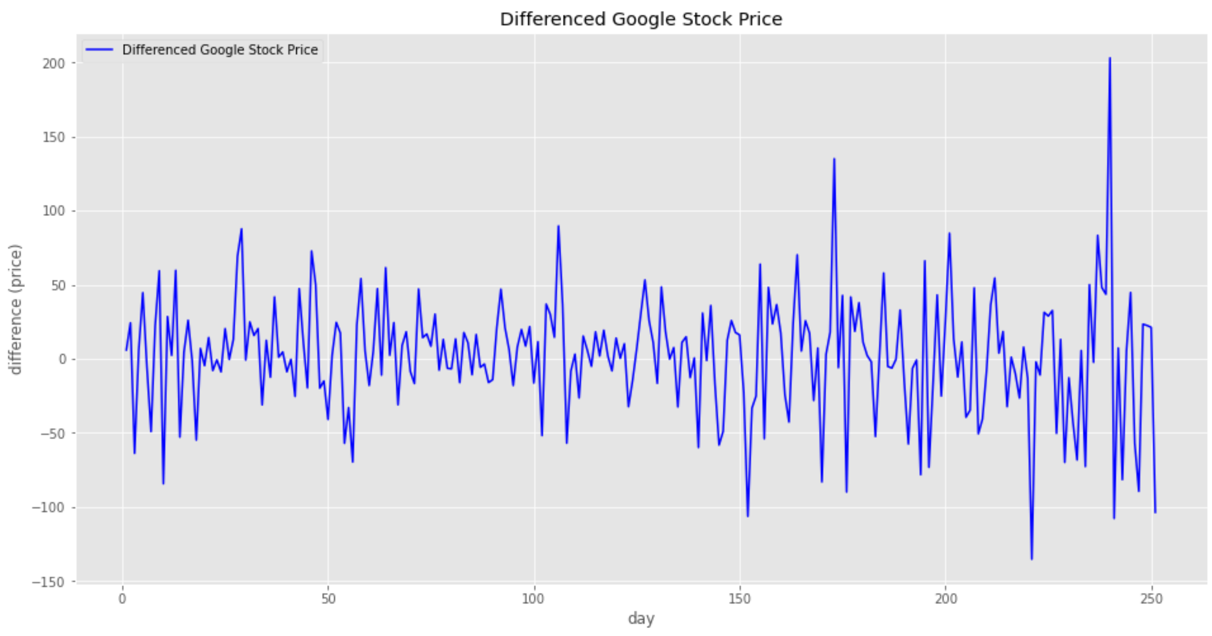


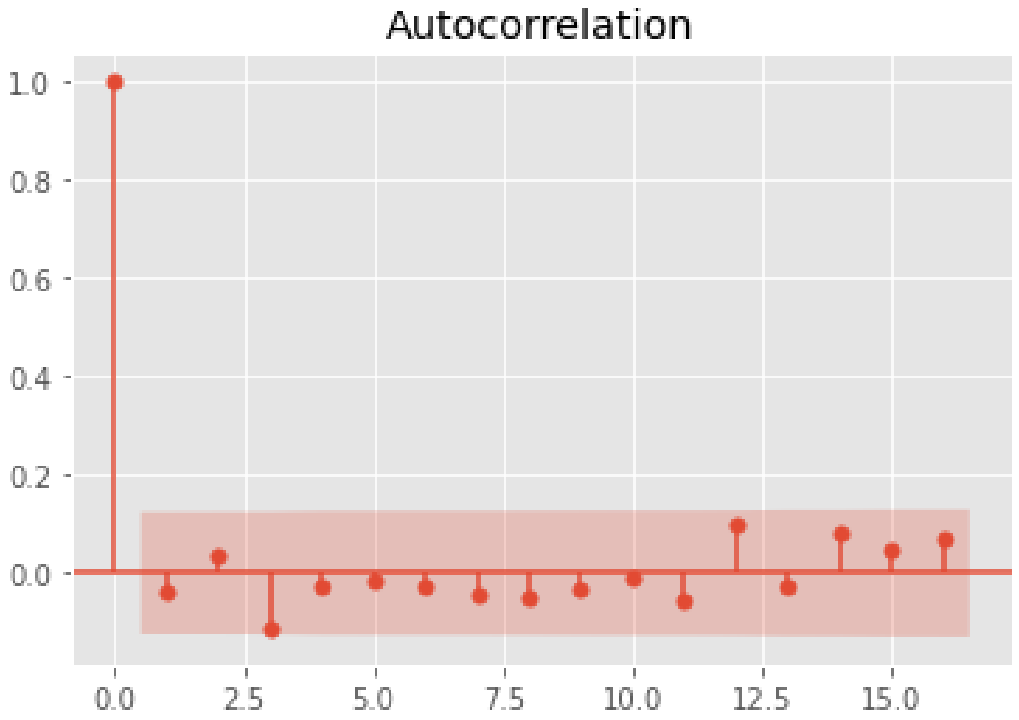


Auto-Correlation: The effects of patterns

- Let’s detrend by taking differences

Differenced Daily Closing Price of Google – Alphabet Stock



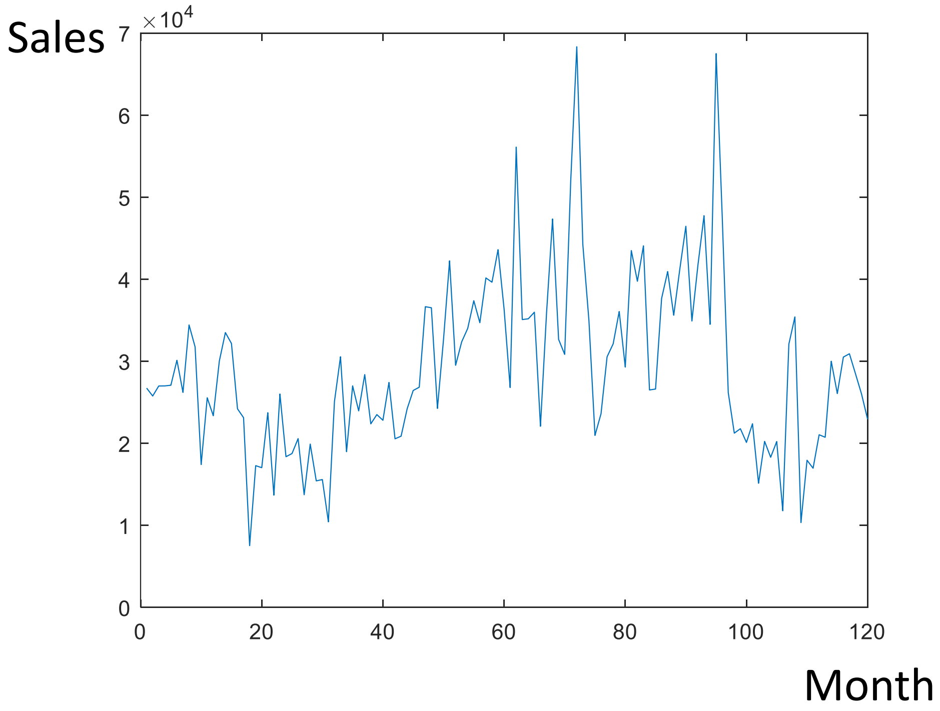


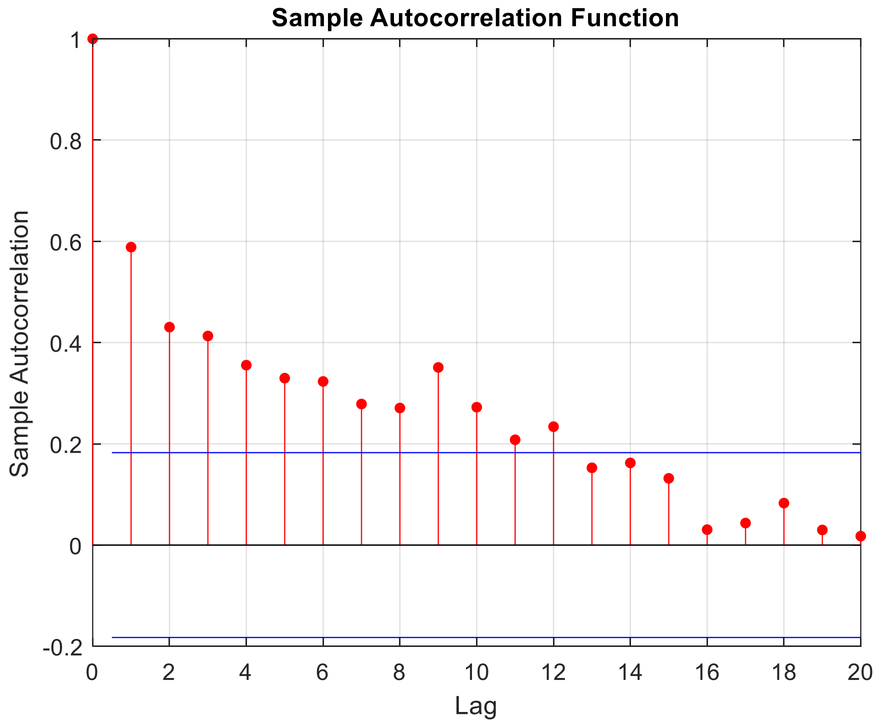
This is usually the case with time series for stock markets, after a transformation there is little AC left.

Auto-Correlation: The effects of patterns

- The effect of patterns:

Domestic Dishwasher Sales

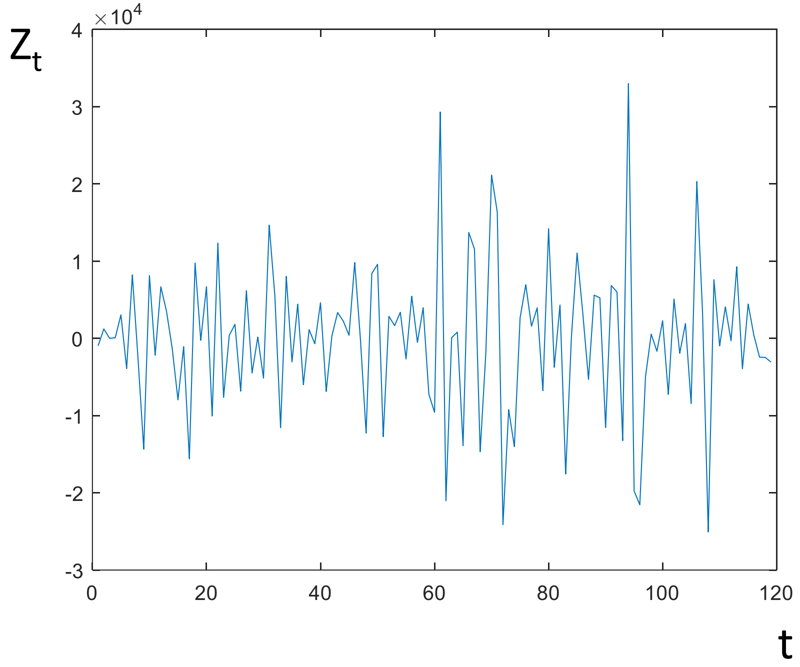


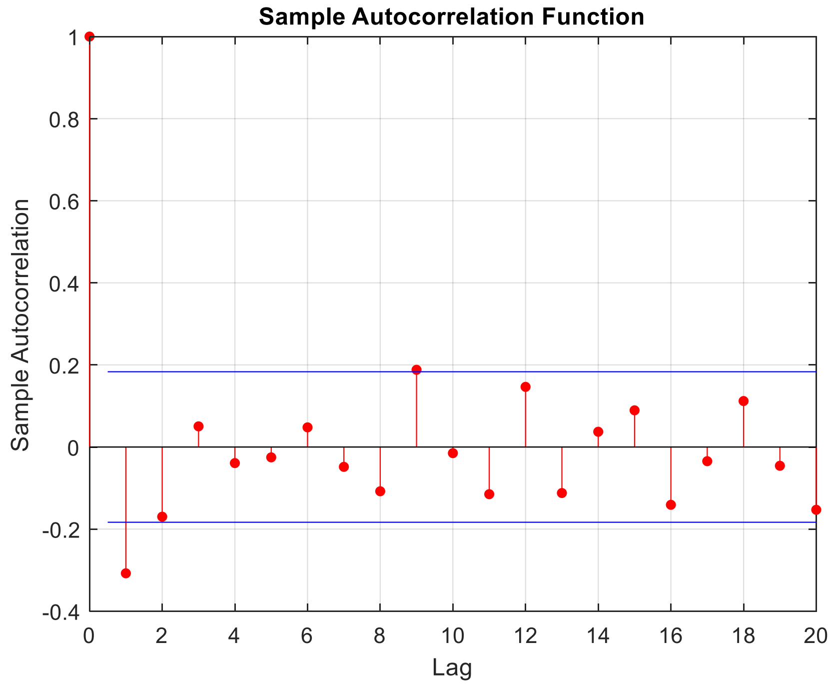


Auto-Correlation: The effects of patterns

- Let's detrend by taking differences

Differenced Sales: $Z_t = Y_t - Y_{t-1}$



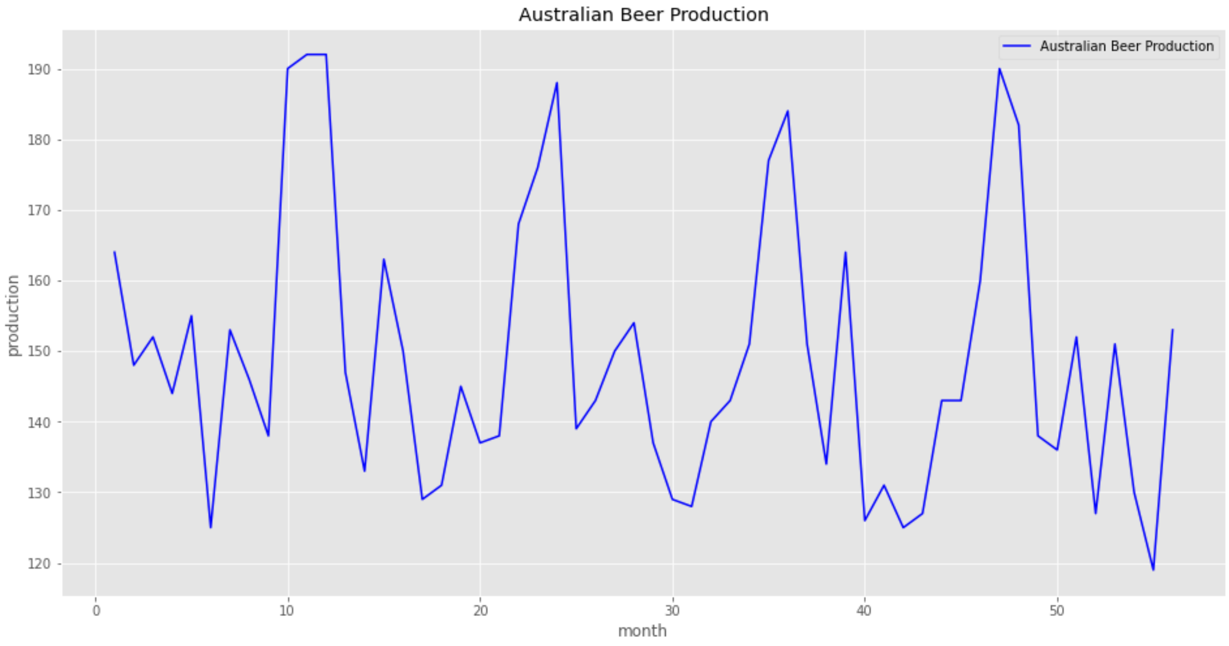


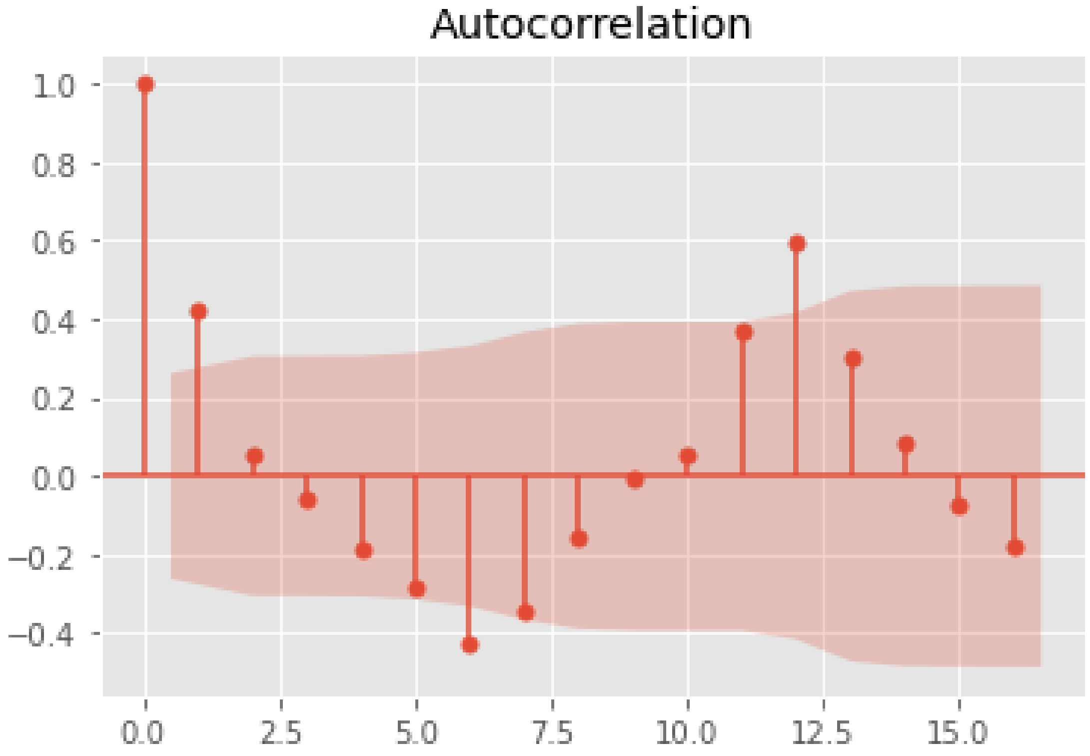
Negative Auto-Correlation in Lag 1!

Auto-Correlation: The effects of patterns

- The effect of patterns:

Australian Monthly Beer Production

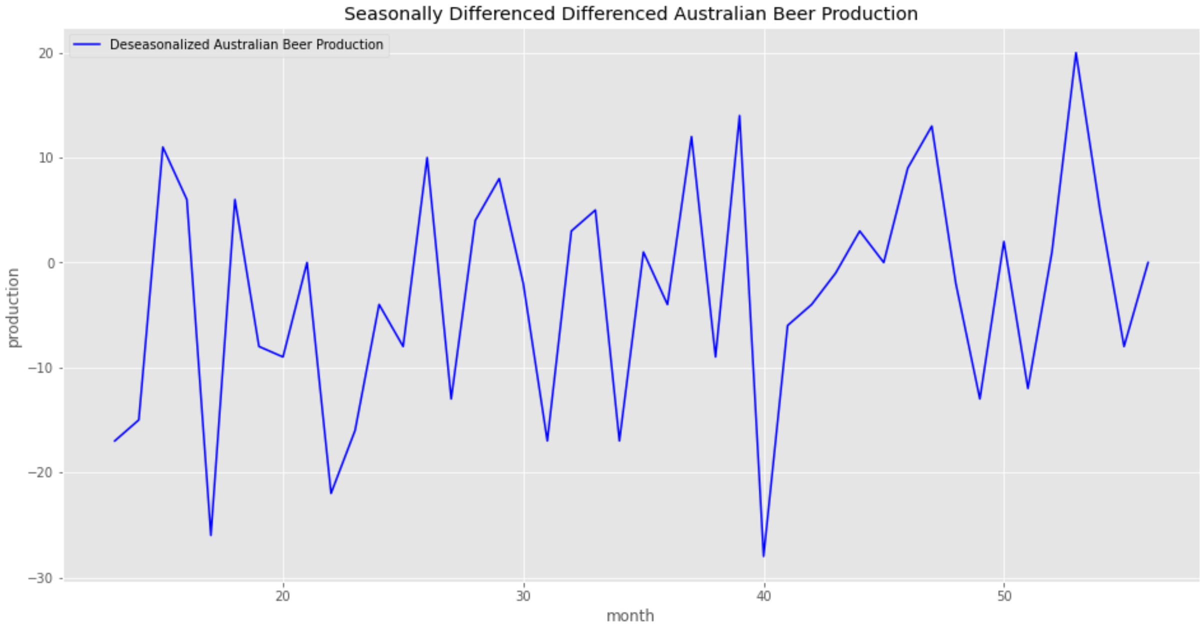


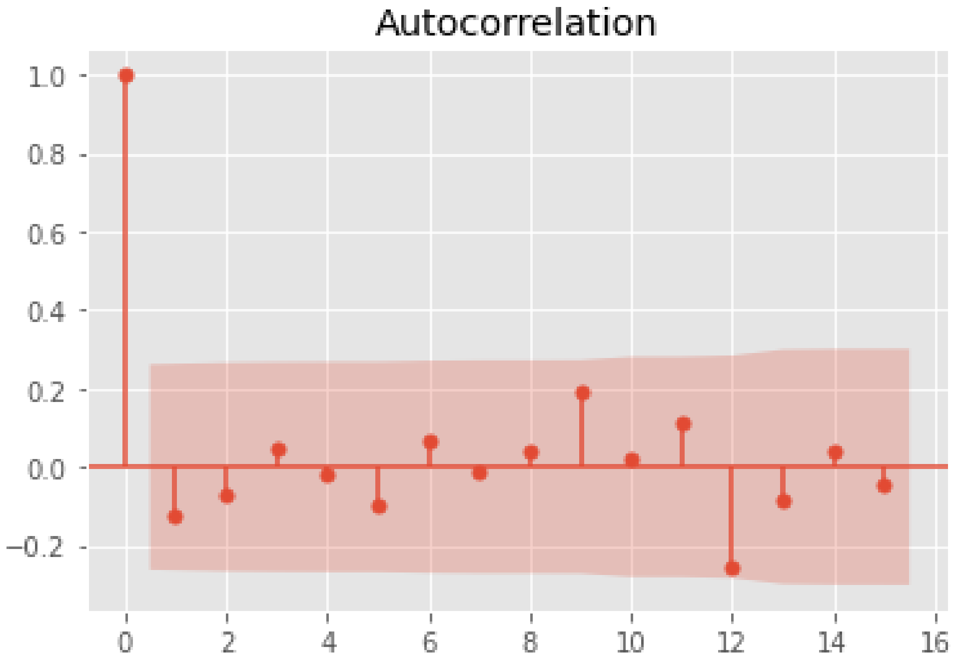


Auto-Correlation: The effects of patterns

- The effect of patterns: deseasonalized Australian Beer Production

Deseasonalized Australian Monthly Beer Production





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

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

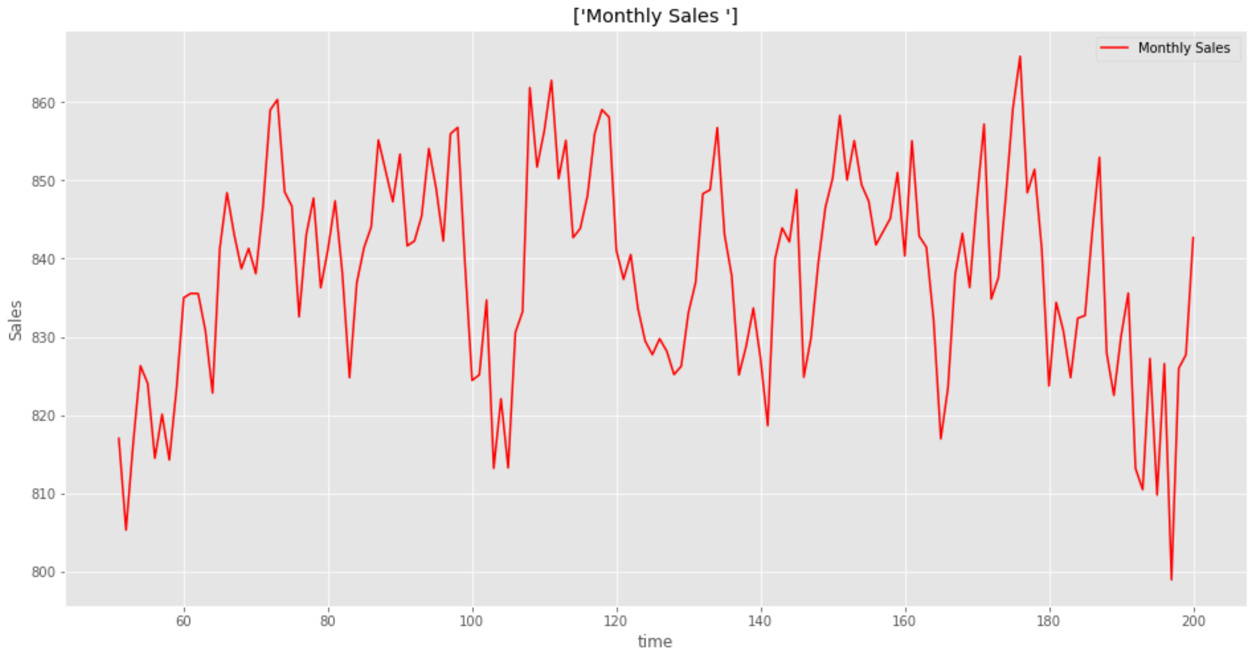
- We can already figure out some of the basic properties. First, we have to have the AR coefficient: $-1 < \phi_1 < 1$, otherwise the series would diverge (in expectation). Note that for general AR-p processes the stability conditions for the parameters are more complicated (please see Hyndman and Athansapoulos, 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 $Corr(Y_{t-k}, Y_t) = \phi_1^k$.

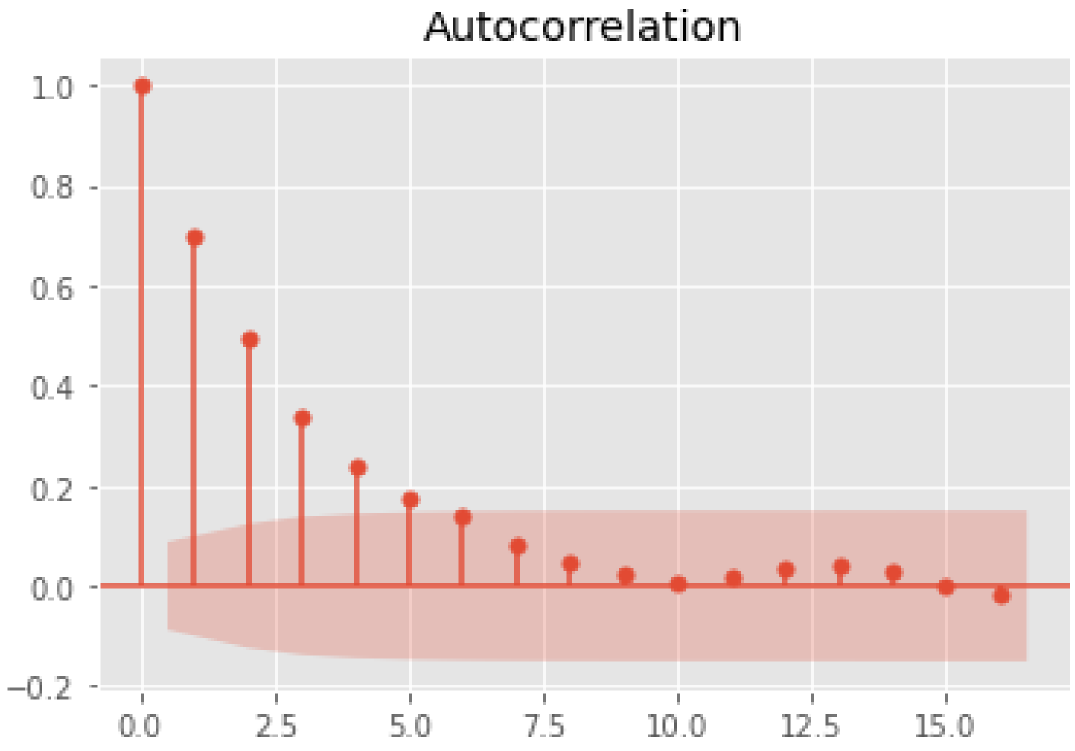
Auto-Regressive models: AR(1)

- 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.

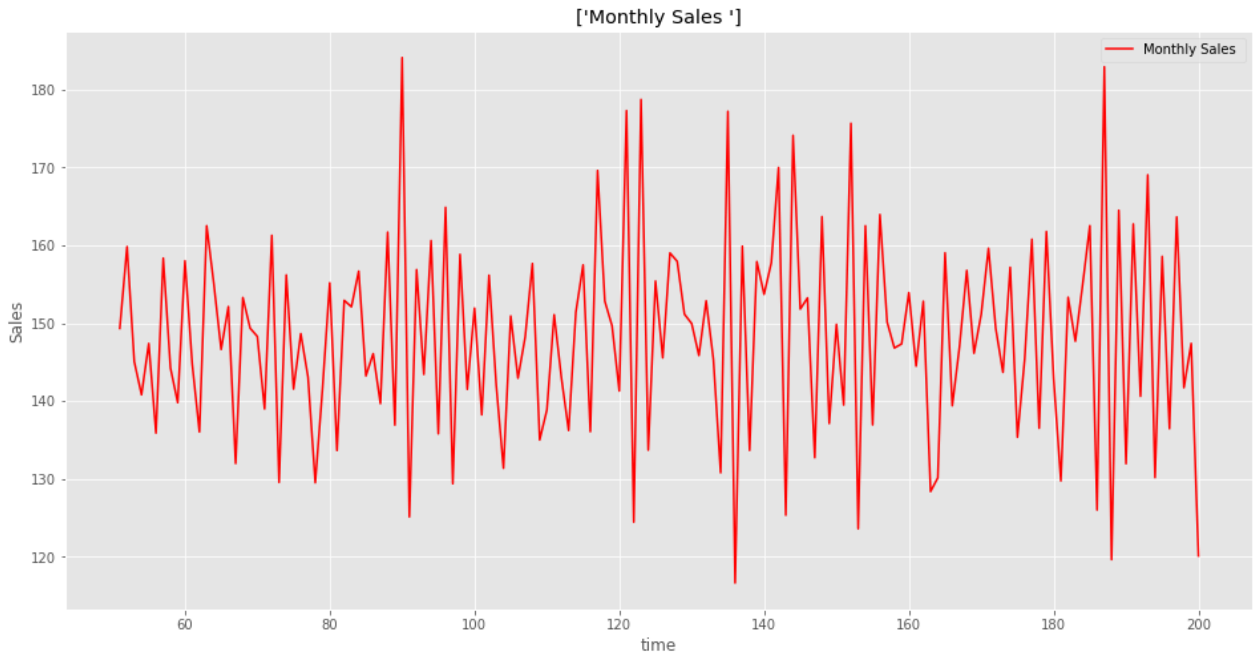


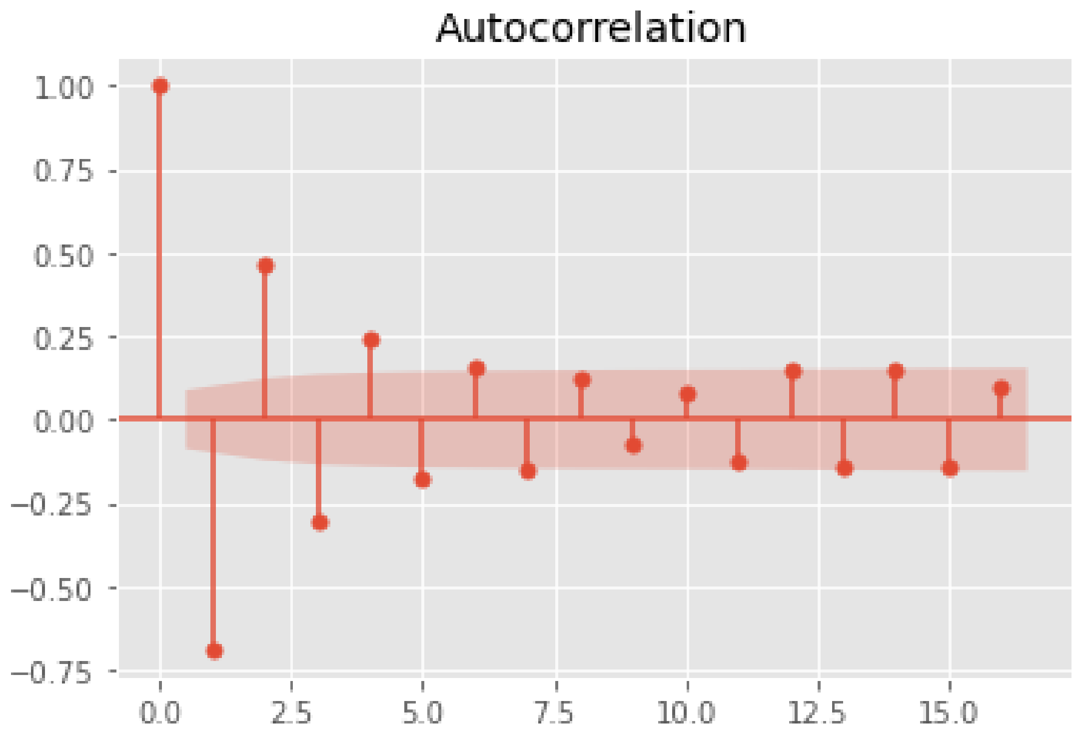


Note the geometric decrease in the AC's starting from lag 1.

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.

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

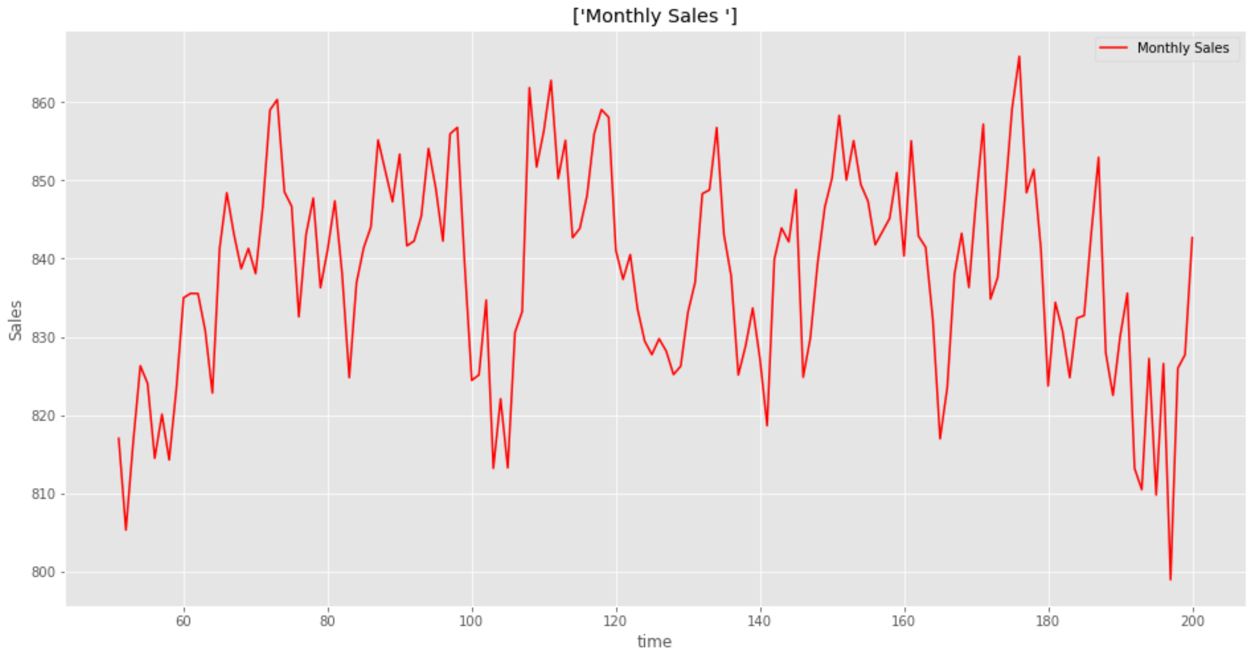
$$Y_t = c + \theta_1 \epsilon_{t-1} + \epsilon_t$$

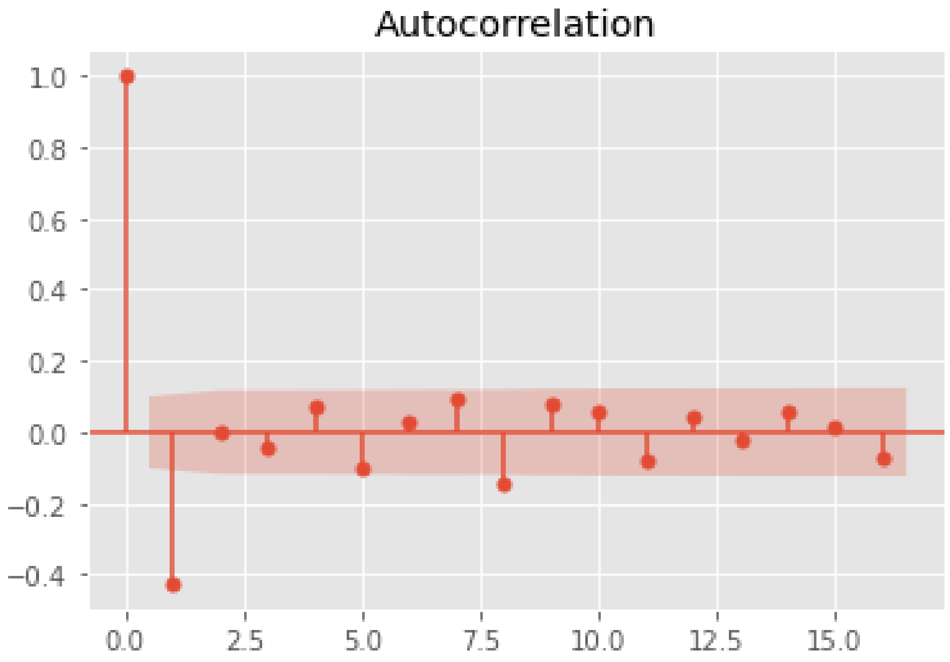
- First, for invertibility (please see Hyndman and Athansapoulos, Chapter 8). we need $|\theta_1| < 1$. Once again for general MA-q processes the invertibility conditions for the parameters are more complicated
- Next, we can verify that Y_{t-1} and Y_t are correlated but Y_{t-2} and Y_t are not. Therefore,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





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.

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

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