Advanced Analytical Theory and Methods: Time Series Analysis

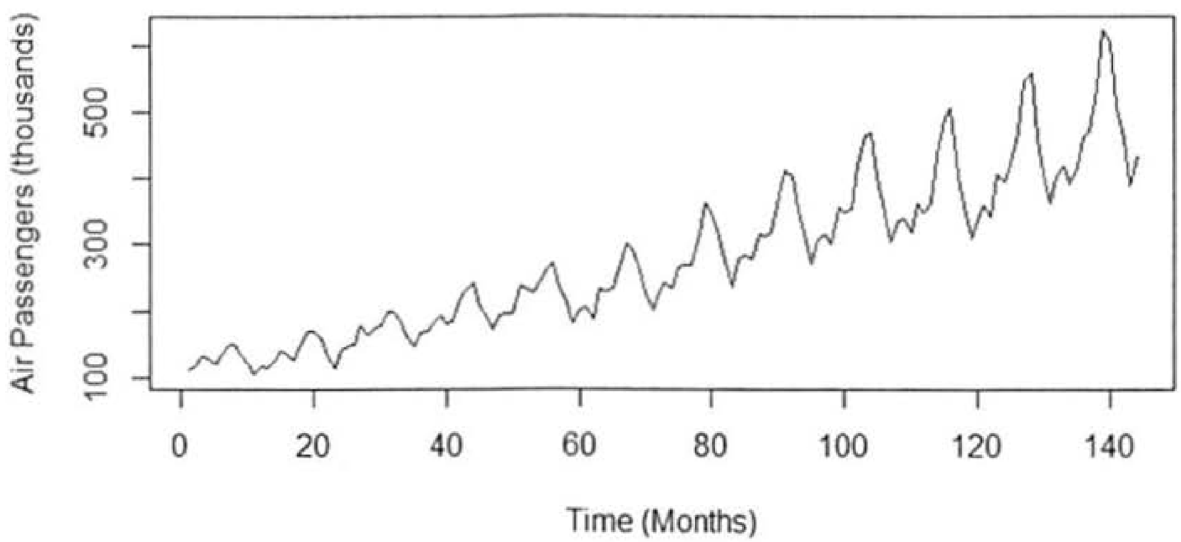
Prof. Dr. Shamim Akhter

Professor, Dept. of CSE

Ahsanullah University of Science and Technology

Overview of Time Series Analysis

- Uncover patterns in historical data that change over time.
  - model the underlying structure of observations
  - Example: monthly unemployment rates, daily website visits, or stock prices every second.



The goals of time series analysis:

- Identify and model the structure of the time series.
- Forecast future values in the time series.

Applications of Time Series Analysis

Components of Time Series

1) **Trend:** A trend is nothing but a movement to relatively higher or lower values over a long period.

- When time series analysis shows a general pattern that is upward, we call it an Up-Trend and when it exhibits a lower pattern, we call it a Down-Trend. Whenever there is no trend or a straight line, we call it a horizontal trend.
- You opened a hardware shop over there and now at the beginning, everyone will buy hardware. So, the sales of the shop are high or we can say the trend is high. But after some time, when everyone has their hardware, the trend may go down.

2) **Seasonality:** It's a **repeating pattern within a fixed period**.

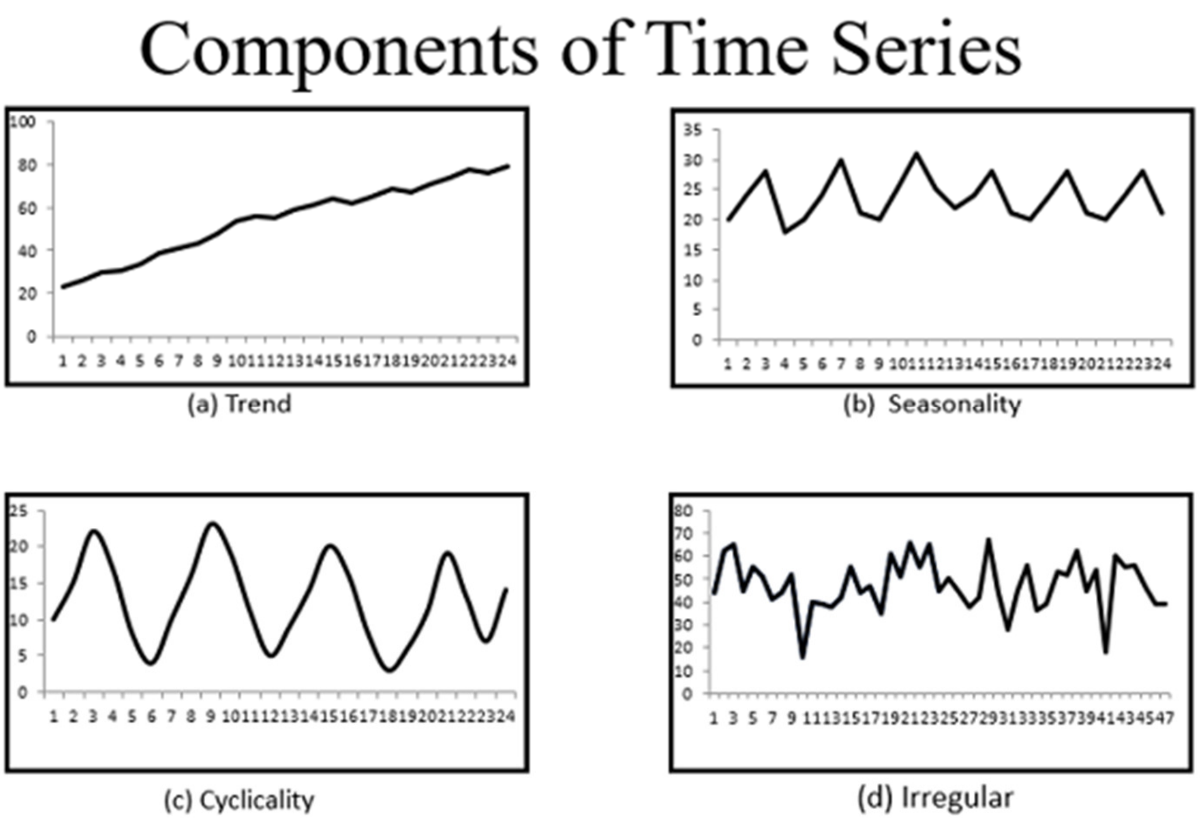
Elish
icecream

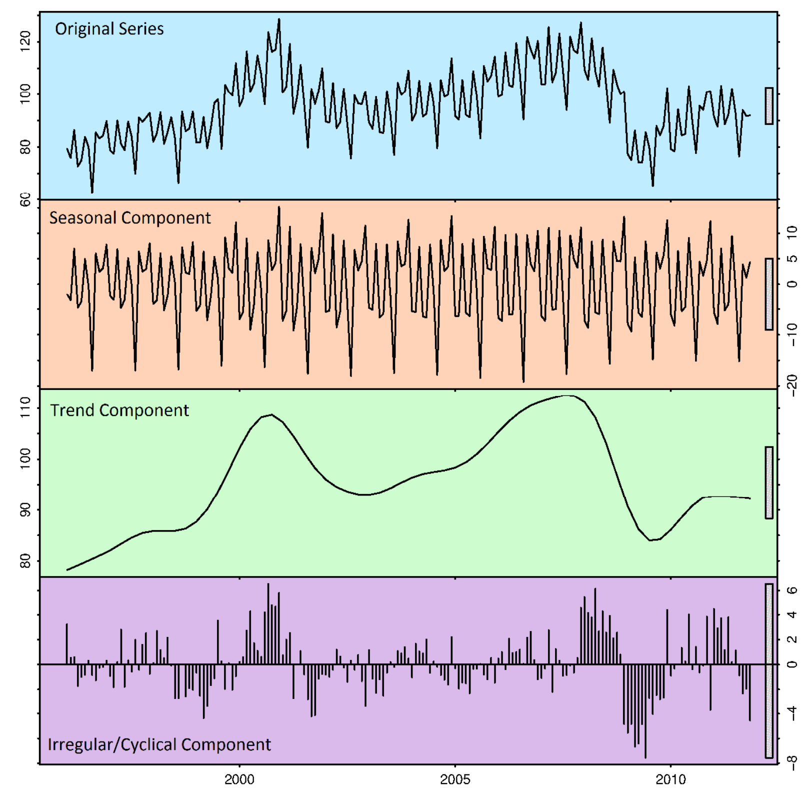
- For example, **Jatka Conservation Week** is celebrated all over Bangladesh in the months of either **March or April**. Now, the sales of Hilsha in these months are very low as compared to other months of the year. This has been noticed for the past two years, five years, ten years, and so on, so it's a repeating pattern within a fixed period, while in trend this is not the case.
- Taking one more example of ice cream, the sales of ice cream go comparatively higher in summers than in winters, so this is a seasonality again.

3) **Irregularity:** This is also known as noise or irrelevant data. It is inconsistent in nature, or, unsymmetric. Irregularity typically occurs for a brief period and does not repeat.

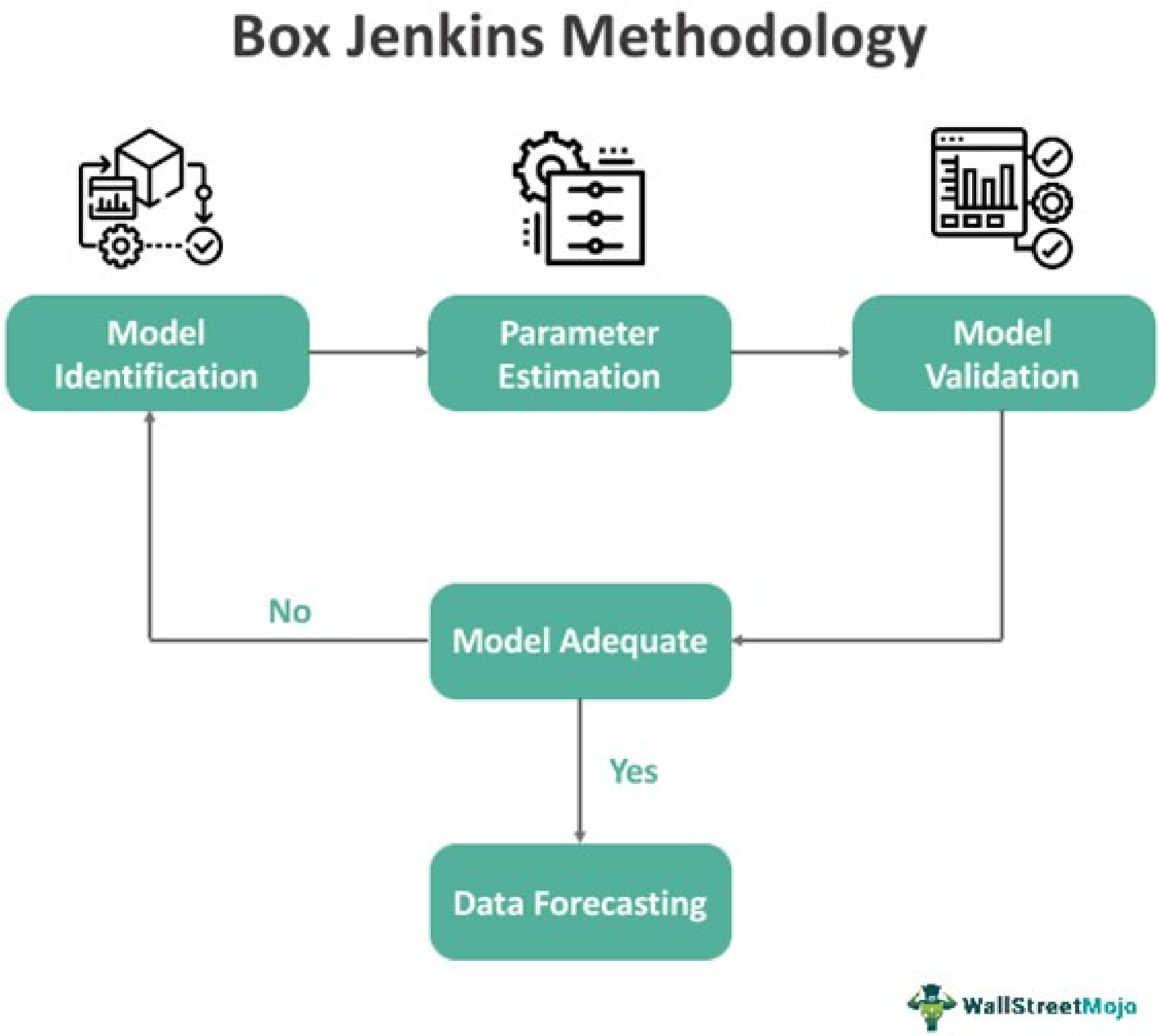
- For example, COVID-19 emerged suddenly within a decade. During the COVID pandemic, sales of sanitizers and masks were high, but after some time, these products have become less common. So, this is all happening erratically. You don't know how many sales will occur, so this represents random variation, which is known as an irregularity.

4) **Cyclic:** It is repeating up and down movements, so this means we can go over more than a year. Cyclic does not have any fixed patterns. They can happen anytime, like in a year in a decade, or maybe within six months. They keep on repeating and as a result, they are much harder to predict.





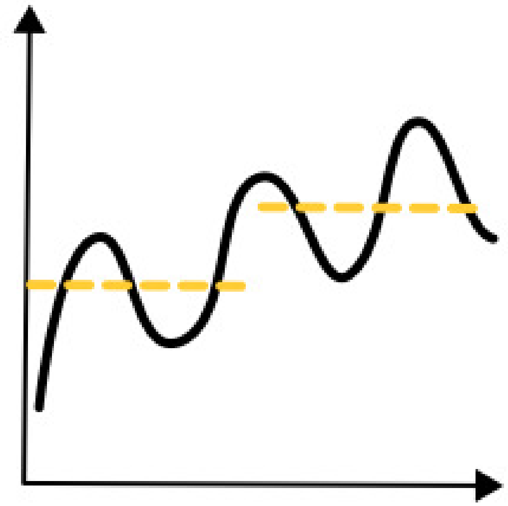
Box-Jenkins Methodology



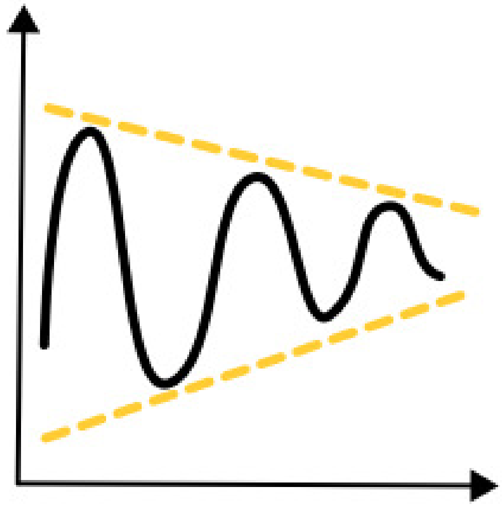
ARIMA Model

- Autoregressive Integrated Moving Average(ARIMA)
  - A statistical model is autoregressive if it predicts future values based on past values.
- The first step:
  - to remove any trends or seasonality from the time series so that achieves certain properties to which autoregressive and moving average models can be applied.
  - Stationary Time Series Y_t for t=1,2,3…if the following three(3) conditions are met:
    - The expected value (mean) of Y_t, is a constant for all values of t.
    - The variance of Y_t is finite.
    - The covariance of Y_t, and Y_{t+h}, depends only on the value of h = 0, 1, 2,... for all t.

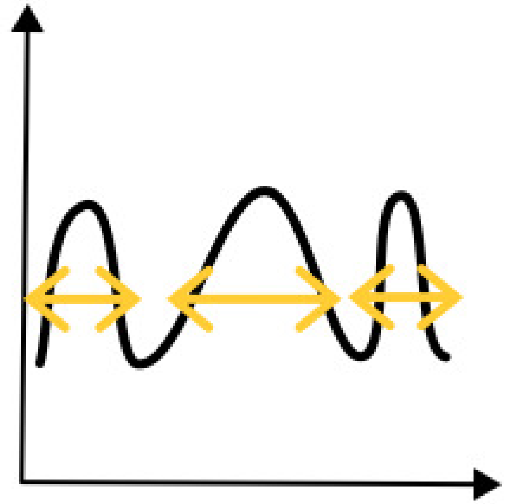
The mean, variance, and covariance are not time-dependent.



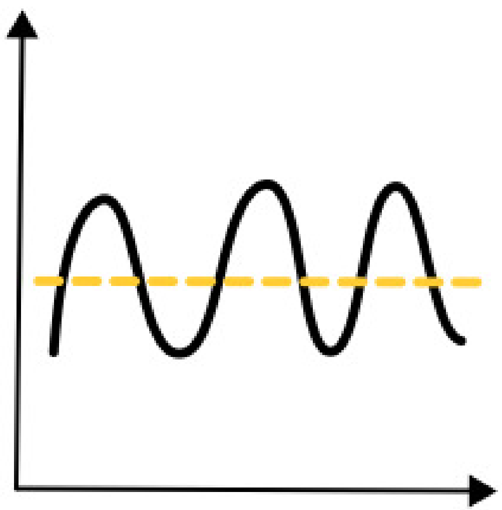
Mean dependent on time



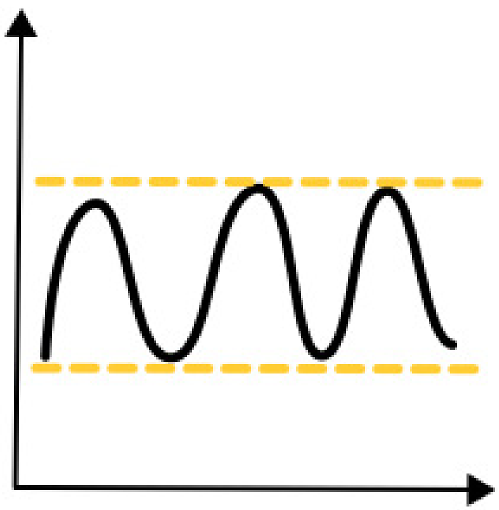
Variance dependent on time



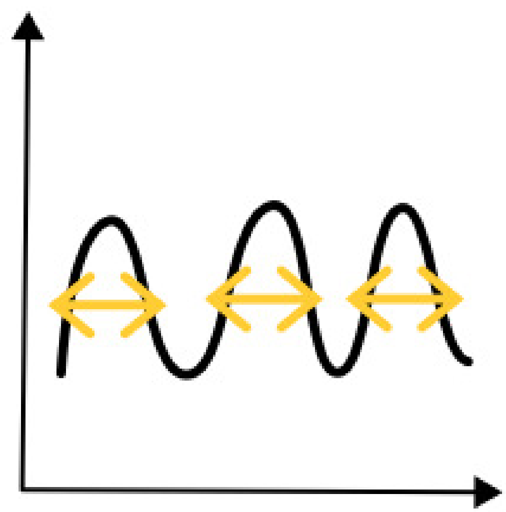
Covariance dependent on time



Mean independent on time



Variance independent on time



Covariance independent on time

Converting Non-Stationary Into Stationary

- There are **three methods available** for this conversion – detrending, differencing, and transformation.

- **Detrending**: Sometimes we want to analyze the data without the trend to capture other patterns like seasonality or noise. In these situations, detrending can help us.
  - We assume the time series is an additive combination of a trend and other components.

$$Y(t)=T(t)+S(t)+e(t)$$

Here:

t is time

$Y(t)$ is the original time series

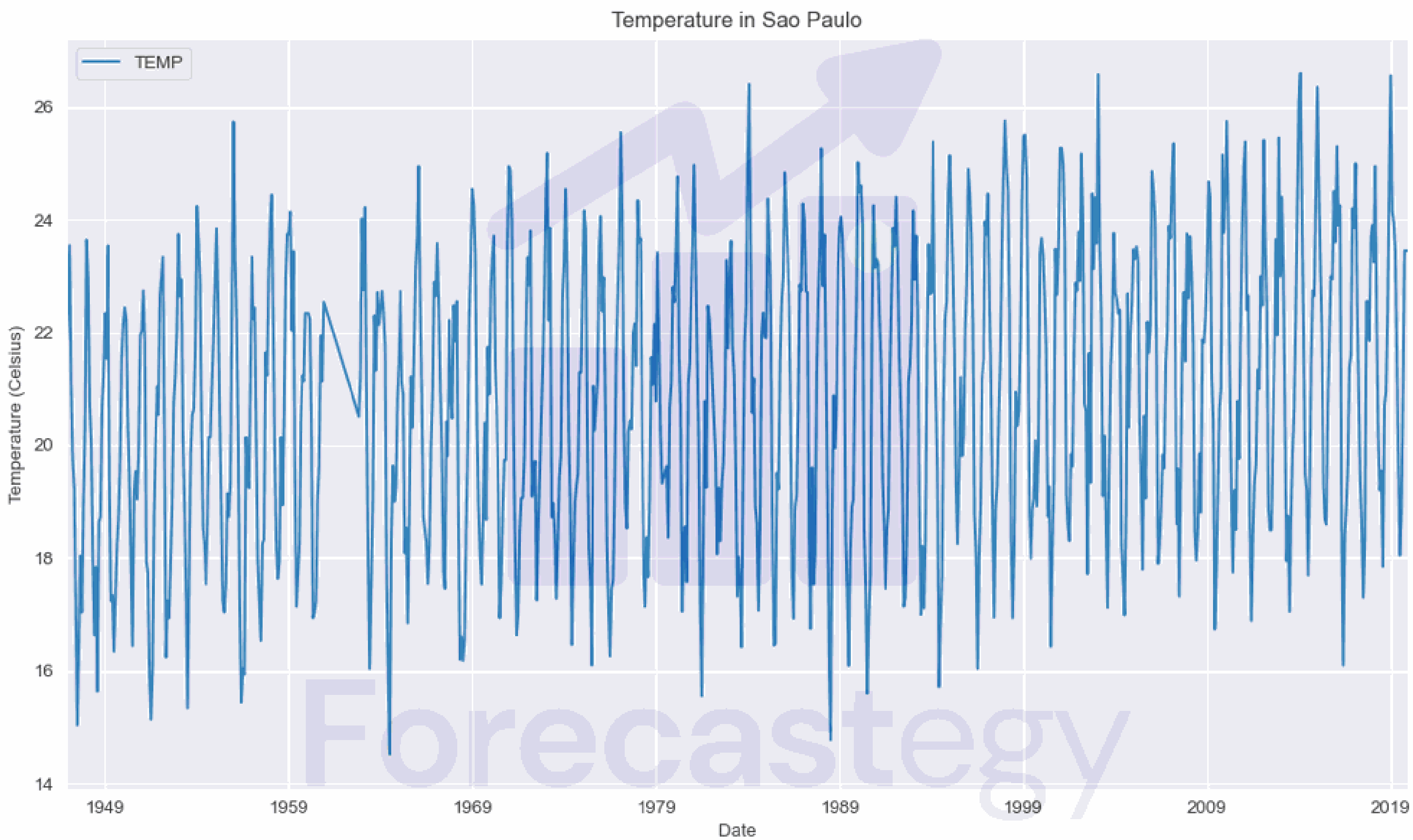
$T(t)$ is the trend

$S(t)$ is the seasonal component

$e(t)$ is the noise

Monthly historical temperature in Sao Paulo, Brazil

TIMESTAMP	TEMP
1980-12-01 00:00:00	22.63
1952-10-01 00:00:00	20.74
1991-01-01 00:00:00	23.3
2012-04-01 00:00:00	22.45
2018-02-01 00:00:00	23.25



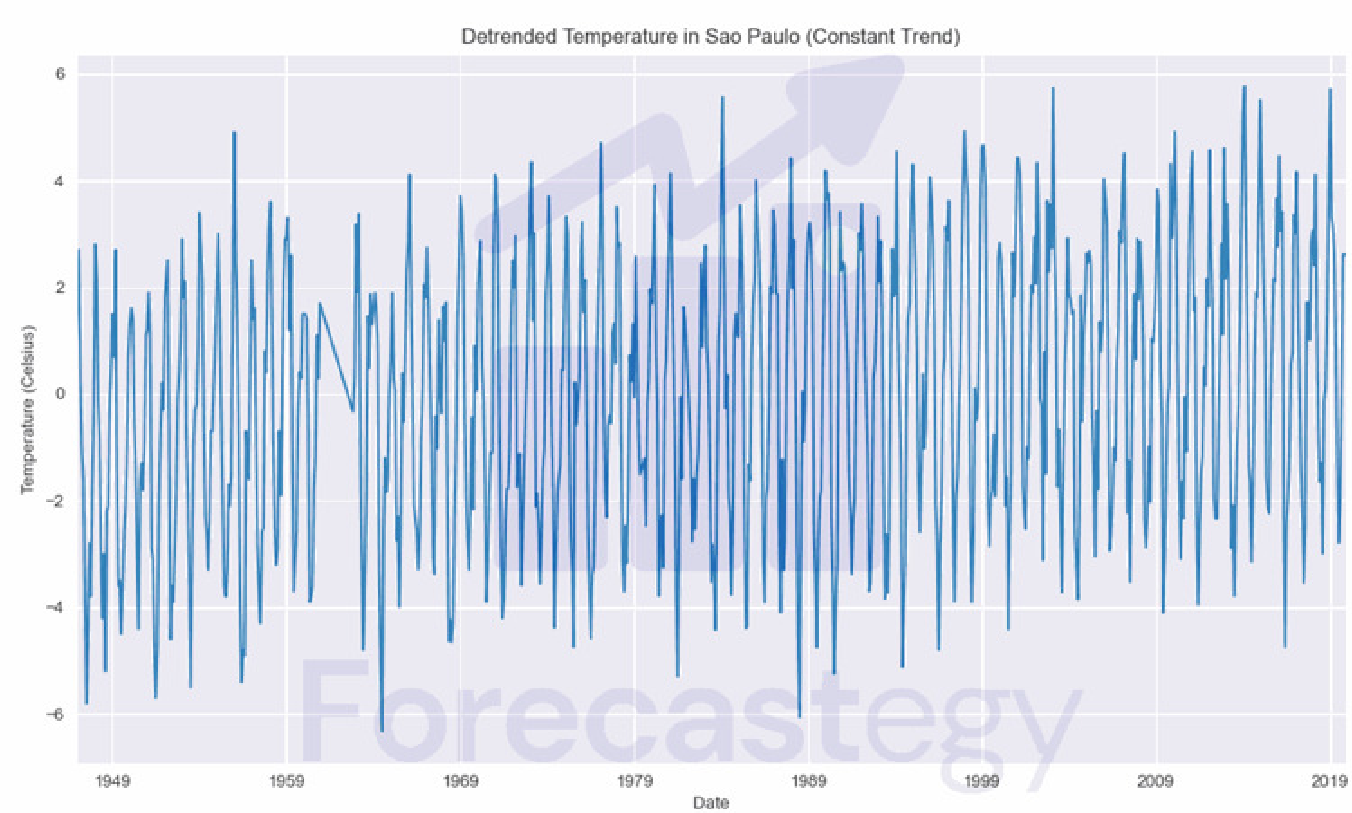
Monthly historical temperature in Sao Paulo, Brazil

https://www.kaggle.com/datasets/volpatto/temperature-timeseries-for-some-brazilian-cities?select=station_sao_paulo.csv

Detrending With A Constant Model (Scipy)

- The simplest way to detrend a time series is by subtracting the mean value of the data.
- This is called a constant model, and it assumes that the trend of the time series is a straight horizontal line.

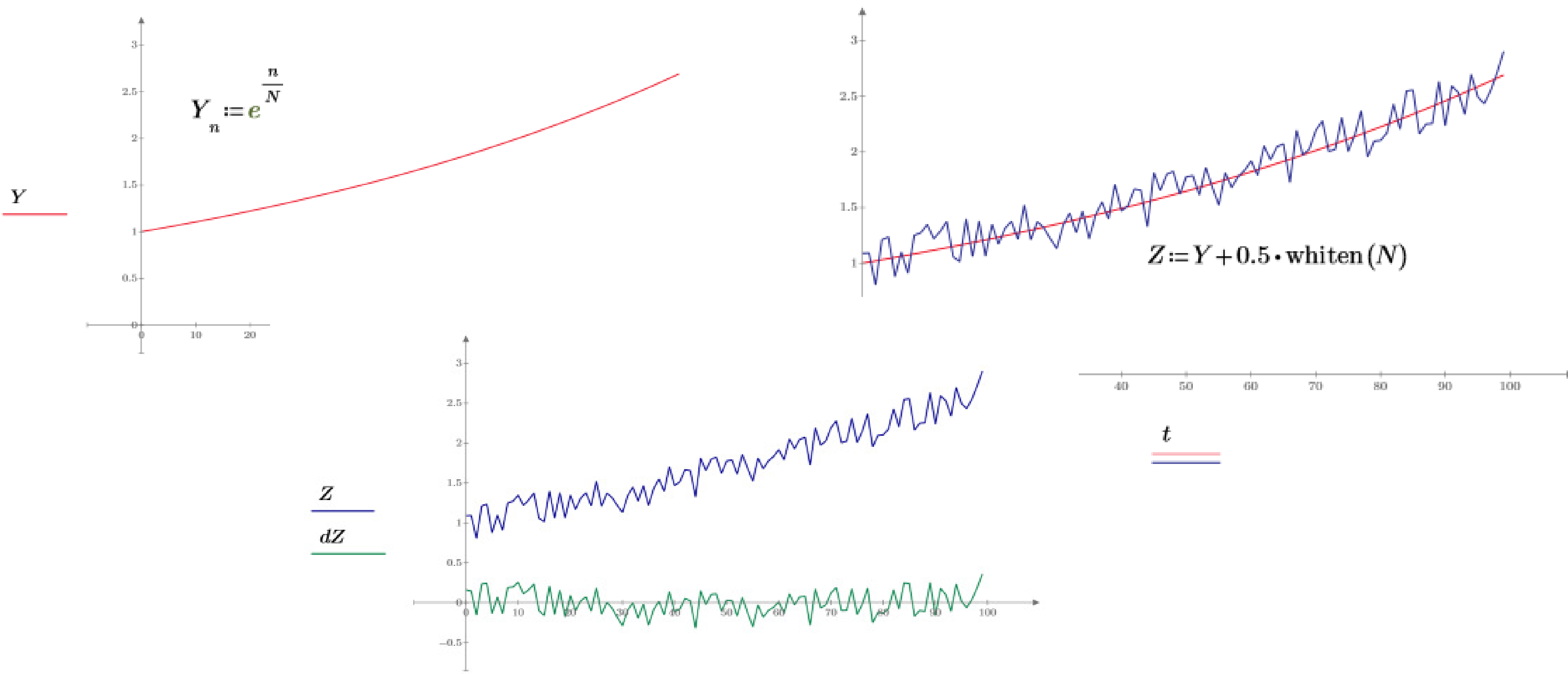
```
from scipy.signal import detrend
detrended = detrend(data['TEMP'], type='constant')
detrended = pd.Series(detrended, index=data.index)
```

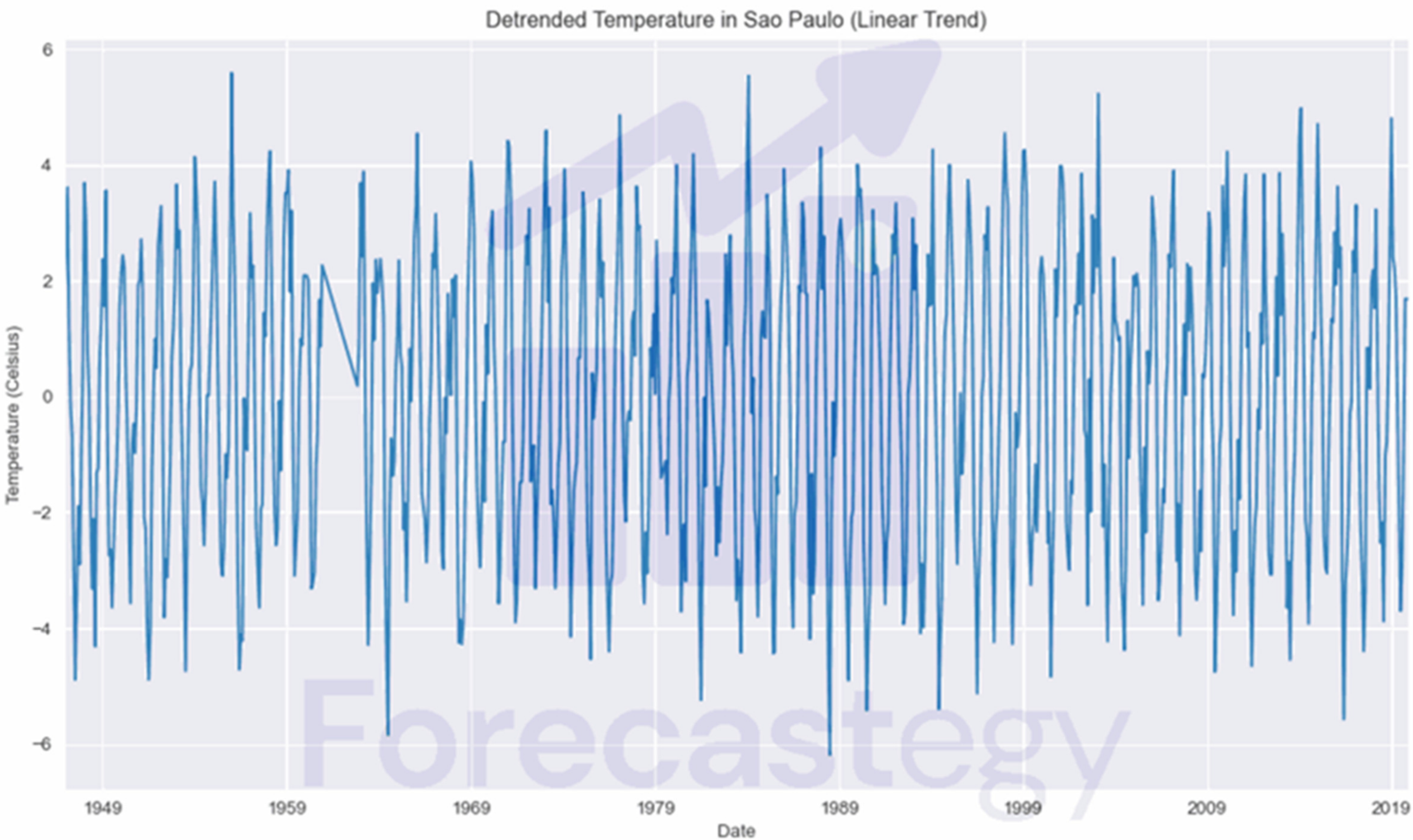


Detrending With A Linear Model (Scipy)

- It’s rare to find a time series where the trend is a simple horizontal line through time, but it’s very common to find a linear trend.

```
detrended = detrend(data['TEMP'], type='linear')
detrended = pd.Series(detrended, index=data.index)
```



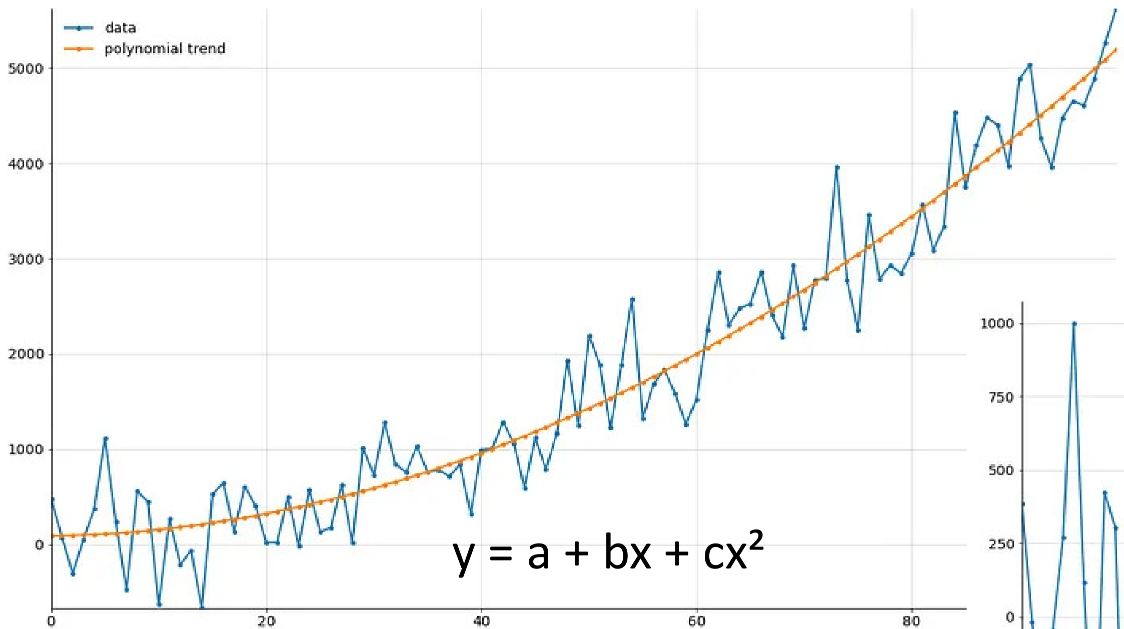


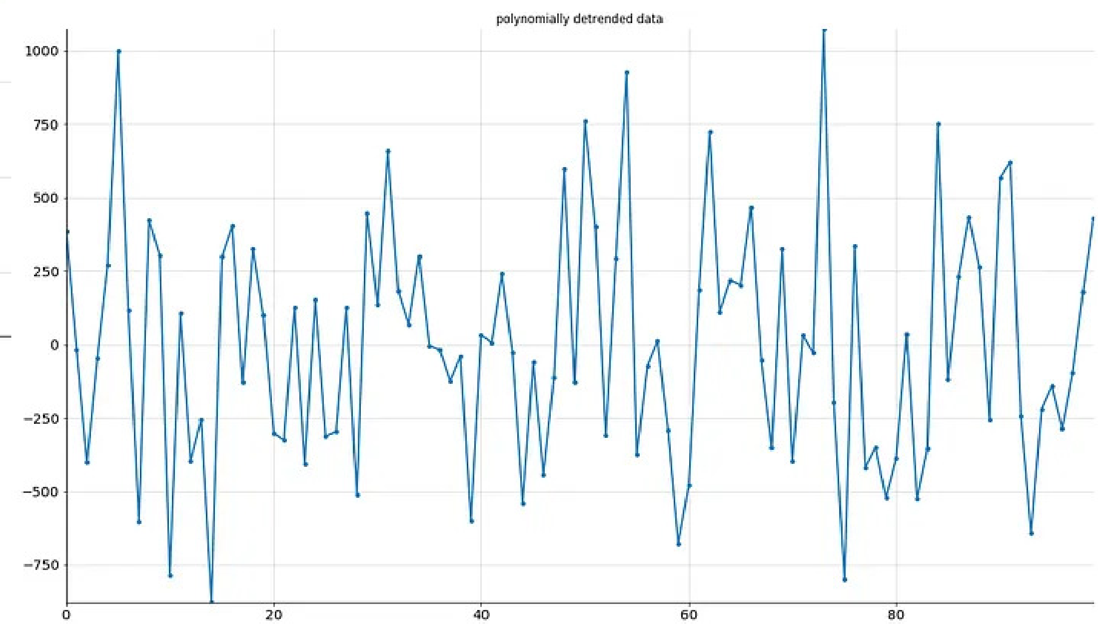
Detrended Temperature in Sao Paulo (Linear Trend)
Temperature (Celsius)
6
4
2
0
-2
-4
-6
1949
1959
1969
1979
1989
1999
2009
2019
Date
Forecastegy

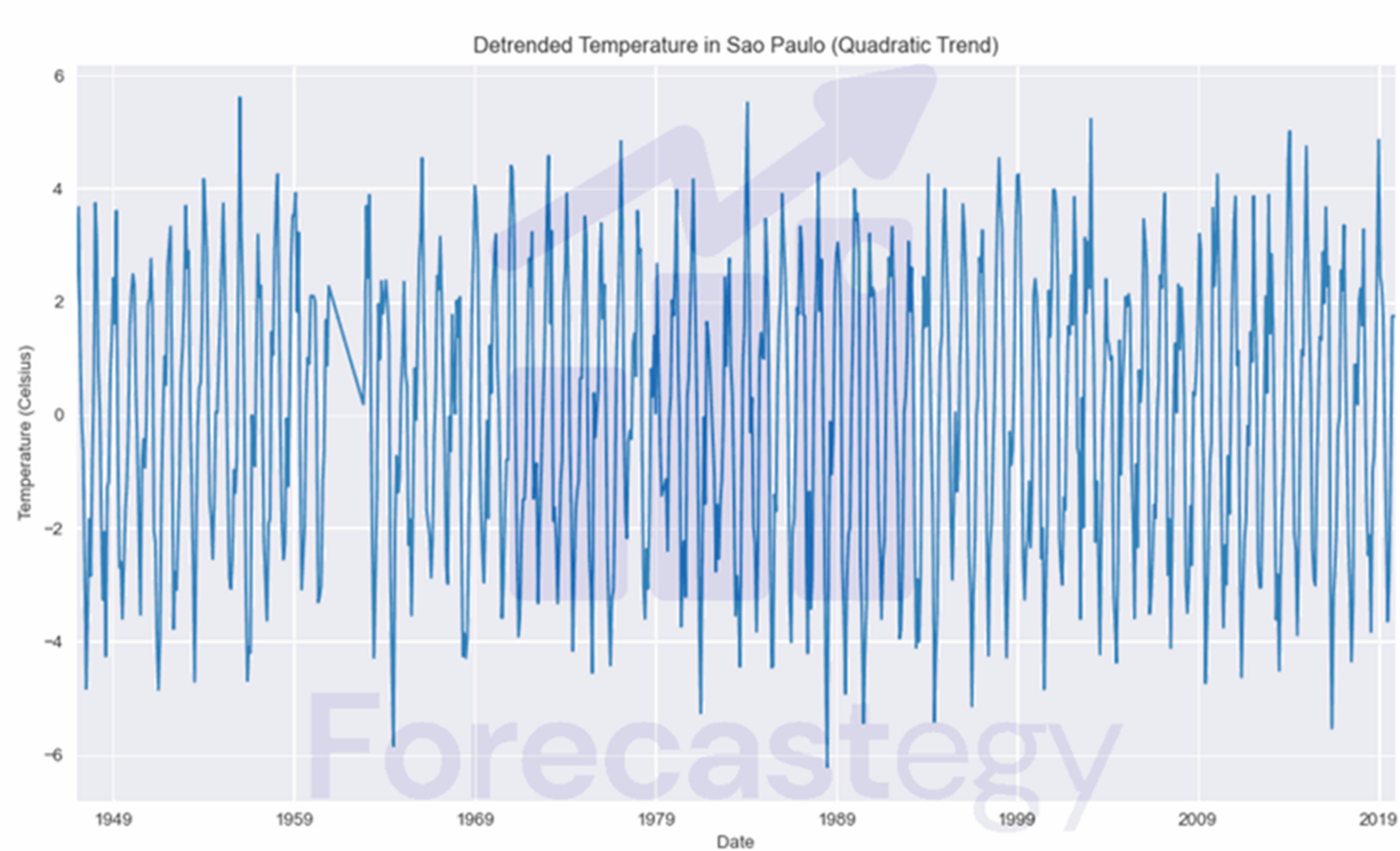
Detrending With A Quadratic Model (Statsmodels)

- If you find that the linear model is not enough to remove the trend, you can try a quadratic model.

```
from statsmodels.tsa.tsatools import detrend
detrended = detrend(data['TEMP'], order=2)
```





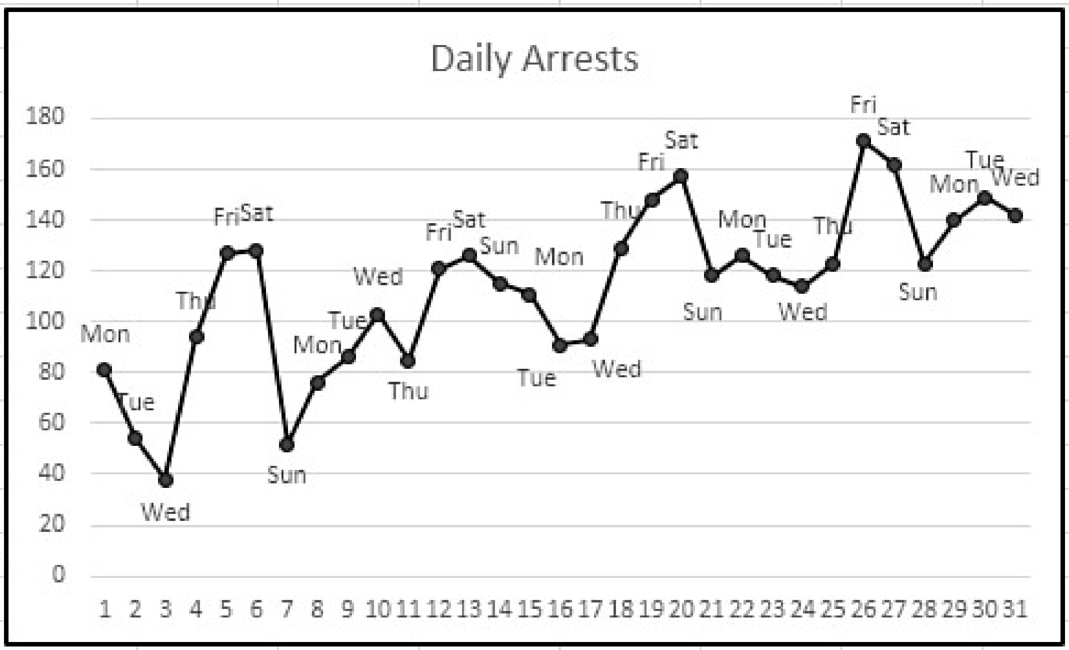

Detrended Temperature in Sao Paulo (Quadratic Trend)
Temperature (Celsius)
6
4
2
0
-2
-4
-6
1949
1959
1969
1979
1989
1999
2009
2019
Date
Forecastegy

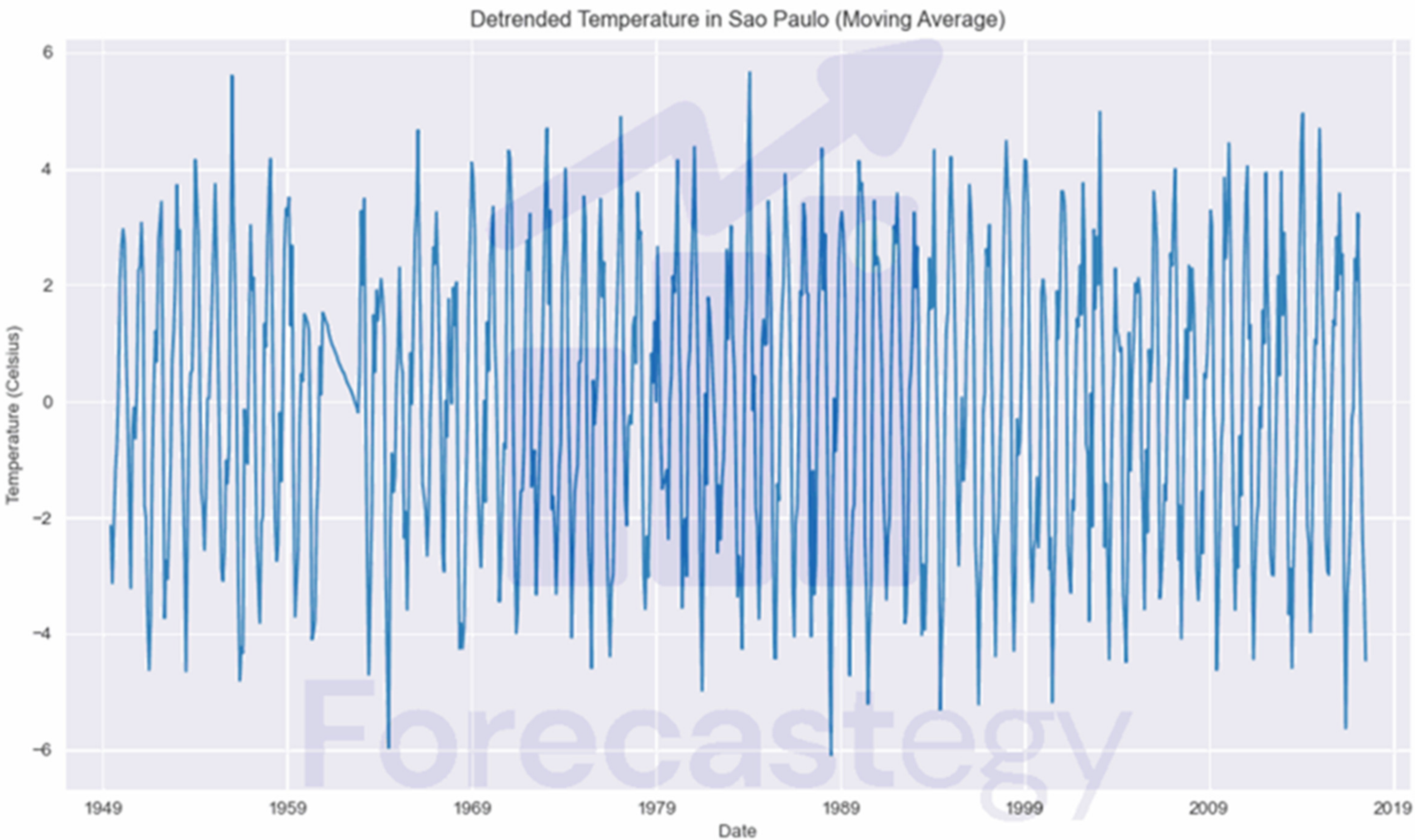
Detrending With A Moving Average Model

- This model calculates the average value of the data over a certain window and subtracts it from each data point, resulting in a smoothed version of the time series.

```
rolling_mean = data['TEMP'].rolling(window=60, center=True).mean()
detrended = data['TEMP'] - rolling_mean
```

Day	t	Arrests	Moving Average
Mon	1	81	
Tue	2	54	
Wed	3	38	
Thu	4	94	82.0
Fri	5	127	81.3
Sat	6	128	85.9
Sun	7	52	95.1
Mon	8	76	93.9
Tue	9	86	93.0
Wed	10	103	92.7
Thu	11	85	101.7
Fri	12	121	106.7
Sat	13	126	107.4
Sun	14	115	106.0
Mon	15	111	112.3





Detrended Temperature in Sao Paulo (Moving Average)
Temperature (Celsius)
6
4
2
0
-2
-4
-6
1949
1959
1969
1979
1989
1999
2009
2019
Date
Forecastegy

How To Select The Best Detrending Model?

- Always start by visually inspecting the data after applying each model.
- This will give you a good idea of how well each model is detrending the data.
- Look for any remaining trends, oscillations, or patterns that might suggest that the model is not fully capturing the underlying trend in the data.

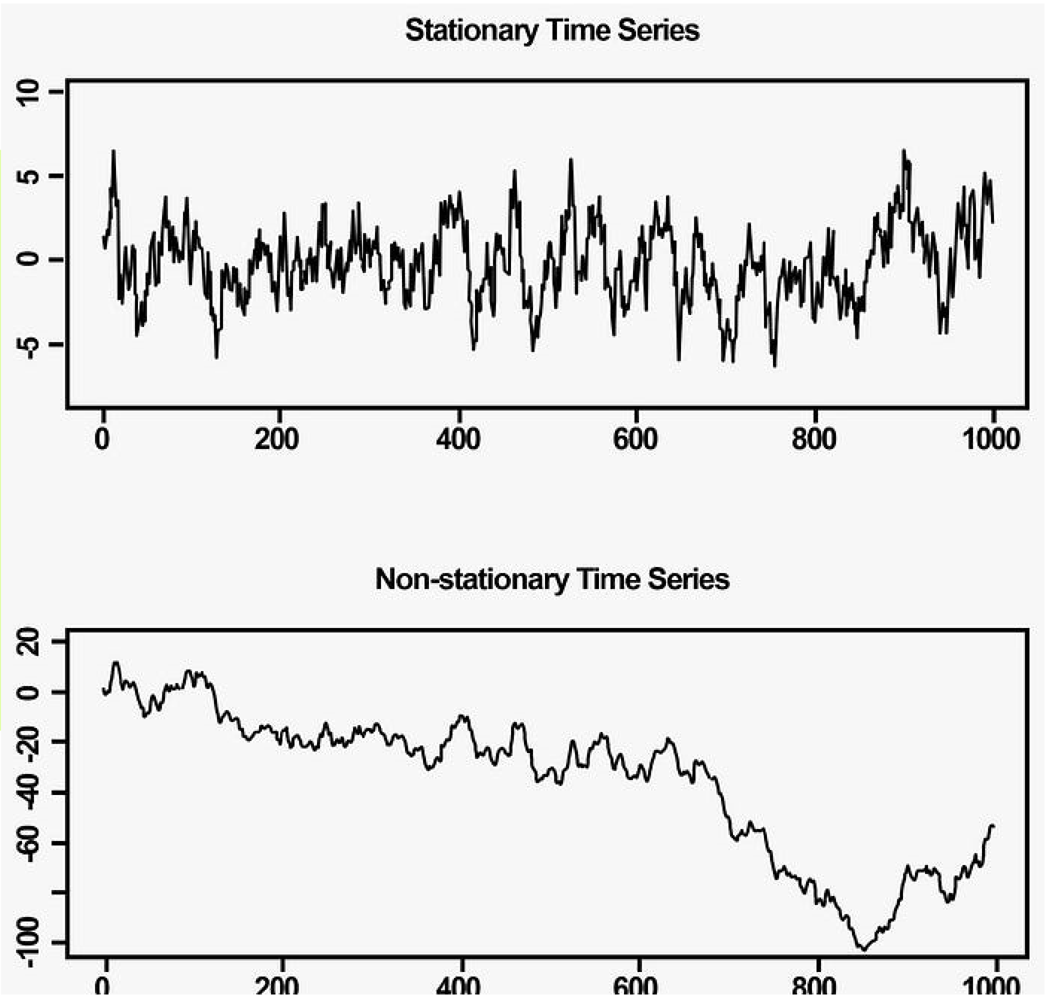
Differencing

This is a simple transformation of the series into a new time series, which we use to remove the series dependence on time and stabilize the mean of the time series, so **trend and seasonality are reduced during this transformation**.

$$\Delta X_t = X_t - X_{t-1}$$

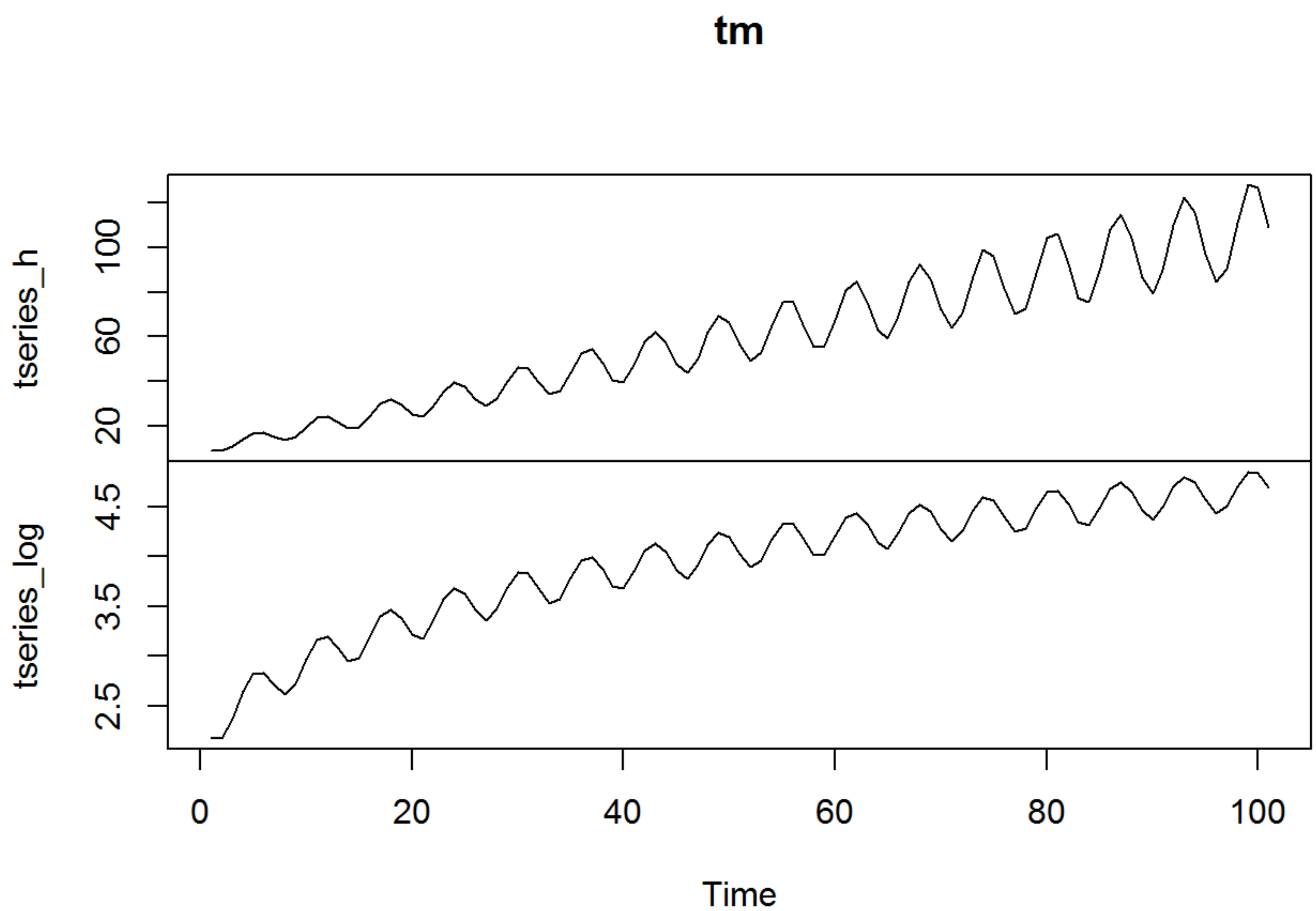
where:

- ΔX_t is the differenced series,
- X_t is the value of the series at time t,
- X_{t-1} is the value of the series at time $t-1$.



Transformation

- This includes three different methods they are Power Transform, Square Root, and Log Transfer. The most commonly used one is Log Transfer.



Exploration of ARIMA Model

An ARIMA model can be understood by outlining each of its components as follows:

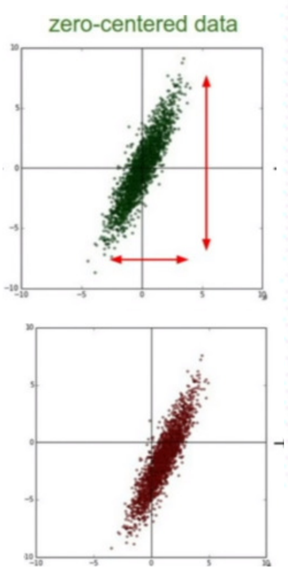
- **Autoregression (AR)**: refers to a model that shows a changing variable that regresses on its own lagged, or prior values.

- **Integrated (I)**: represents the differencing of raw observations to allow the time series **to become stationary** (i.e., data values are replaced by the difference between the data values and the previous values).

- **Moving average (MA)**: incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Autoregressive (AR) Models

- A statistical model is autoregressive if **it predicts future values based on past values.**
- For example, an autoregressive model might seek to predict a stock's future prices based on its past performance.
- For a stationary time series, y, t= 1, 2, 3, ... , an autoregressive model of order p, denoted AR(p) then



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

where δ is a constant for a nonzero-centered time series:

ϕ_j is a constant for j = 1, 2, . . ., p   are the parameters of the model

y_{t-j} is the value of the time series at time $t-j$

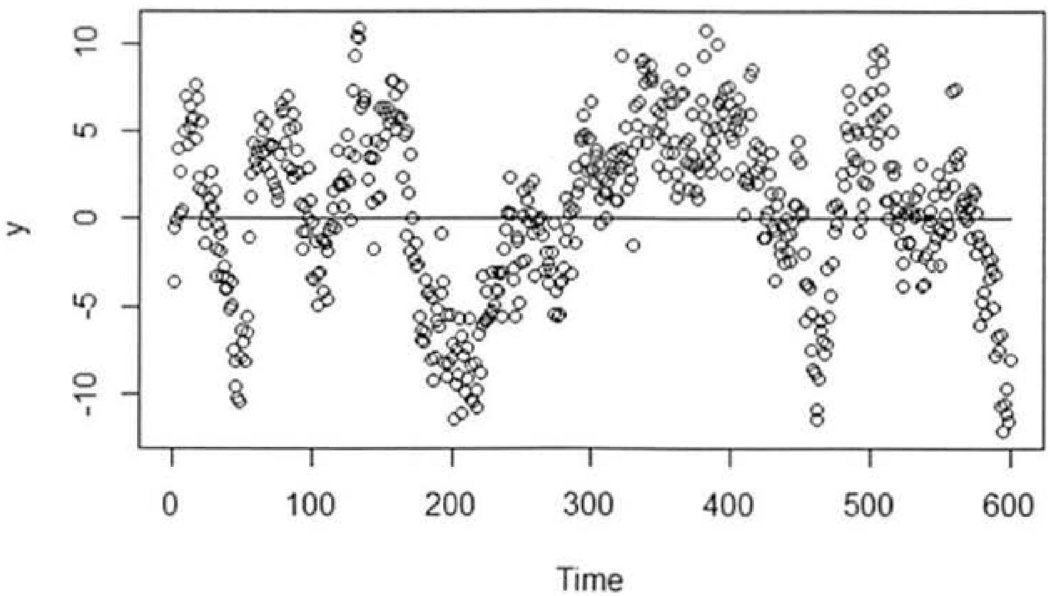
$\phi_p \neq 0$

$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ for all t white noise or error – independently and identically distributed with a mean zero and same variance.

Epsilon-Random error

Autocorrelation Function (ACF)

- Statistical correlation summarizes the strength of the relationship between two variables. Pearson's correlation coefficient [-1 and 1]
- We can calculate the correlation for time series observations with observations with previous time steps, called lags.
- Because the correlation of the time series observations is calculated with values of the same series at previous times, this is called a serial correlation, or an autocorrelation.



There is no overall trend in the time series plotted in Figure.

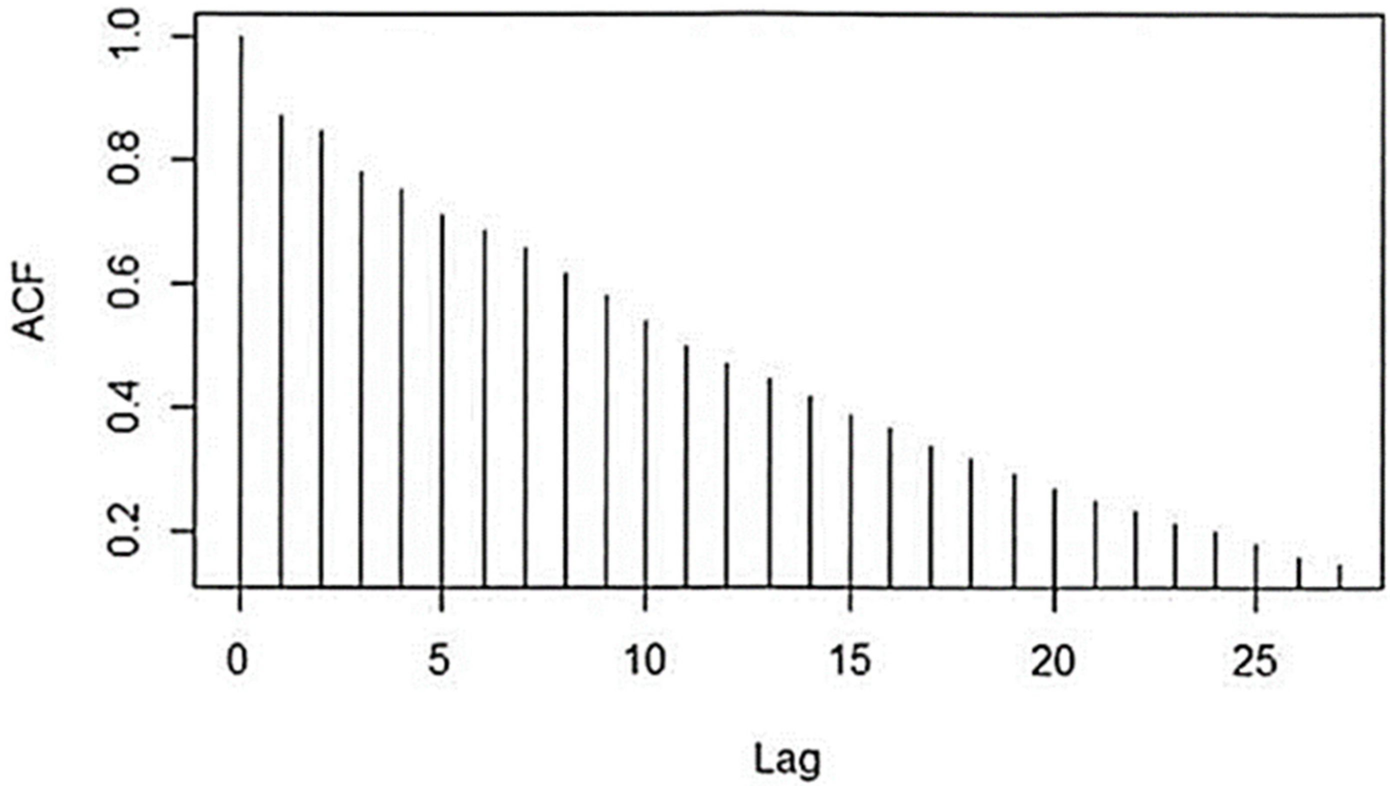
It appears that each point is somewhat dependent on the past points.

The difficulty is that the plot does not provide insight into the variables' covariance in the time series and its underlying structure.

The autocorrelation function (ACF) provides this insight.

$$ACF(h) = \frac{cov(y_t, y_{t+h})}{\sqrt{cov(y_t, y_t)\, cov(y_{t+h}, y_{t+h})}} = \frac{cov(h)}{cov(0)}$$

A plot of the autocorrelation of a time series by lag is called the **AutoCorrelation Function**, or the ACF. This plot is sometimes called a correlogram or an autocorrelation plot.



By convention, the quantity h in the ACF is referred to as the *lag*, the difference between the time points t and $t+h$. At lag 0, the ACF provides the correlation of every point with itself. So ACF(0) always equals 1. According to the ACF plot, at lag 1 the correlation between y_t and y_{t-1} is approximately 0.9, which is very close to 1. So y_{t-1} appears to be a good predictor of the value of y_t. Because ACF(2) is around 0.8, y_{t-2} also appears to be a good predictor of the value of y_t. A similar argument could be made for lag 3 to lag 8. (All the autocorrelations are greater than 0.6.) In other words, a model can be considered that would express y_t as a linear sum of its previous 8 terms. Such a model is known as an autoregressive model of order 8.

Partial Autocorrelation Function(PACF)

- Partial correlation is a statistical method used to measure
  - how strongly two variables are related while considering and adjusting for the influence of one or more additional variables.
- The correlation between two variables indicates how much they change together.
- Nonetheless, partial correlation takes an additional step by considering the potential influence of other variables affecting this relationship.

'Partial correlation goes a step further by removing the influence of other factors.

Exposes Direct Connections:
• PACF helps you see the direct relationship between points in a time series, removing the effects of the points in between. It's like understanding how today's weather affects tomorrow's weather, without being influenced by yesterday's weather.
• Eliminates Shared Variability:
• By considering the effect of other variables, PACF gives a clearer picture of the unique relationship between two points in a time series..

- This way, partial correlation exposes the distinctive connection between two variables by eliminating the shared variability with the control variables.

From the earlier example[slide 25], the autocorrelations are quite high for the first several lags. Although it appears that an AR(8) model might be a good candidate to consider for the given dataset, examining an AR(1) model provides further insight into the ACF and the appropriate value of p to choose. For an AR(1) model, centered around, $\delta = 0$, the simplified form is:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \qquad y_{t-1} = \phi_1 y_{t-2} + \varepsilon_{t-1}.$$

$$\begin{aligned} y_t &= \phi_1(\phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \end{aligned}$$

Therefore, in a time series that follows an AR(1) model, considerable autocorrelation is expected at lag 2. As this substitution process is repeated, Y_t, can be expressed as a function of Y_{t-h}, for h = 3, 4 ... and a sum of the error terms.

This observation means that even in the simple AR(1) model, there will be considerable autocorrelation with the larger lags even though those lags are not explicitly included in the model.

What is needed is a measure of the autocorrelation between Y_t, and Y_{t+h}, for h = 1, 2, 3 ... with **the effect of the Y_{t+1} to Y_{t+h-1} values excluded from the measure**. The **partial autocorrelation function (PACF) provides** such a measure.

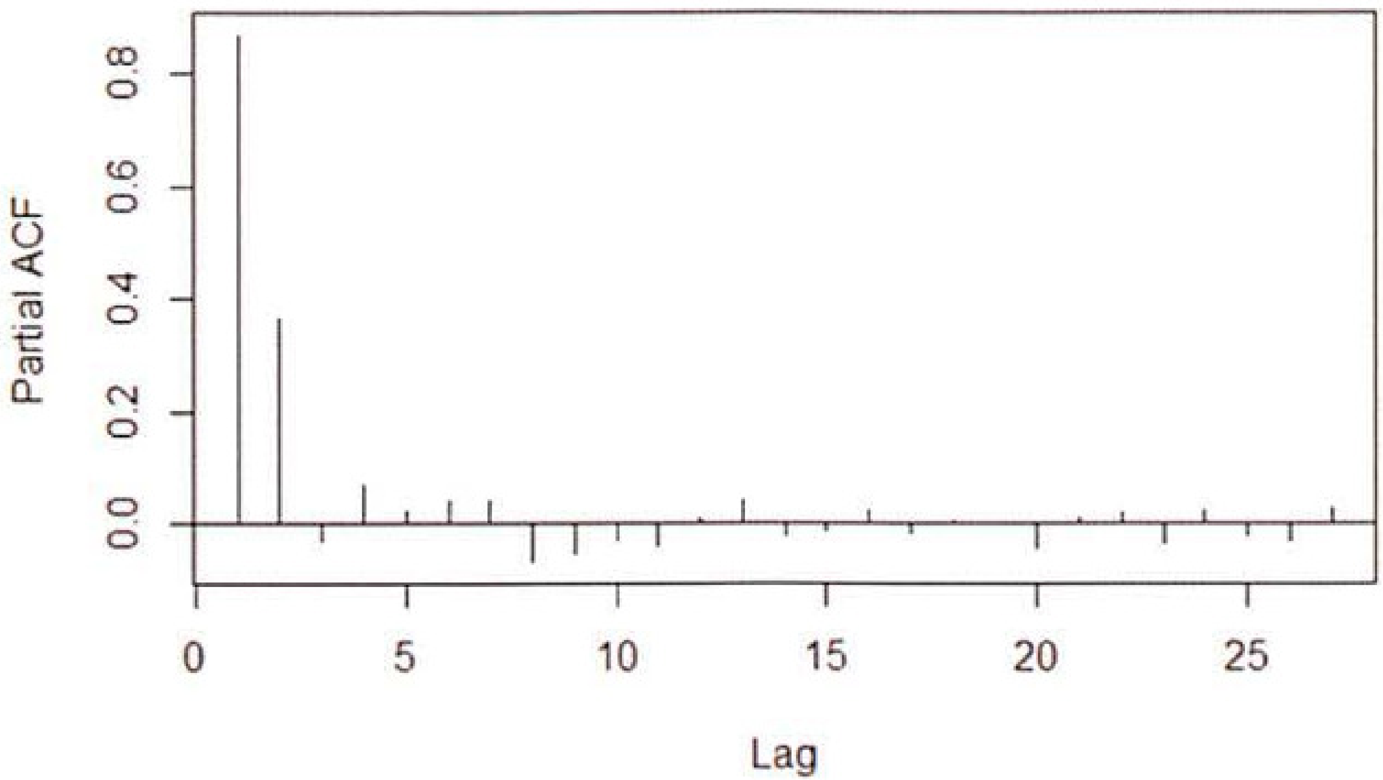
Given $y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$

$$PACF(h) = corr(y_t - y_t^*, y_{t+h} - y_{t+h}^*) \text{ for } h \geq 2$$
$$= corr(y_t, y_{t+1}) \qquad \text{for } h = 1$$

where $y_t^* = \beta_1 y_{t+1} + \beta_2 y_{t+2} \cdots + \beta_{h-1} y_{t+h-1}$,
$y_{t+h}^* = \beta_1 y_{t+h-1} + \beta_2 y_{t+h-2} \cdots + \beta_{h-1} y_{t+1}$, and
the h − 1 values of the βs are based on linear regression.

In other words, **after linear regression is used to remove the effect of the variables between Y_t, and Y_{t+h} on Y_t, and Y_{t+h}, the PACF is the correlation of what remains**.

For h = 1, there are no variables between Y_t, and Y_{t+1}. So the PACF(1) equals ACF(1). Although the computation of the PACF is somewhat complex, many software tools hide this complexity from the analyst.



For the earlier example, the PACF plot in Figure illustrates that after lag 2, the value of the PACF is sharply reduced. Thus, after removing the effects of y_{t+1} and y_{t+2}, the partial correlation between y_t and y_{t+3} is relatively small. Similar observations can be made for h = 4, 5, ... Such a plot indicates that an AR(2) is a good candidate model for the time series plotted in Figure 8-2. In fact, the time series data for this example was randomly generated based on $y_t = 0.6y_{t-1} + 0.35y_{t-2} + \varepsilon_t$ where $\varepsilon_t \sim N(0, 4)$.

Moving Average(MA) Models

- Rather than using **past values of the forecast variable in a regression**, a moving average model uses past forecast errors in a regression-like model.
- For a time series, y_t, centered at zero, a moving average model of order q, denoted MA(q), is expressed as:

  $$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

  Where θ_k is a moving average constant for k=1,2,3, … , q

  $\theta_q \neq 0$

  $\epsilon_t \sim N(0, \sigma^2)$ for all t are the white noise terms[mean zero and same variance]

- For a simulated MA(3) time series of the form

  $y_t = \epsilon_t - 0.4\, \epsilon_{t-1} + 1.1 \epsilon_{t-2} - 2.5\, \epsilon_{t-3}$ where ϵ_t ~N(0, 1)

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3}$$

$$y_{t-1} = \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3} + \theta_3 \varepsilon_{t-4}$$

$$y_{t-2} = \varepsilon_{t-2} + \theta_1 \varepsilon_{t-3} + \theta_2 \varepsilon_{t-4} + \theta_3 \varepsilon_t$$

$$y_{t-3} = \varepsilon_{t-3} + \theta_1 \varepsilon_{t-4} + \theta_2 \varepsilon_{t-5} + \theta_3 \varepsilon$$

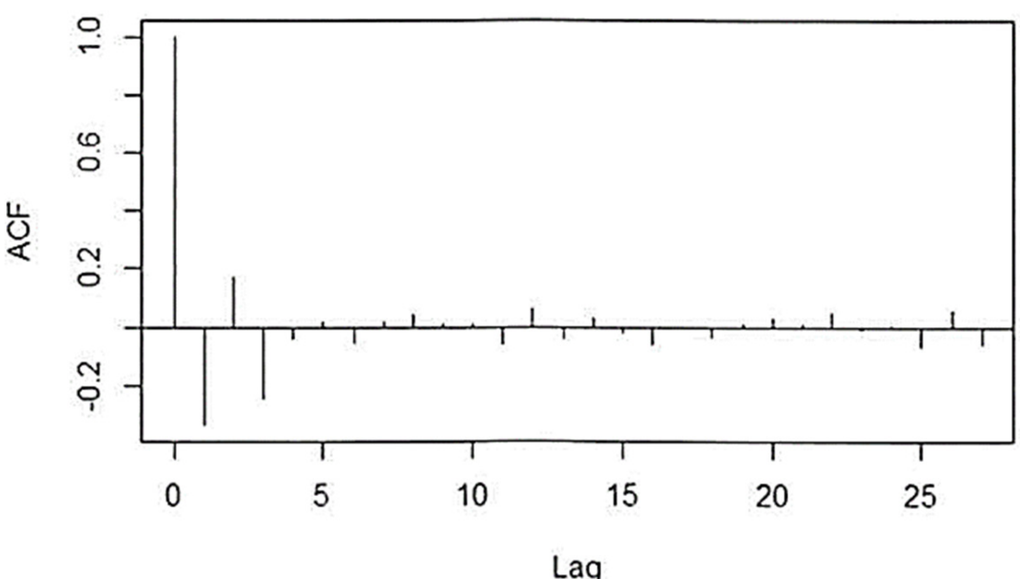
$$y_{t-4} = \varepsilon_{t-4} + \theta_1 \varepsilon_{t-5} + \theta_2 \varepsilon_{t-6} + \theta_3$$

A	B	C	D
Day	t	Arrests	Moving Average
Mon	1	81	
Tue	2	54	
Wed	3	38	
Thu	4	94	82.0
Fri	5	127	81.3
Sat	6	128	85.9
Sun	7	52	95.1
Mon	8	76	93.9
Tue	9	86	93.0
Wed	10	103	92.7
Thu	11	85	101.7
Fri	12	121	106.7
Sat	13	126	107.4
Sun	14	115	106.0
Mon	15	111	112.3

Moving Average(MA) Models

Because the expression of Y_t shares specific white noise variables with the expressions for Y_{t-1} through Y_{t-3}, inclusive, those three variables are correlated to Y_t .

However, the expression of Y_t does not share white noise variables with Y_{t-4}. So the theoretical correlation between Y_t and Y_{t-4} is zero.



Of course, when dealing with a particular dataset, the theoretical autocorrelations are unknown, but the observed autocorrelations should be close to zero for lags greater than q when working with an MA(q) model.

The ACF(0) equals 1, because any variable is perfectly correlated with itself.
At lags 1, 2, and 3, the value of the ACF is relatively large in absolute value compared to the subsequent terms. In an autoregressive model, the **ACF slowly decays**, but for an MA(3) model, the ACF somewhat abruptly cuts off after lag 3. In general, this pattern can be extended to any MA(q) model.

ARMA and ARIMA Models

- In general, the data scientist does not have to choose between an AR(p) and an MA(q) model to describe a time series. It is often useful to combine these two representations into one model.
- The combination of these two models for a stationary time series results in an Autoregressive Moving Average model,

$$
\begin{aligned}
y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} \\
+ \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}
\end{aligned}
\quad (8\text{-}15)
$$

where δ is a constant for a nonzero-centered time series

ϕ_j is a constant for j = 1, 2, ..., p

$\phi_p \neq 0$

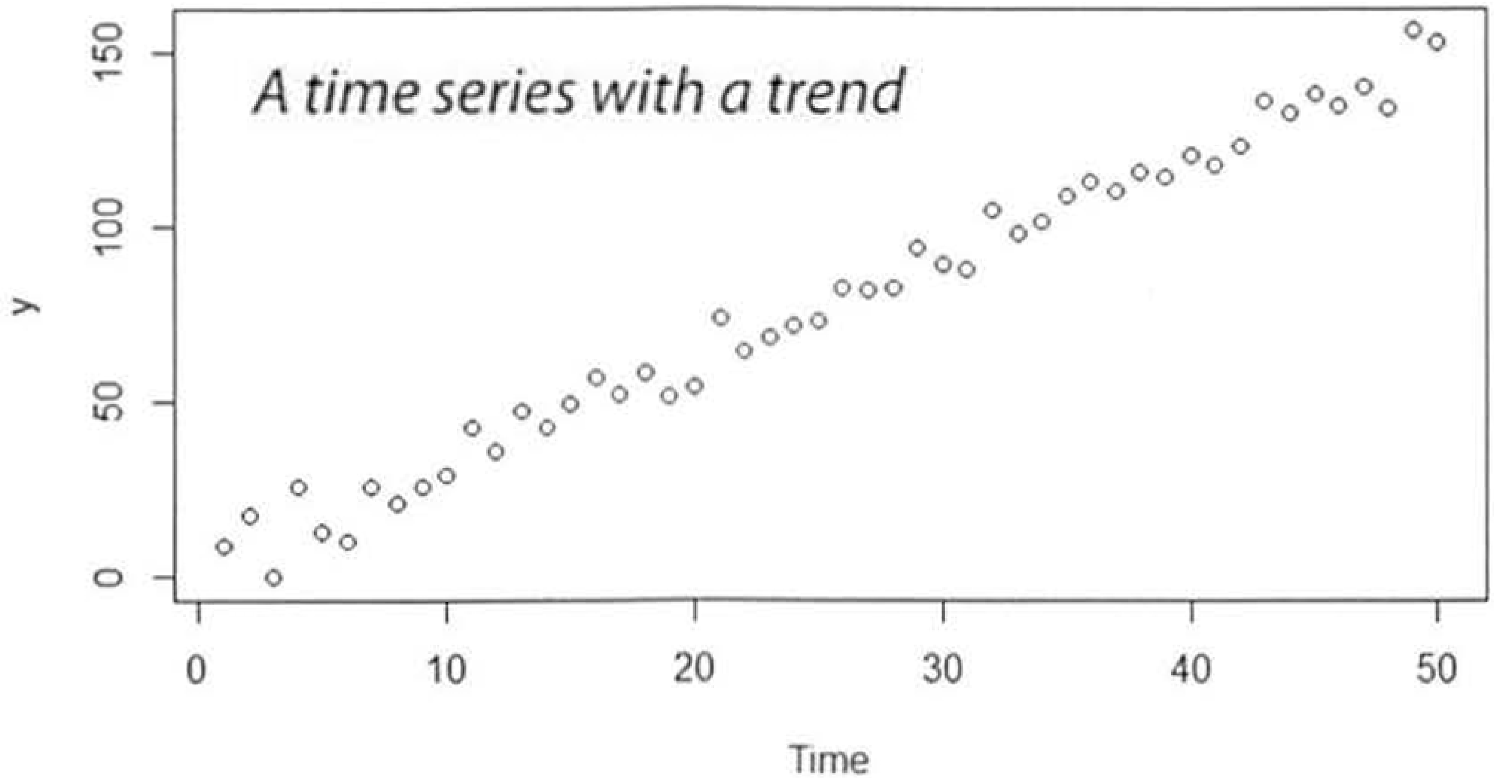
θ_k is a constant for k = 1, 2, ..., q

$\theta_q \neq 0$

$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ for all t

If p != 0 and q =0, then the ARMA(p,q) model is simply an AR(p) model. Similarly, if p = 0 and q !=0, then the ARMA(p,q) model is an MA(q) model.

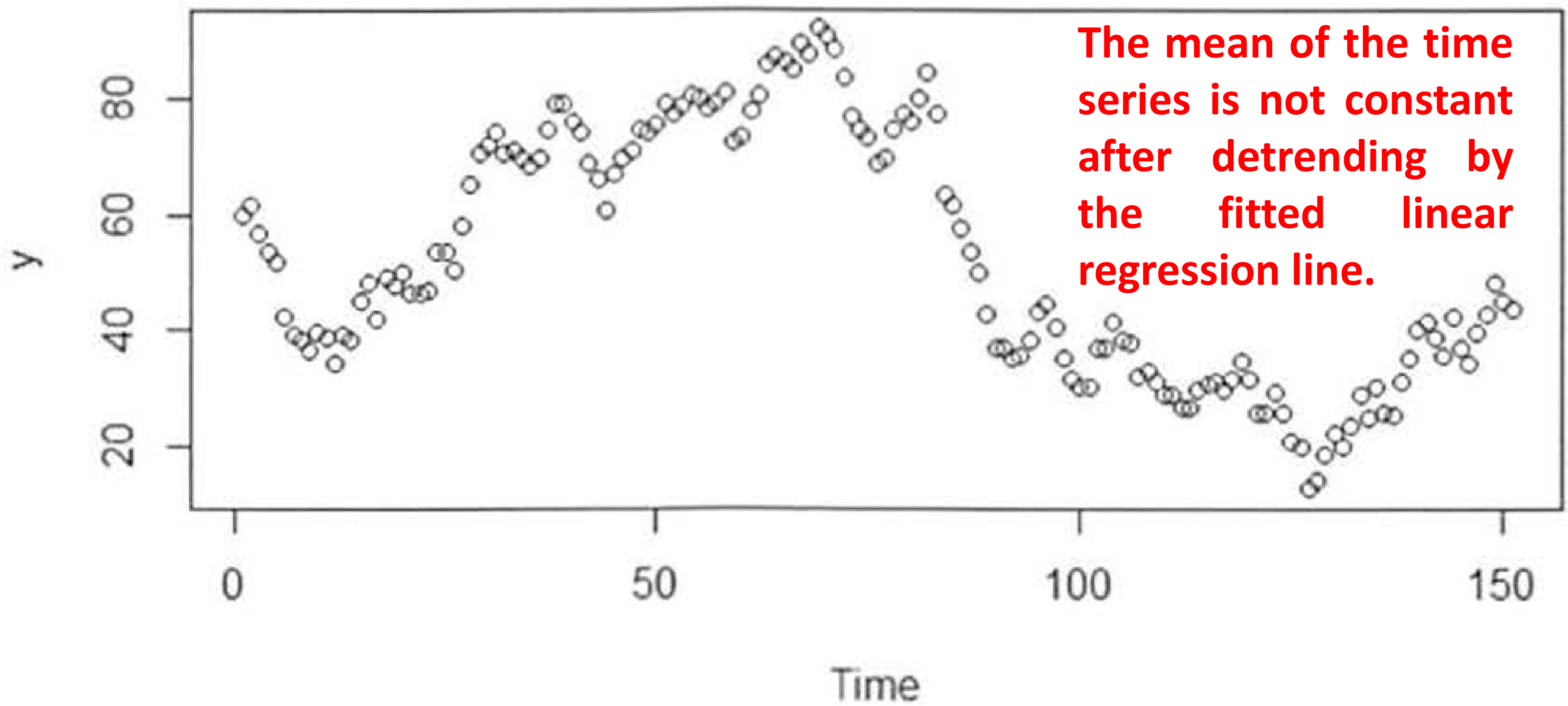
To apply an ARMA model properly, the time series must be stationary. However, any time series exhibit some trends over time.



To apply an ARMA model properly, the time series must be a stationary one. However, many time series exhibit some trend over time. Figure illustrates a time series with an increasing linear trend over time. Since such a time series does not meet the requirement of a constant expected value (mean), the data needs to be adjusted to remove the trend. One transformation option is to perform a regression analysis on the time series and then to subtract the value of the fitted regression line from each observed y-value.

If detrending using a linear or higher order regression model does not provide a stationary series, a second option is to compute the difference between successive y-values. This is known as ***differencing***. In other words, for the n values in a given time series compute the differences as shown in Equation 8-16.

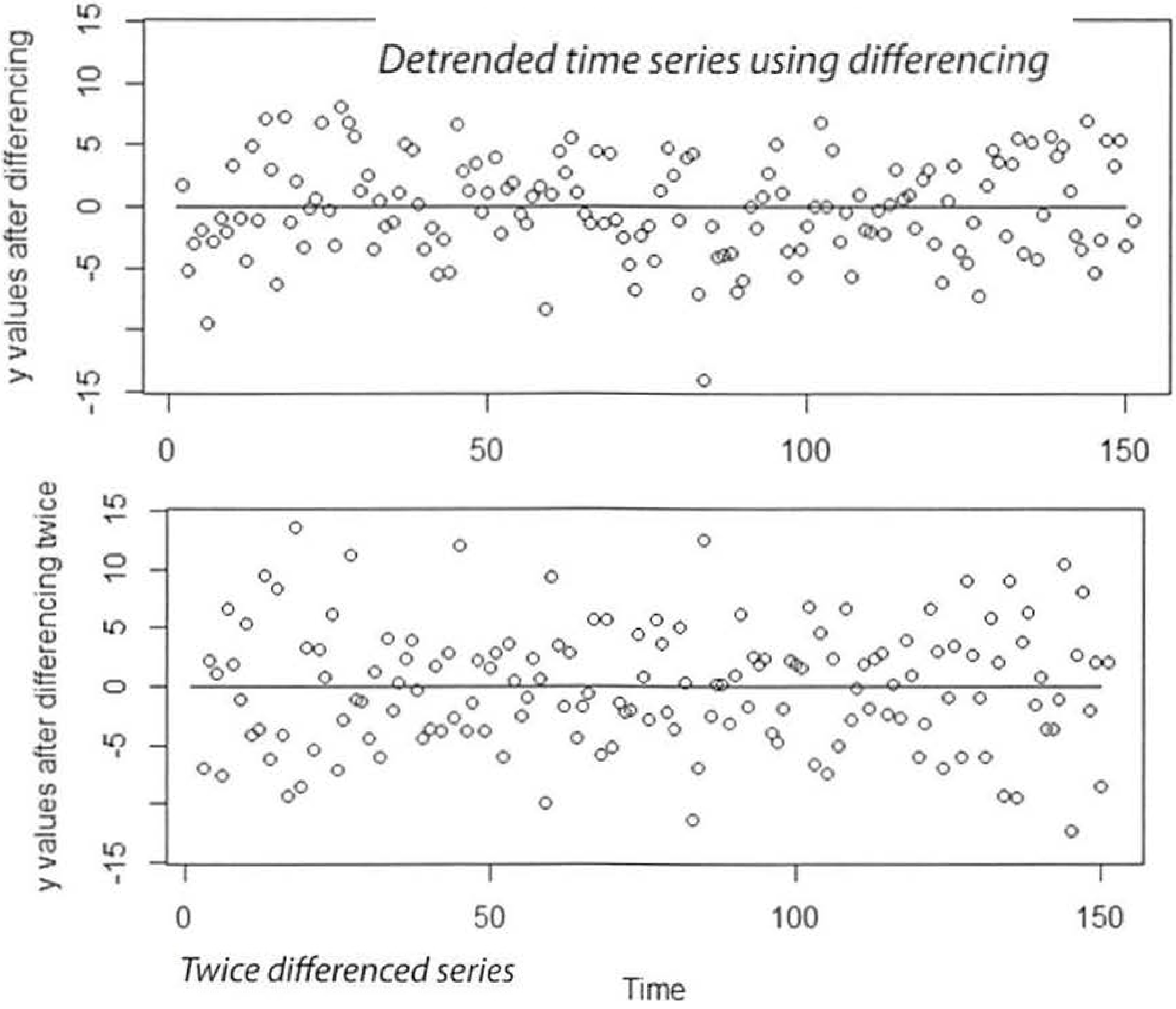
$$d_t = y_t - y_{t-1} \quad \text{for } t = 2,3,\ldots,n \tag{8-16}$$



If the differenced series is not reasonably stationary, applying differencing additional times may help. Equation 8-17 provides the twice differenced time series for $t = 3, 4, \ldots n$.

$$\begin{aligned} d_{t-1} - d_{t-2} &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2} \end{aligned} \tag{8-17}$$

Successive differencing can be applied, but over-differencing should be avoided. One reason is that over-differencing may unnecessarily increase the variance. The increased variance can be detected by plotting the possibly over-differenced values and observing that the spread of the values is much larger, as seen in Figure 8-10 after differencing the values of y twice.


Detrended time series using differencing
y values after differencing
15
10
5
0
-5
-15
0
50
100
150
y values after differencing twice
15
10
5
0
-5
-15
0
50
100
150
Twice differenced series
Time

- Because the need to make a time series stationary is common, the differencing can be included (integrated) into the ARMA model definition by defining the **Autoregressive Integrated Moving Average model, denoted ARIMA(p,d,q)**.
- The structure of the ARIMA model is identical to the expression in Equation 8-15. Still, the ARMA(p,q,d) model is applied to the time series, Y_t, after applying differencing d times.

Additionally, it is often necessary to account for seasonal patterns in time series. For example, in the retail sales use case example in Section 8.1, monthly clothing sales track closely with the calendar month. Similar to the earlier option of detrending a series by first applying linear regression, the seasonal pattern could be determined and the time series appropriately adjusted. An alternative is to use a ***seasonal autoregressive integrated moving average model***, denoted ARIMA(p,d,q) $\times$ $(P,D,Q)_s$ where:

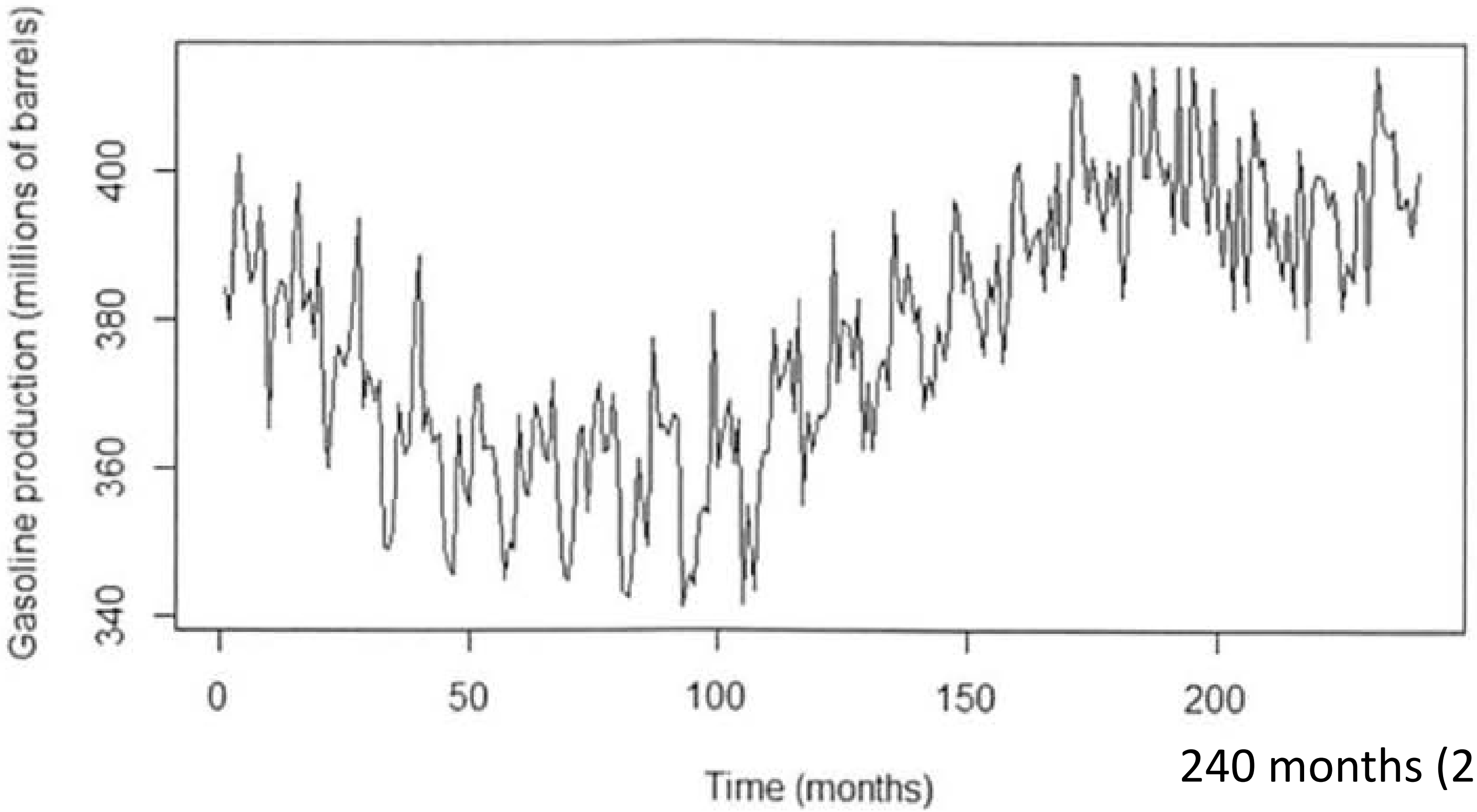
- p, d, and q are the same as defined previously.
- s denotes the seasonal period.
- P is the number of terms in the AR model across the s periods.
- D is the number of differences applied across the s periods.
- Q is the number of terms in the MA model across the s periods.

For a time series with a seasonal pattern, following are typical values of s:

- 52 for weekly data
- 12 for monthly data
- 7 for daily data

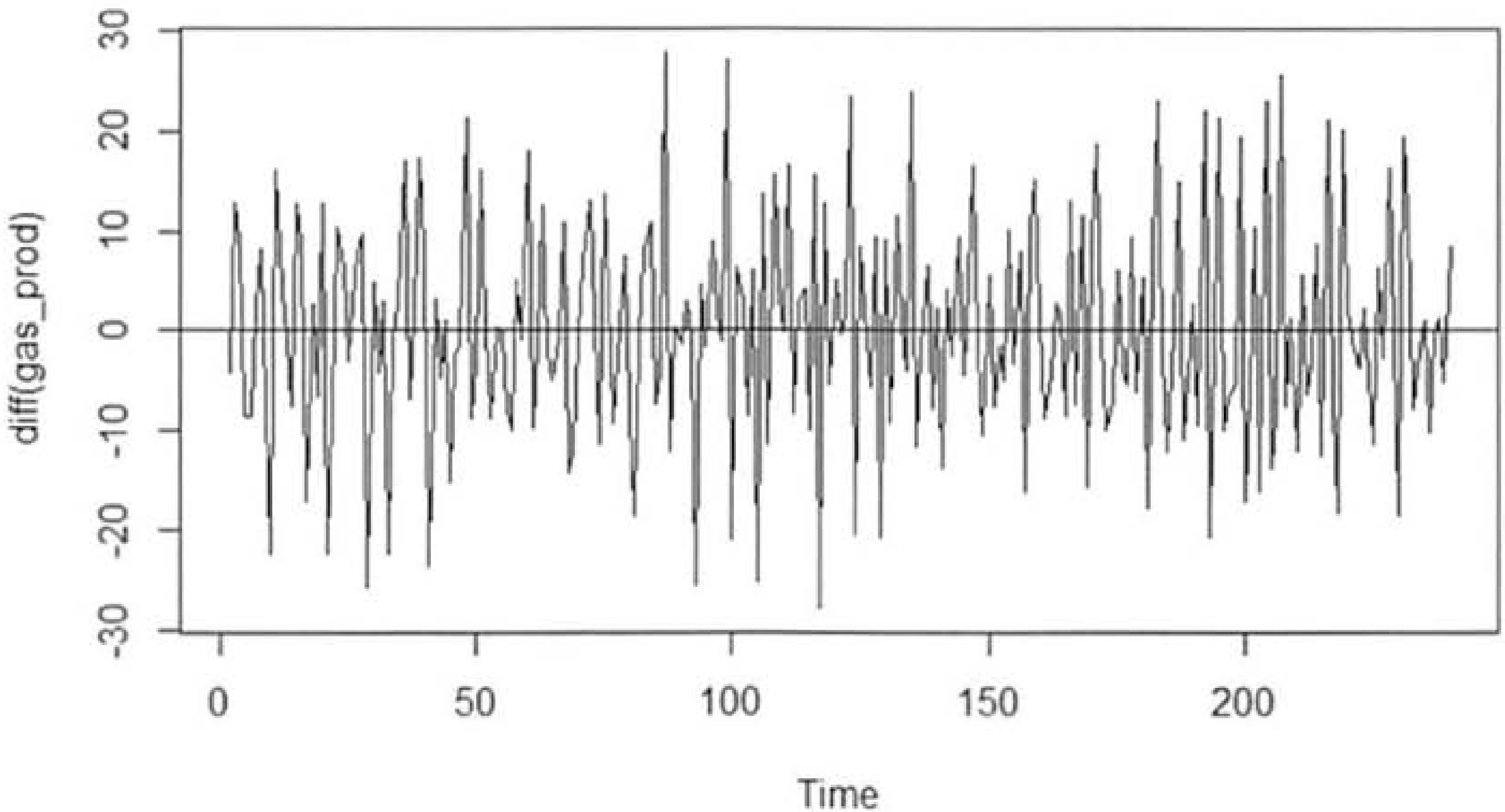
Evaluating an ARIMA Model

- Monthly gasoline production



240 months (20 years).

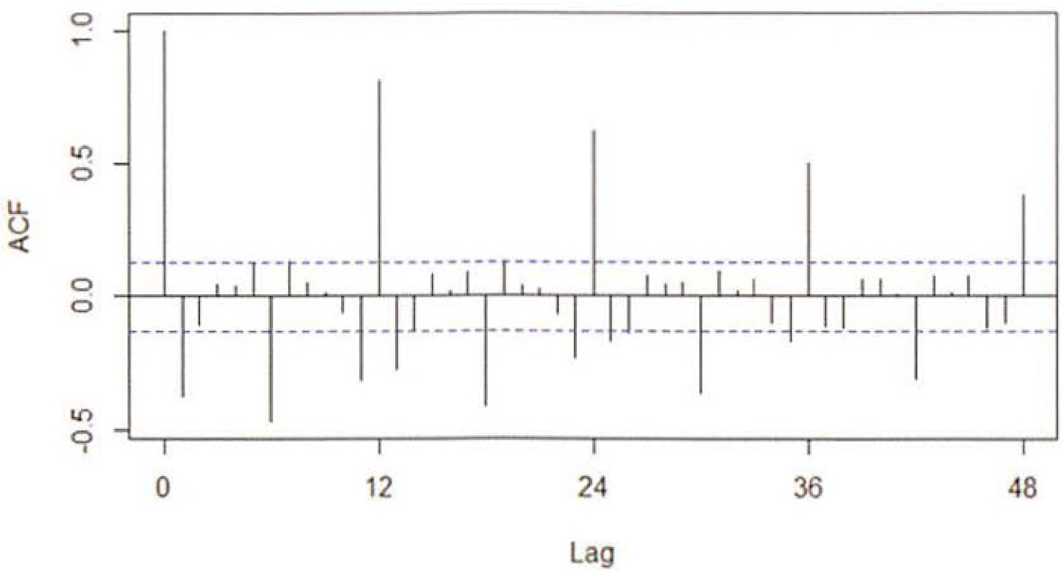
- Differenced gasoline production time series



The differenced time series has a constant mean near zero with a fairly constant variance over time. Thus, a stationary time series has been obtained. The ACF and PACF plots **for the differenced series** are provided in the next slide.

The dashed lines provide upper and lower bounds at a 95% significance level. Any value of the ACF or PACF outside of these bounds indicates that the value is significantly different from zero.

- ACF of the differenced gasoline time series

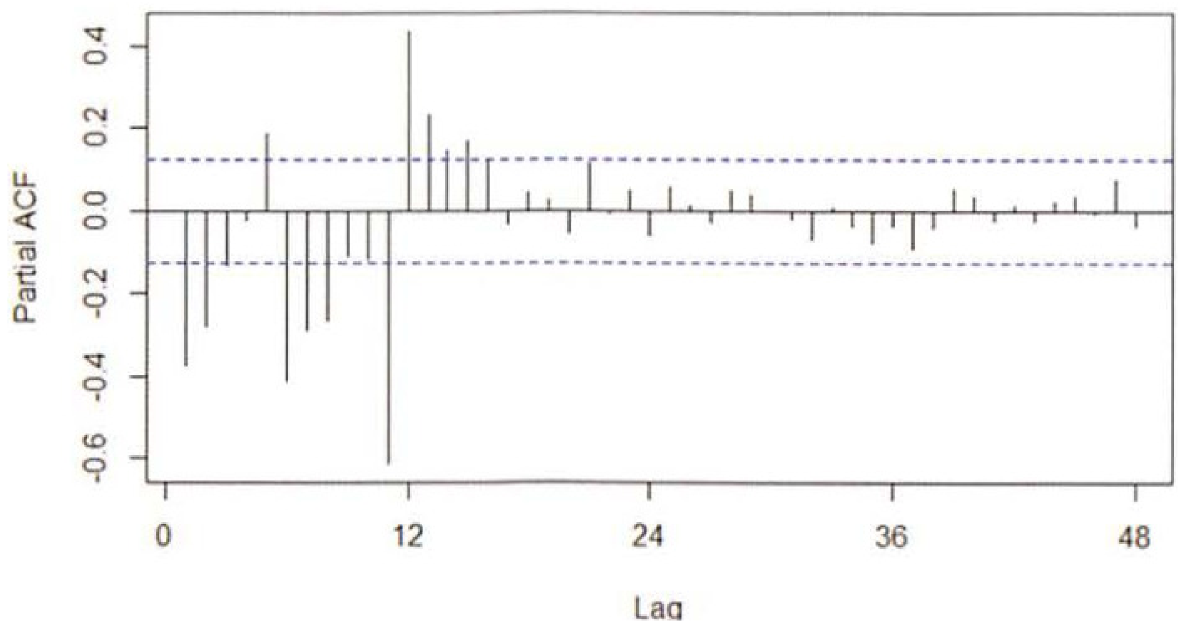


- The figure shows several significant ACF values. The slowly decaying ACF values at lags 12, 24, 36, and 48 are of particular interest.
- The figure indicates a seasonal autoregressive pattern every 12 months

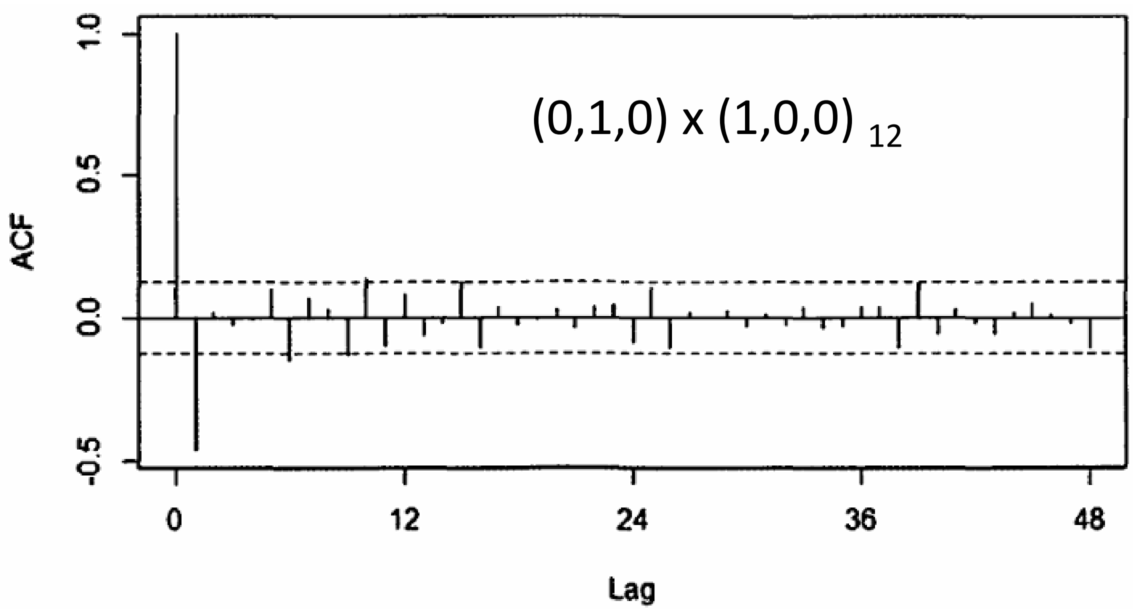
- PACF of the differenced gasoline time series

- Examining the PACF plot in the figure, the PACF value at lag 12 is quite large, but the PACF values are close to zero at lags 24, 36, and 48.

- Thus, a seasonal AR(1) model with period = 12 will be considered.

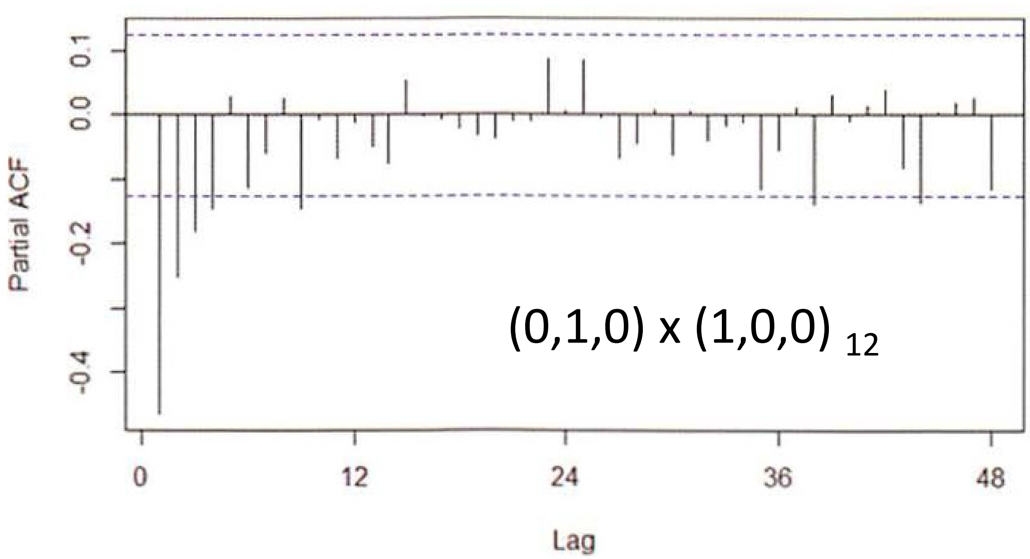


- ACF of residuals from seasonal AR(1) model



The ACF plot of the residuals in Figure indicates that the autoregressive behavior at lags 12, 24, 26, and 48 has been addressed by the seasonal AR(1) term. The only remaining ACF value of any significance occurs at lag 1.

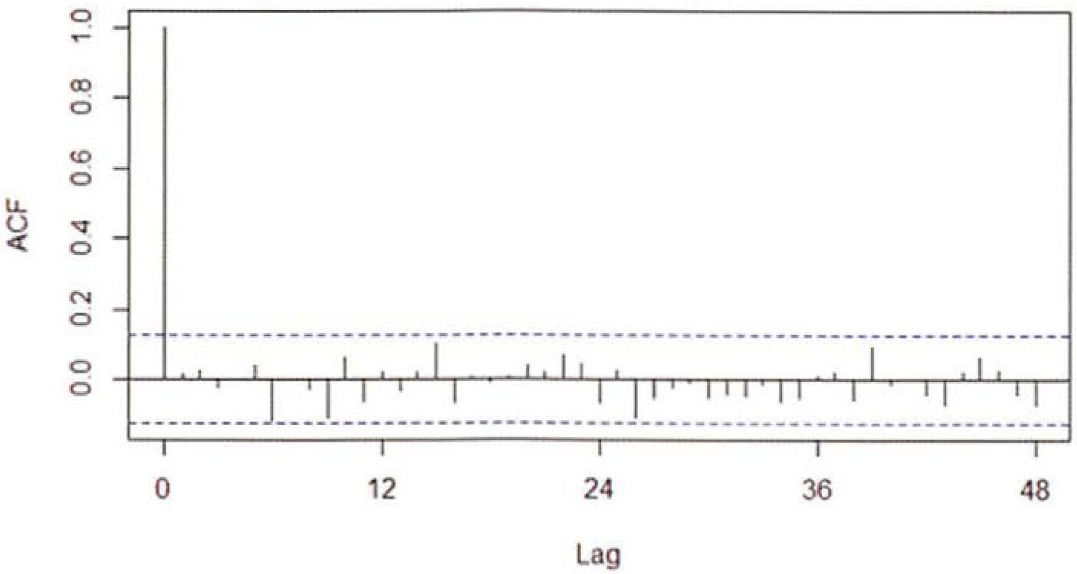
- PACF of residuals from seasonal AR(1) model



The figure shows several significant PACF values at lags 1, 2, 3, and 4. The PACF plot in the figure exhibits a slowly decaying PACF, and the ACF cuts off sharply at lag 1.

MA(1) model should be considered for the nonseasonal portion of the ARMA model on the differenced series. In other words, a (0,1,1) x $(1,0,0)_{12}$ ARIMA model will be fitted to the original gasoline production time series.

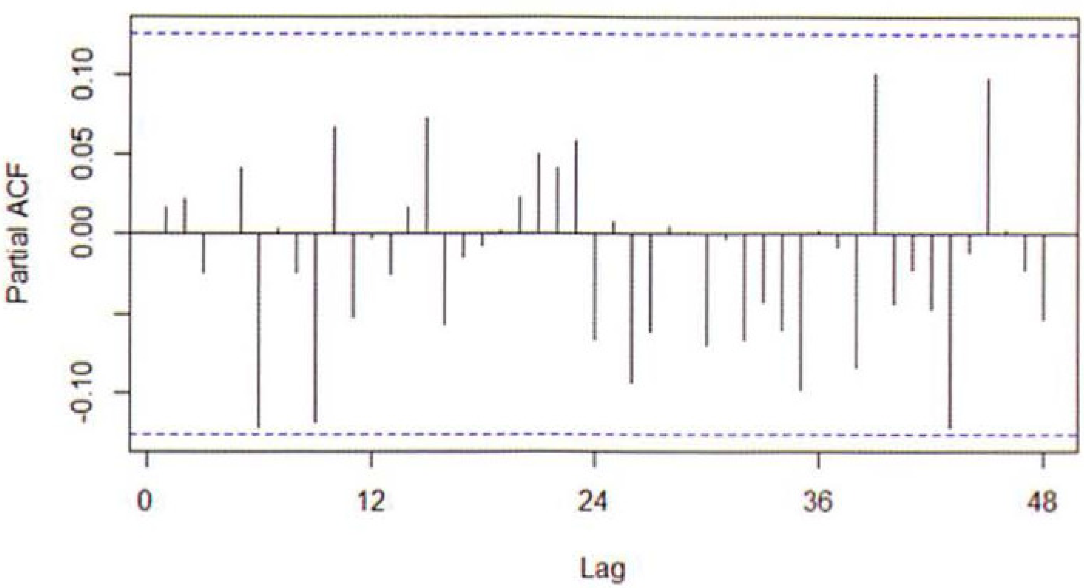
- ACF for the residuals from the $(0, 1, 1) \times (1,0,0)_{12}$ model



Based on the standard errors associated with each coefficient estimate, the coefficients are significantly different from zero.

In the figures, respective ACF and PACF plots for the residuals from the second pass ARIMA model indicate that no further terms need to be considered in the ARIMA model.

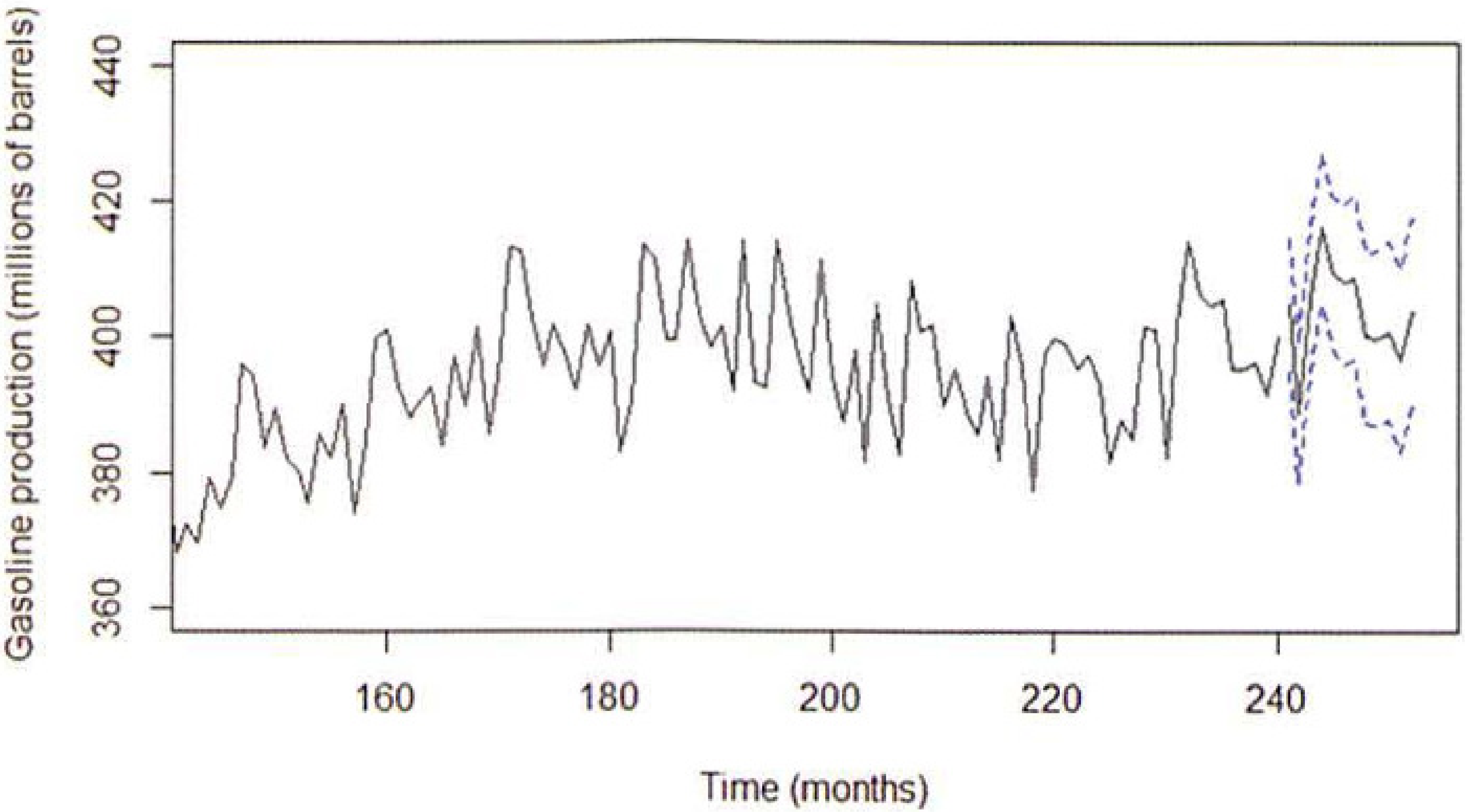
- PACF for the residuals from the $(0, 1, 1) \times (1,0,0)_{12}$ model



It should be noted that the ACF and PACF plots each have several points that are **close to the bounds at a 95% significance level.** However, these points occur **at relatively large lags**.

To avoid overfitting the model, these values are attributed to random chance. So no attempt is made to include these lags in the model. However, it is advisable to compare a reasonably fitting model to slight variations of that model.

Actual and forecasted gasoline production



Comparing Fitted Time Series Models

A model that finds the best balance will be predictive.

But how can we evaluate a model's complexity and its fitting numerically?

- AIC (Akaike Information Criterion)
- AICc (Akaike Information Criterion, corrected)
- BIC (Bayesian Information Criterion)

$$AIC = -2\ln(L) + 2k$$

$$\text{AIC} = N * ln(\frac{SS_e}{N}) + 2K$$

N: *Number of observations*
SS_e: *Sum square of errors*
K: *Number of parameters*

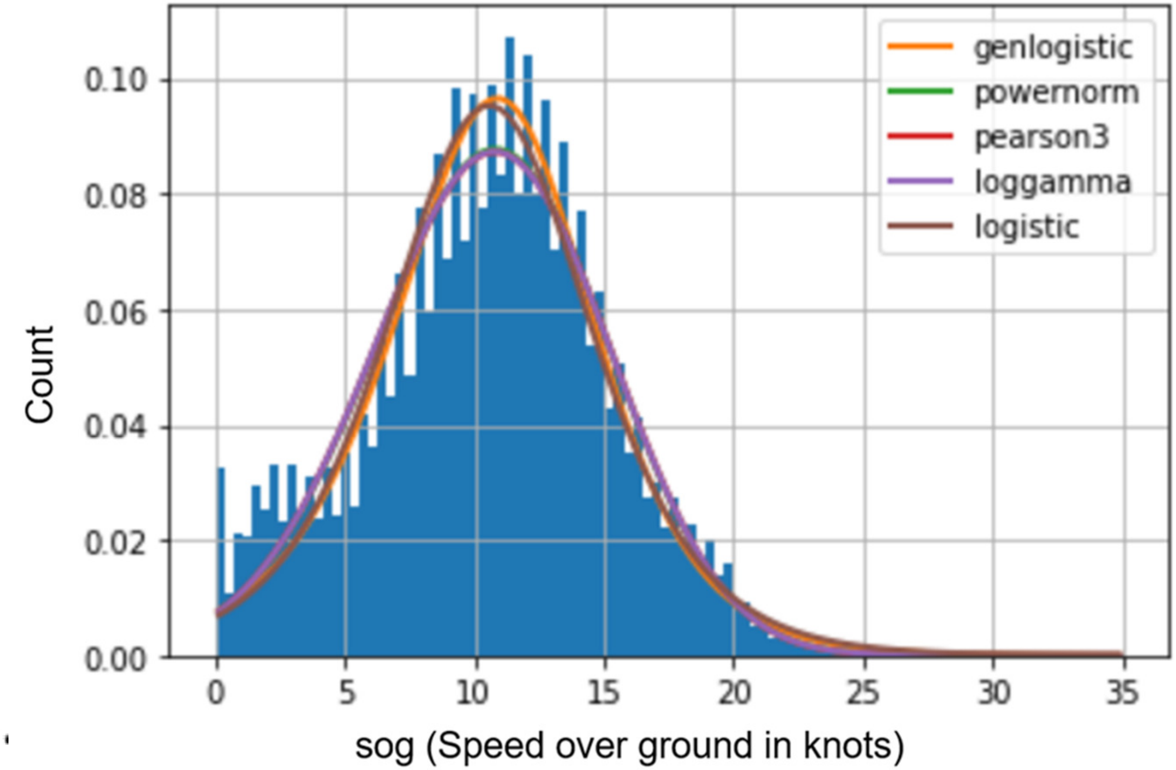
AIC is comprised of two important aspects:

- Maximum log-likelihood (measures how well the given model has captured the variance in the dependent variable)
- Number of parameters=K

Since a smaller AIC score is preferred, based on this formula adding more parameters penalizes the score. So if two models equally explain the variance in the given data, the model with fewer parameters will have a lower AIC score and be selected as the better fit model.

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$
$$= \frac{2kn}{n-k-1} - 2\ln(L)$$

$$\mathrm{BIC} = k\ln(n) - 2\ln(\widehat{L}).$$



Top 5 Fitted Distributions (Cargo Speed: 197475 reports)

	sumsquare_error	aic	bic
genlogistic	0.005878	1048.725316	-3.422171e+06
powernorm	0.006401	1296.360186	-3.405346e+06
pearson3	0.006494	1347.257116	-3.402498e+06
loggamma	0.006511	1328.668346	-3.401980e+06
logistic	0.006684	1016.553496	-3.396812e+06

In terms of mathematical expression, the partial correlation coefficient which assesses the relationship between variables X and Y while considering the influence of variable Z, is typically calculated using the given formula:

$$\rho_{XY.Z} = \frac{\rho_{XY} - \rho_{XZ} \cdot \rho_{YZ}}{\sqrt{(1-\rho^2_{XZ})(1-\rho^2_{YZ})}}$$

Here, ρ_{XY}

is the correlation coefficient between X and Y.

ρ_{XZ}

is the correlation coefficient between X and Z.

ρ_{YZ}

is the correlation coefficient between Y and Z.

The **numerator represents** the correlation between X and Y after accounting for their relationships with Z. The **denominator normalizes** the correlation by removing the effects of Z.