Time Series Models

2023. 8

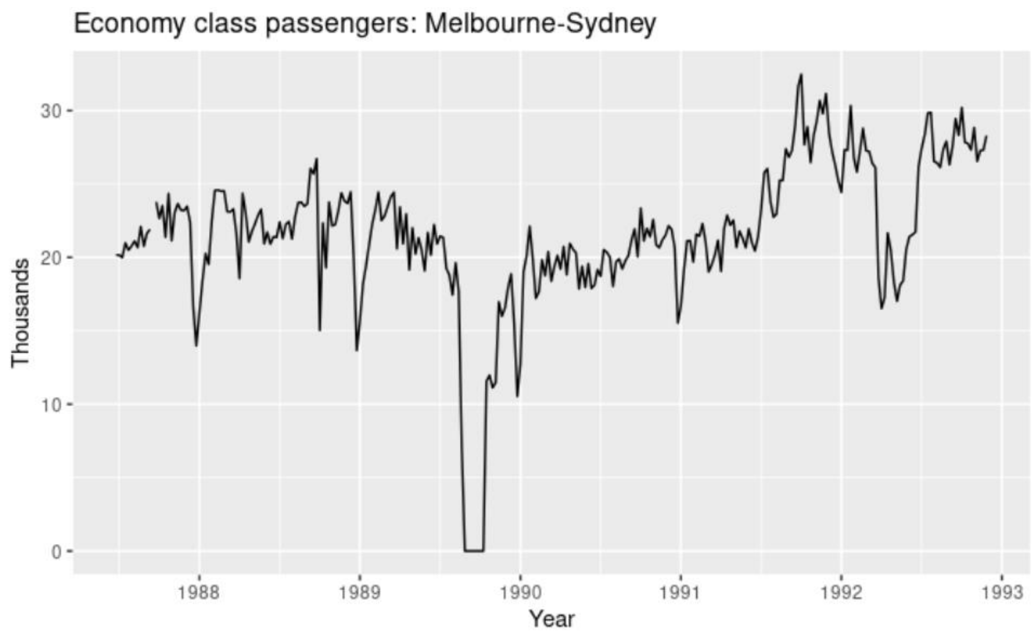
Yongjin Jeong, KwangWoon University

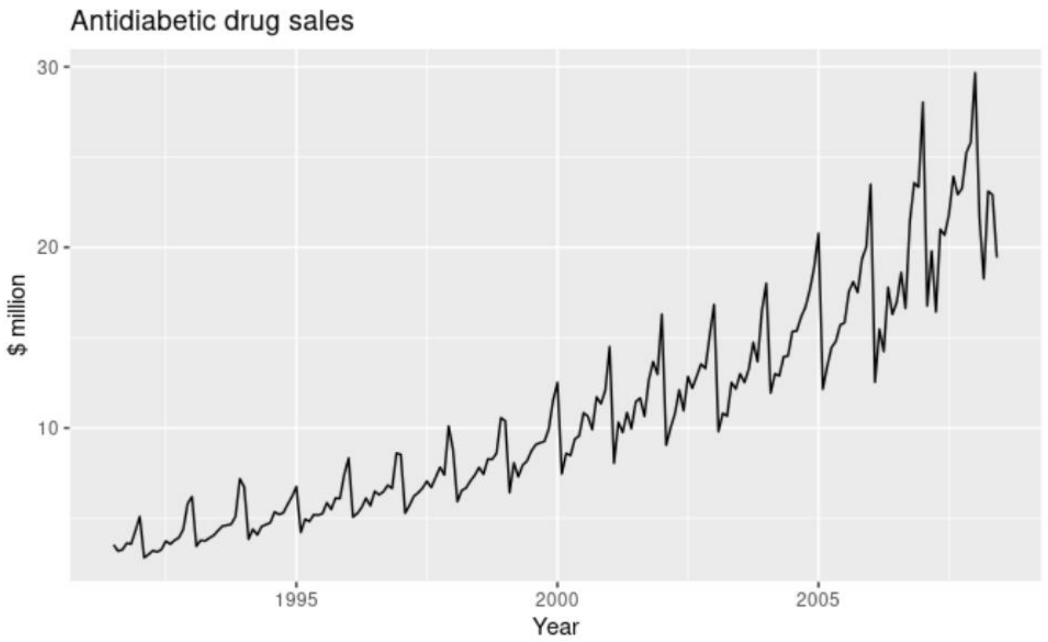
[reference]

- Hyndman, R.J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on February 2022
- Also available in https://otexts.com/fpp2

Time Plot

- **What is time plot?**
  - the observations are plotted against the time of observation
- **Examples**





- Industrial dispute (노동쟁의) – late 1989
- Reduced load in 1992 – replacement policy (economy->business)
- Large increase in the 2nd half of 1991
- Large dips around the start of each year – holiday effect

- clear increasing trend
- Sudden drop at the start of each year – government subsidization scheme (makes the patients stockpile drugs at the end of the year)
- Strong seasonal pattern - size increases as level of series increases
- Trend is changing slowly

Time series patterns

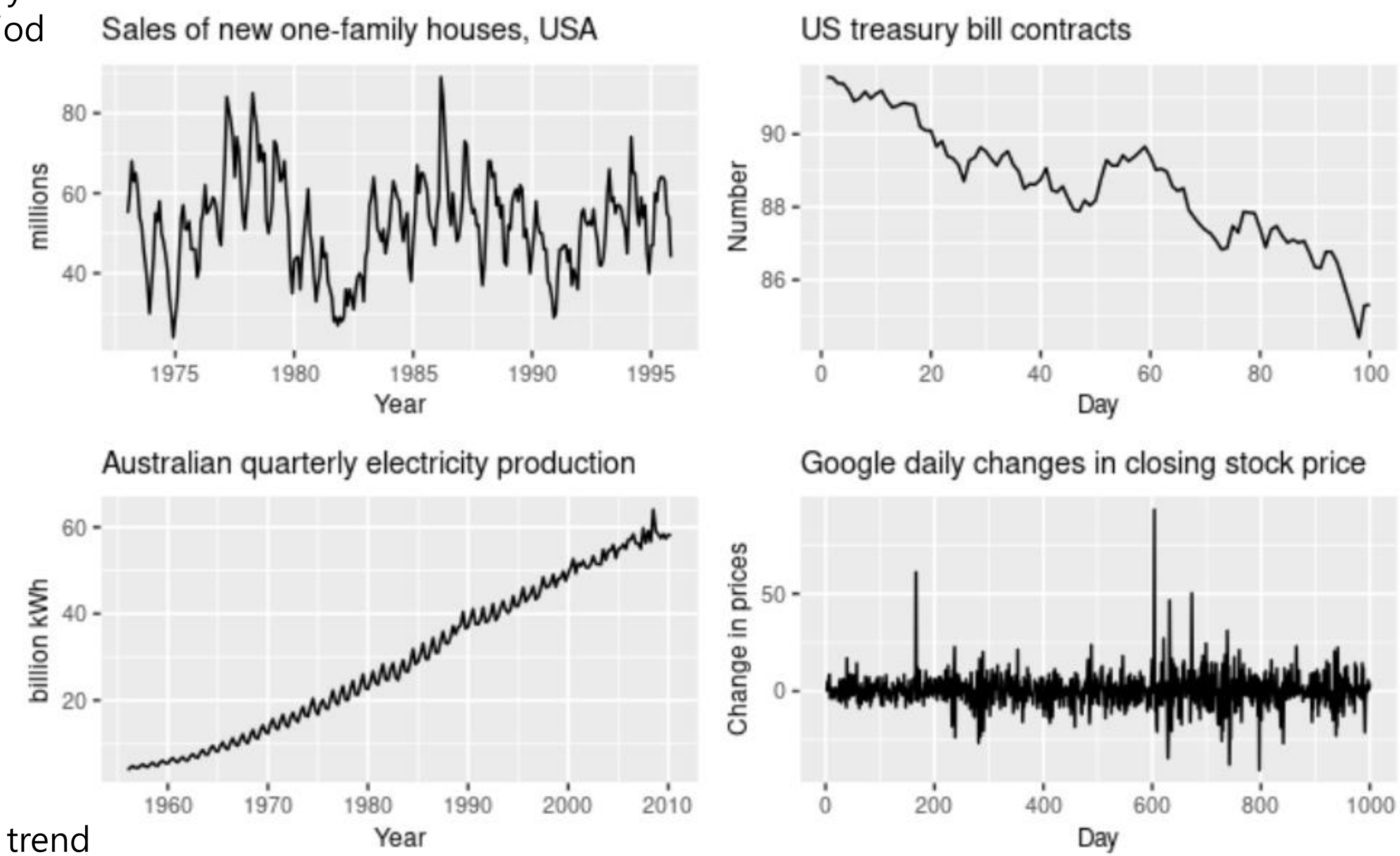
- **Trend (or changing direction)**
  - a long-term increase or decrease in the data
- **Seasonal** ← regular (period)
  - A time series is affected by seasonal factors such as the time of the year or the day of the week
  - Seasonality is always of a fixed and known frequency.
- **Cyclic** ← w/o fixed (or predictable) freq.
  - A *cycle* occurs when the data exhibit rises and falls that are not of a fixed frequency, usually due to economic conditions and often related to the "business cycle"
  - The duration of these fluctuation is usually at least 2 years.

Time series patterns: Examples

- Strong seasonality within each year,
- Also, some strong cyclic behavior with a period of about 6–10 years.

- (Chicago market for 100 consecutive days in 1981)
- No seasonality
- Obviously downward trend

국채계약



- Strong increasing trend
- Strong seasonality
- No evidence of any cyclic behavior

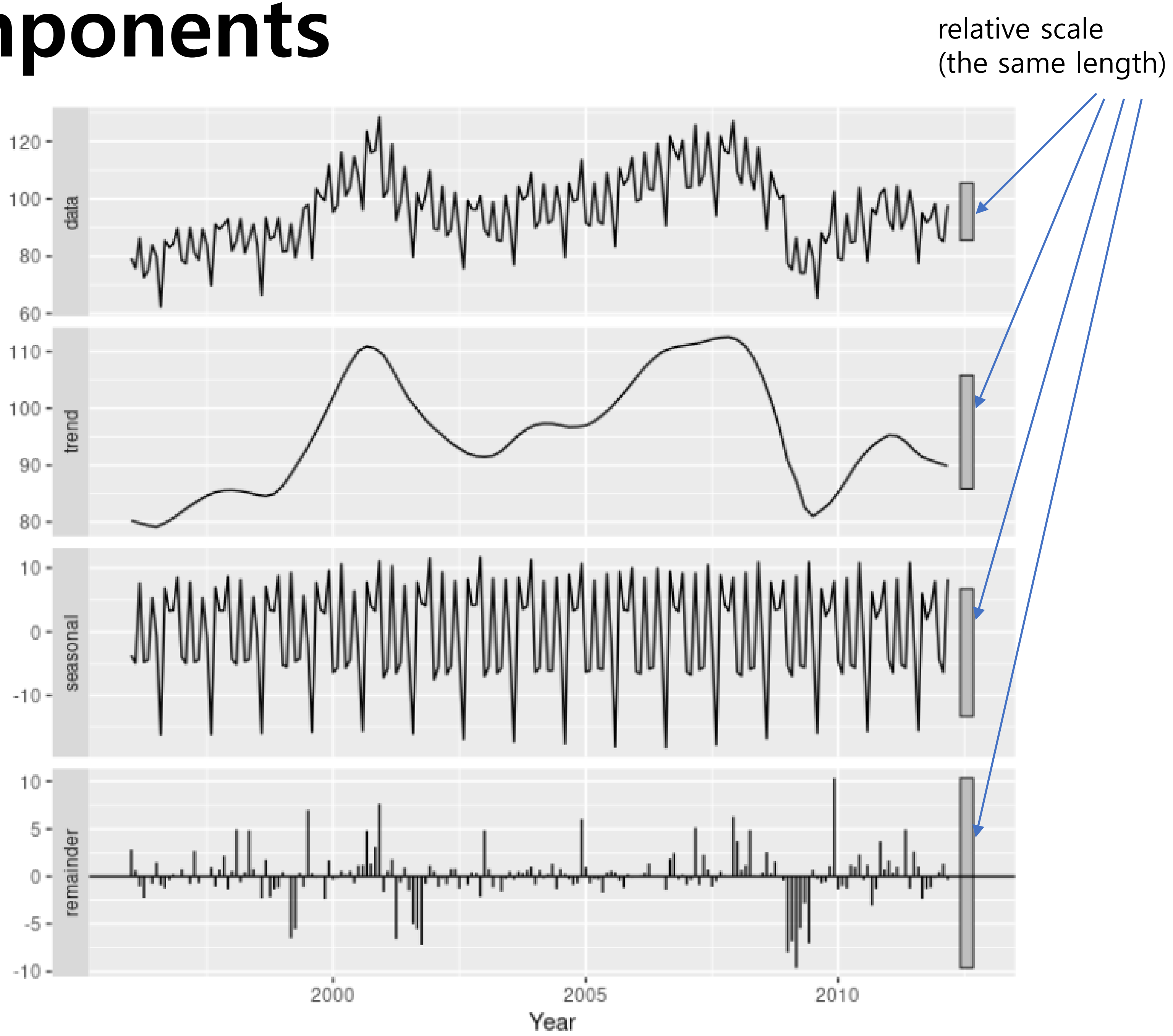
- No trend, seasonality, cyclic
- Random fluctuations (hard to predict)
- No strong patterns helpful for forecasting

Time Series Components

- **Additive decomposition**

$$y_t = S_t + T_t + R_t$$

y_t : data,
S_t : seasonal component,
T_t : trend-cycle component,
R_t : remainder component
(for all period of t)



Time Series Analysis using ARIMA

- Additive Components ← for understanding (simple)

$$Y_t = T_t + S_t + R_t$$

- Y_t is the observed time series data.
- T_t represents the trend component.
- S_t represents the seasonal component.
- R_t represents the residual or irregular component.

- ARIMA (AR+ I + MA) ← for forecasting
  - It aims to capture the regular, predictable patterns in the data using the AR, I, and MA components
  - The irregular term (R_t) typically contains noise, short-term variations, and any unpredictable factors that are not accounted for by the trend and seasonal components.
  - In a well-fitted ARIMA model, the residual term should ideally be close to white noise.

Autocorrelation (자기상관)

- **Correlation and Autocorrelation**
  - **Correlation** - measures the extent of a linear relationship between two variables
  - **Autocorrelation** - measures the linear relationship between *lagged values* of a variable
  - r_k : measures relationship between y_t and y_{t-k}

$$r = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum (x_t - \bar{x})^2} \sqrt{\sum (y_t - \bar{y})^2}}.$$

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

, where T is length of the time series

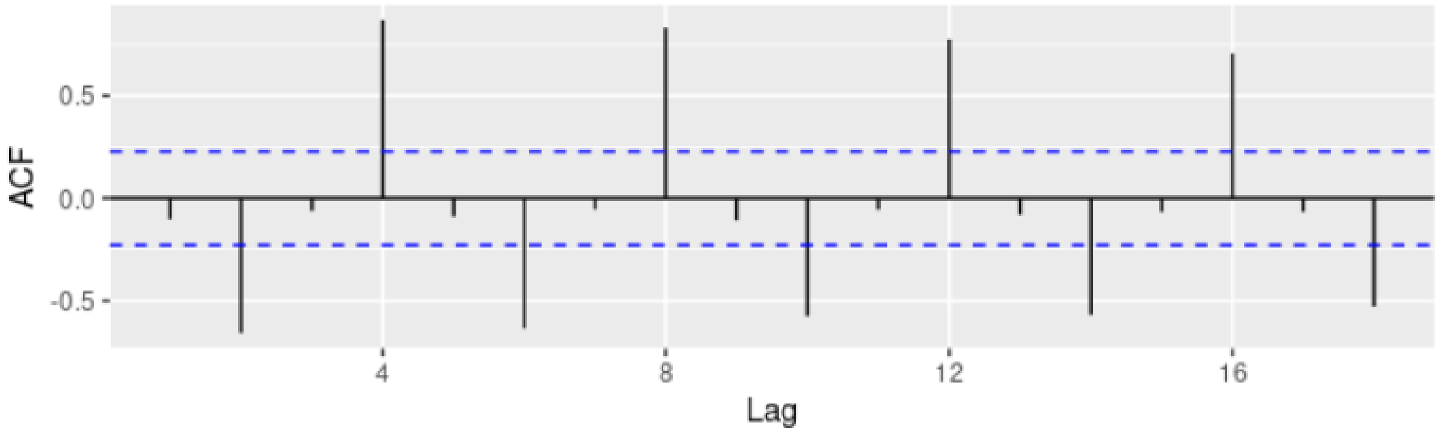
(in fact, the t-lagged signal is y_{t+k})
i.e. y_{t-k} : past observations

ACF plot

- **ACF (Auto-Correlation Function) or Correlogram**
  - Let's see first 9 autocorrelation coefficients.

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

r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9
-0.102	-0.657	-0.060	0.869	-0.089	-0.635	-0.054	0.832	-0.108



- r_4 is higher than other lags – due to seasonal pattern in the data (peaks and troughs(최저점) are 4 quarters apart)
- r_2 is more negative than the other lags because troughs tend to be two quarters behind peaks.
- The dashed blue lines indicate whether the correlations are significantly different from zero.

ACF plot

- **Trend and Seasonality in ACF plots**
  - When data have a trend:

    올라가는 trend 라면 아주 조금 떨어진 신호는 동일한 신호
    에 가까움, 즉 large correlation

    - The autocorrelations for small lags tend to be large and positive (because observations nearby in time are also nearby in size.) -> So the ACF of trended time series tend to have positive values that slowly decrease as the lags increase.
  - When data are seasonal:

    Season 만큼 떨어지면 똑같은 신호, 즉 large correlation

    - the autocorrelations will be larger for the seasonal lags (at multiples of the seasonal frequency) than for other lags.

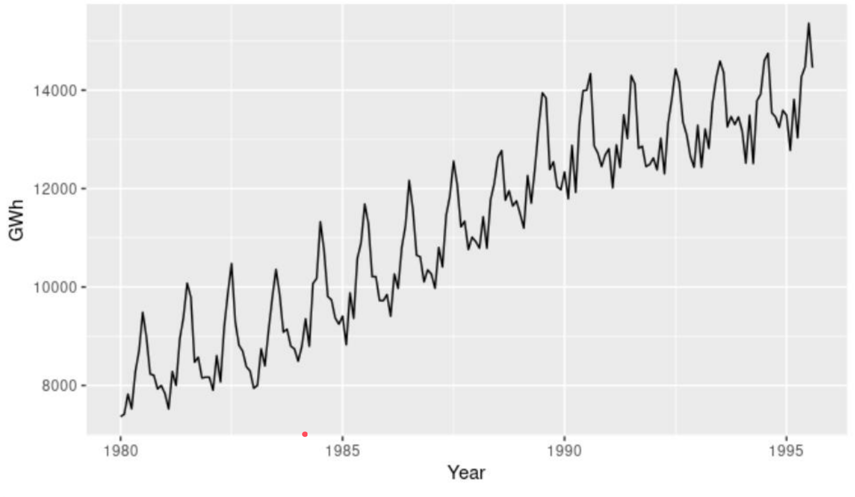
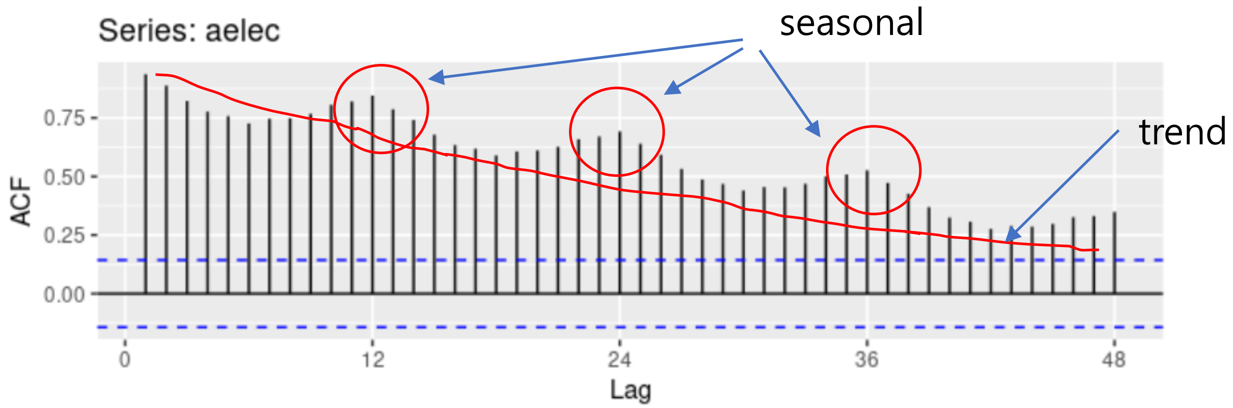


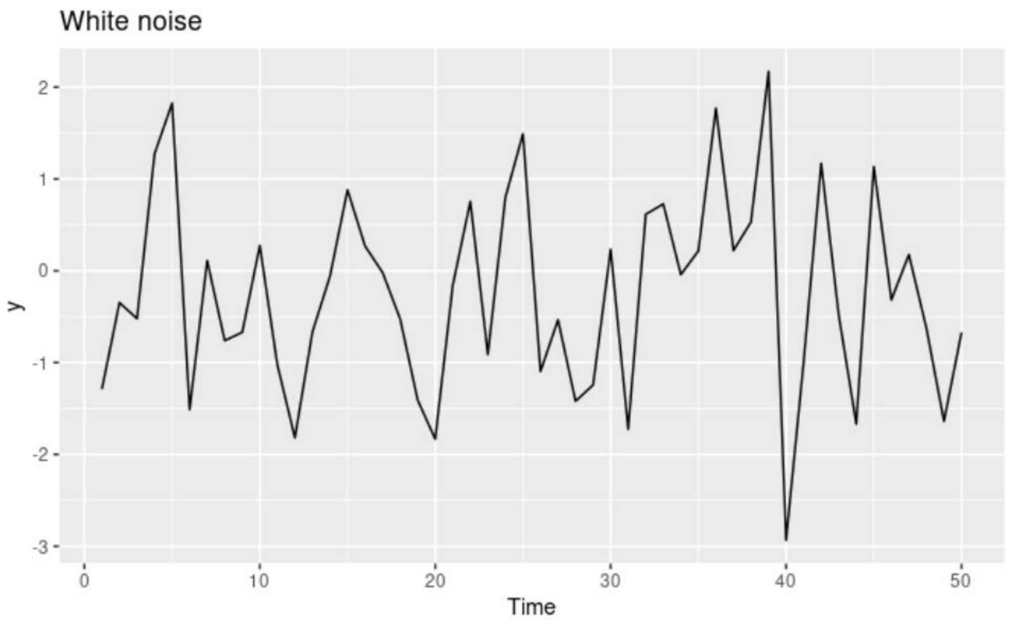
Figure 2.15: Monthly Australian electricity demand from 1980–1995.

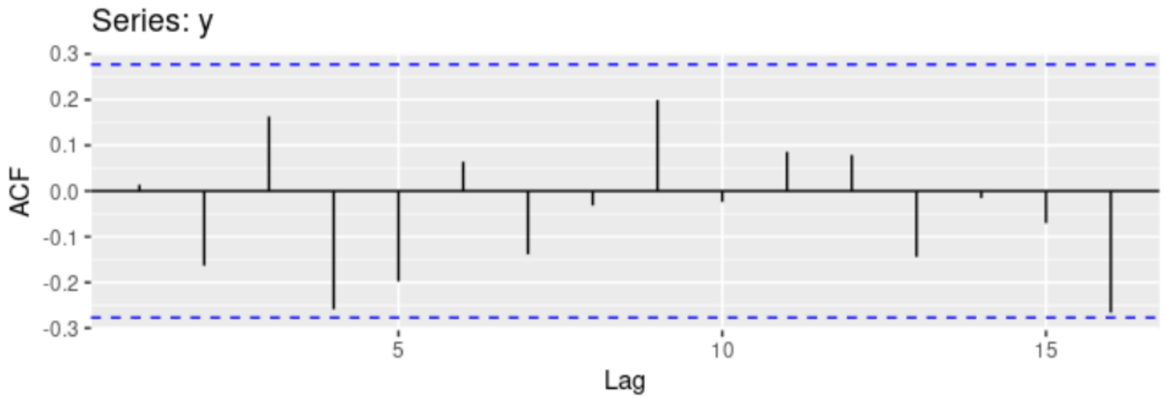


<Example of data set having both trend and seasonality>

ACF plot

- **White noise**
  - shows no autocorrelation -> each autocorrelation to be close to zero (not be exactly equal to zero because of some random variation)





Partial Autocorrelation (PAC) (그 사이는 제거)

- **PAC (Partial AC) Functions** t=i 와 t=i−k 시점 사이의 상관관계
  - measure the linear dependence of one variable after removing the effect of other variable(s) that affect to both variables.
  - Insights into the direct influence of past observations on the current observation (AR(p,d,q) 의 p 를 결정하는 데 중요한 역할)
  - For a time series, the partial autocorrelation between x_t and x_{t-h} is defined as **the conditional correlation between x_t and x_{t-h}, conditional on $x_{t-h+1}, \dots, x_{t-1}$**, the set of observations that come between the time points t and $t-h$.
    - The 1st order PAC is equal to the 1st order AC.
    - The 2nd order (lag) PAC is the correlation between values two time periods apart conditional on knowledge of the value in between.

$$\frac{\text{Covariance}(x_t, x_{t-2}|x_{t-1})}{\sqrt{\text{Variance}(x_t|x_{t-1})\text{Variance}(x_{t-2}|x_{t-1})}}$$

    - The 3rd order (lag) PAC is

$$\frac{\text{Covariance}(x_t, x_{t-3}|x_{t-1}, x_{t-2})}{\sqrt{\text{Variance}(x_t|x_{t-1}, x_{t-2})\text{Variance}(x_{t-3}|x_{t-1}, x_{t-2})}}$$

ACF and PACF

- **ACF (자기상관계수)**
  - ACF measures the correlation between a variable and its lagged values.

$$\mathrm{ACF}(k) = \mathrm{corr}(y_t, y_{t-k})$$

  - ACF gives us information about the correlation between the current observation and its lagged values.
  - *ACF* (0)=1
- **PACF (편자기상관계수) of lag *k***
  - PACF measures the correlation between a variable and its lag, <u>after removing the contributions of the intermediate lags</u>. It quantifies the direct relationship between an observation and its specific lag.

$$\mathrm{PACF}(k) = \mathrm{corr}(y_t, y_{t-k} | y_{t-1}, y_{t-2}, \ldots, y_{t-k+1})$$

  - PACF provides insights into the direct influence of past observations on the current observation, while controlling for the effects of intermediate lags.

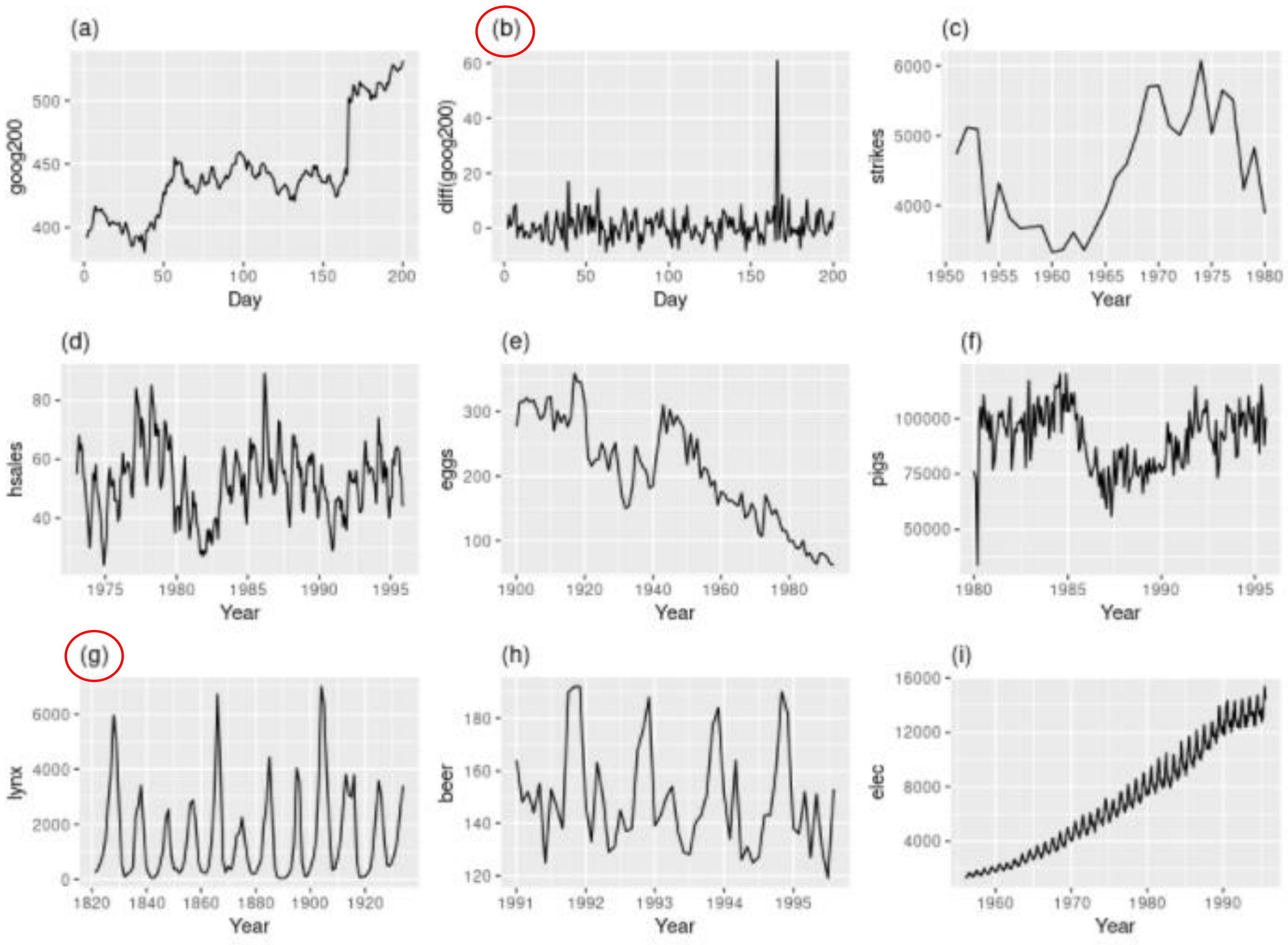
Stationarity

- **Stationarity**
  - If y_t is a stationary time series, then its statistical properties (such as mean, variance, and auto-covariance) remain constant over time.
  - Many statistical techniques and models, such as ARIMA models, rely on the assumption of stationarity
  - When dealing with non-stationary data, it's often necessary to apply techniques such as differencing or transformations to achieve stationarity before applying certain modeling methods.

(*) 전통적으로 시계열 분석은 현재 시점까지 시계열 데이터의 확률적 특성이 시간이 지나도 그대로 유지 될 것을 가정함. 즉, 시간과 관계 없이 평균과 분산이 불변이고, 두 개의 시점 간의 공분산이 다른 시점 과는 무관해야 함.

- How to identify stationarity
  - Time plot
  - The ACF of stationary data drops to zero relatively quickly
  - The ACF of non-stationary data decreases slowly.
  - For non-stationary data, the value of $r1$ is often large and positive.

Stationarity: Examples



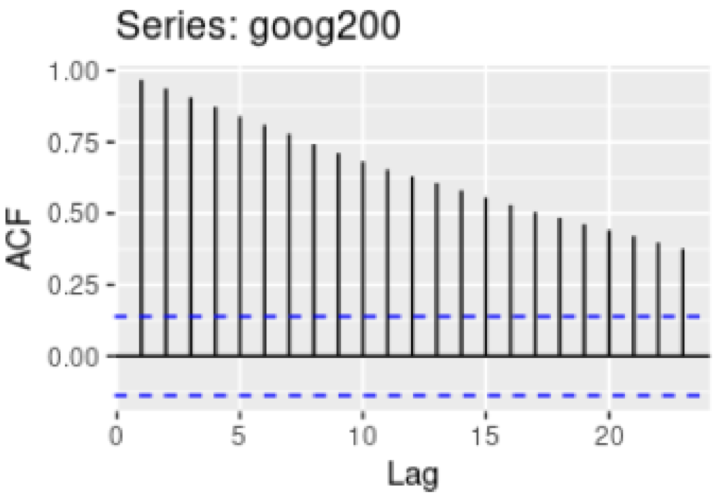
(d),(h),(i) – seasonality
(a),(c),(e),(f),(i) – trends
(b),(g) - stationary

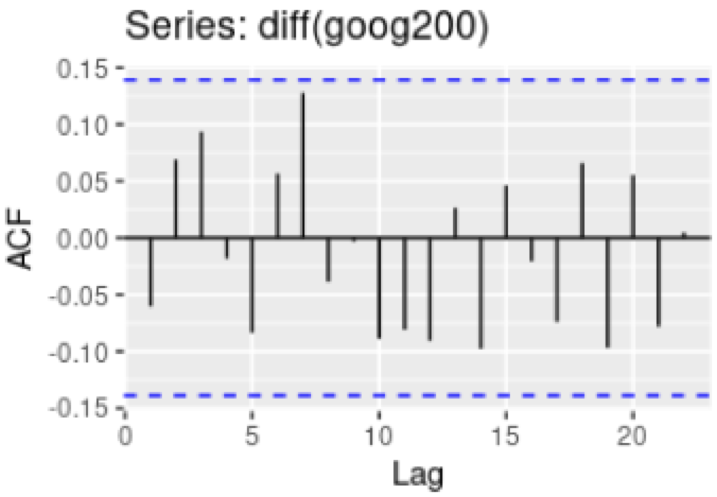
(g) Looks like a strong-cycles hence non-stationary, but these cycles are aperiodic. In the long-term, the timing of these cycles is not predictable. Hence the series is stationary.

(a) Google stock price for 200 consecutive days;
(b) Daily change in the Google stock price for 200 consecutive days

Differencing

- **Differencing (차분)**
  - one way to make a non-stationary time series stationary — compute the differences between consecutive observations. ((a) was non-stationary, but (b) was stationary)
  - Differencing can help stabilize the mean of a time series by removing changes in the level of a time series, and therefore eliminating (or reducing) trend and seasonality.





Differencing

- **Differencing**
  - Used to transform a non-stationary time series into a stationary one. (This stability provides a more predictable framework for modeling and forecasting.)
  - First-order differencing
    - each value represents the change from the previous value

$$y'_t = y_t - y_{t-1}$$

  - Higher-order differencing
    - If the resulting series after first-order differencing is still not stationary, you can apply differencing again.

$$\begin{aligned} y''_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2}. \end{aligned}$$

  - Inverse differencing (integration)
    - After modeling, when you want to make forecasts in the original scale of the data, you need to "reverse" the differencing.

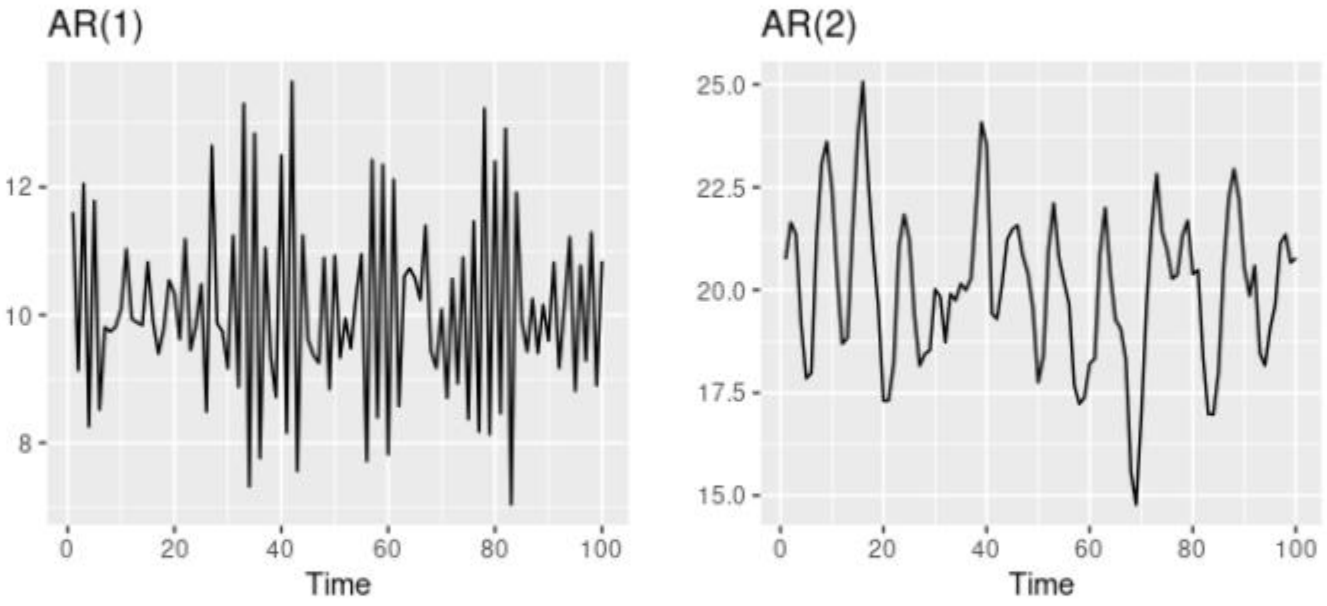
$$y_t = y_{t-1} + y'_t$$

AR(AutoRegressive) model

- **What is AR model?**
  - forecast the variable of interest using a linear combination of past values of the variable (multiple regression with lagged values of y_t as predictors)
  - AR(p) model:

$$\boxed{y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t}$$

white noise (only changes the scale of the series, not the pattern)



$$y_t = 18 - 0.8 y_{t-1} + \varepsilon_t.$$

$$y_t = 8 + 1.3 y_{t-1} - 0.7 y_{t-2} + \varepsilon_t.$$

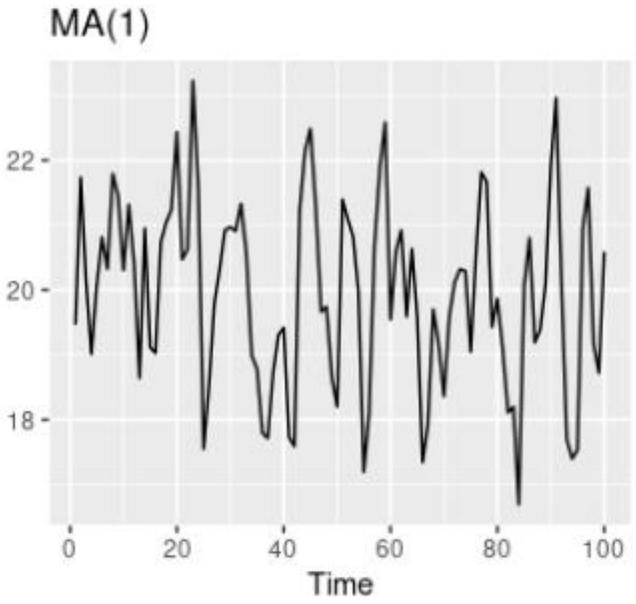
ε_t is normally distributed white noise, N(0,1)

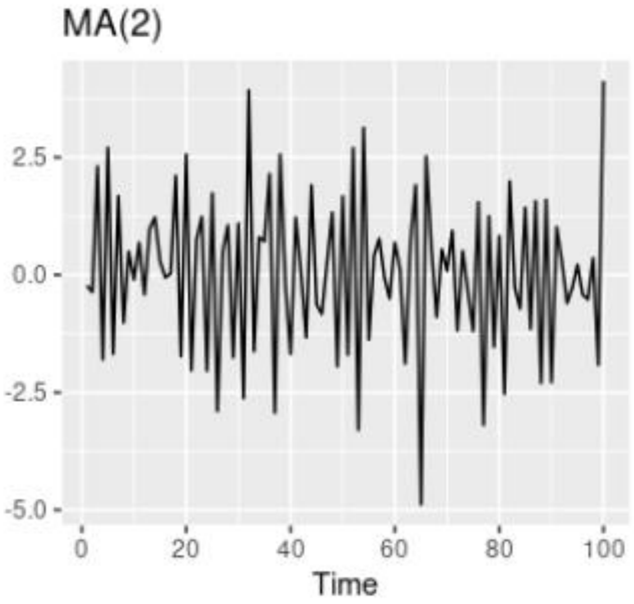
MA(Moving Average) model

- **MA(Moving Average) (**≠ moving average smoothing for reducing noise**)**
  - Rather than using past values of the forecast variable in a regression, a MA model uses **past forecast errors** (between observation and a residual error term from a previous time step) – try to adjust for the errors encountered in previous forecasts.
  - MA(q) model:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

ε_t is white noise





$$y_t = 20 + \varepsilon_t + 0.8\varepsilon_{t-1}$$

$$y_t = \varepsilon_t - \varepsilon_{t-1} + 0.8\varepsilon_{t-2}$$

ARMA and ARIMA models

- **ARMA(Auto-Regressive Moving Average) Model**

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

  – Predictors include both **lagged values of y_t and lagged errors.**
  – Conditions on coefficients ensure stationarity and invertibility.
  – (*) Any MA(q) process can be written as an AR(1) process if we impose some constraints on the MA parameters. Then the MA model is called “invertible”.

- **ARIMA(Auto-Regressive Integrated Moving Average) Models**
  – Combine ARMA model with **differencing**
  – **ARIMA(p,d,q) model**
    - AR: p = order of the autoregressive part
    - I: d = degree of first differencing involved
    - MA: q = order of the moving average part

Non-seasonal ARIMA model

- **ARIMA(Auto-Regressive Integrated Moving Average)**
  - Integration is the reverse of differencing
  - Combining differencing with AR and MA, non-seasonal ARIMA model is:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

(y_t' is the difference series)

- ARIMA(*p*,*d*,*q*) model:

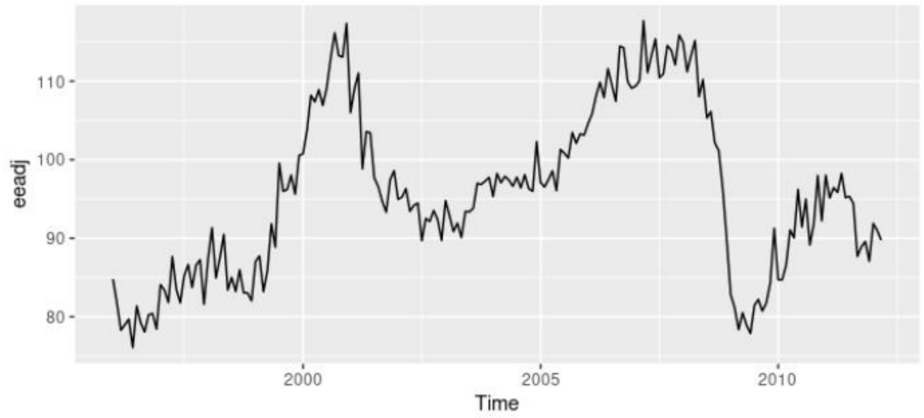
$p =$	order of the autoregressive part;
$d =$	degree of first differencing involved;
$q =$	order of the moving average part.

- Special cases:

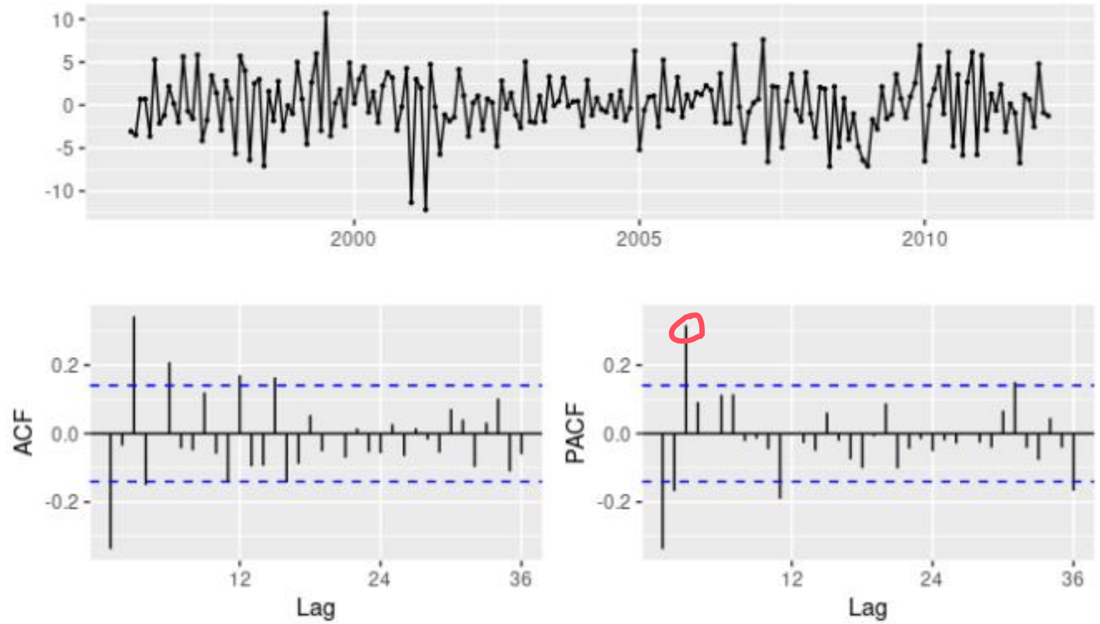
White noise	ARIMA(0,0,0)
Random walk	ARIMA(0,1,0) with no constant
Random walk with drift	ARIMA(0,1,0) with a constant
Autoregression	ARIMA(p,0,0)
Moving average	ARIMA(0,0,q)

ARIMA model: Example

- **Example:**



Time series data: non-stationary



First-difference – look stationary
(no more further difference needed)

- PACF suggests AR(3) model – try ARIMA(3,1,0)

p d

Seasonal ARIMA model

- **Seasonal ARIMA**
  - Include additional seasonal terms in the ARIMA model

$$\text{ARIMA} \quad \underbrace{(p,d,q)}_{\substack{\uparrow \\ \text{Non-seasonal part} \\ \text{of the model}}} \quad \underbrace{(P,D,Q)_m}_{\substack{\uparrow \\ \text{Seasonal part of} \\ \text{of the model}}}$$

m : number of observations per year

  - For example, ARIMA(1,1,1)(1,1,1)$_4$ model is a quarterly data (*m*=4)

- ARIMA & SARIMA
  - It aims to find the appropriate values of the parameters (p, d, q) and (P,D,Q,S) in order to minimize the remaining signal, leaving residuals that resemble white noise.
  - Model refinement: If the residuals exhibit any patterns or systematic behavior, further adjustments to the model may be needed.

SARIMA model and Prophet

	SARIMAX	Prophet
strength	- Solid Theoretical Foundation: suitable for capturing linear relationships and stationary time series patterns - Interpretability: interpretable parameter estimates, such as autoregressive (AR) and moving average (MA) coefficients, - Exogenous Variables: SARIMAX can incorporate exogenous variables, which is useful if you have additional information that can help improve forecasting accuracy.	- Ease of Use: user-friendly and requires minimal tuning. - More Flexibility: more robust against anomalies and nonlinearity compared to traditional models like SARIMAX. - Automatic Seasonality Detection: it simplifies the modeling process. - Forecast Components: separate components for trends, seasonality, and holidays, making it easier to interpret the underlying patterns.
weakness	- Limited Flexibility: may not perform well at abrupt changes or unusual patterns - Manual Tuning: requires manual selection of model orders (p, d, q, P, D, Q, and s) and exogenous regressors, which can be time-consuming and require domain expertise.	- Limited Theoretical Foundation: it lacks the deep theoretical foundation of traditional time series models like SARIMAX. - Limited Control: simplicity can limit its customization options for advanced users who want more control over the modeling process.
If your data has clear linear relationships and you have domain expertise to manually tune the model, SARIMAX might be a good choice. On the other hand, if you're looking for an easy-to-use and flexible model that can handle various data characteristics, Prophet could be more suitable. It's also worth noting that other advanced forecasting techniques, such as machine learning methods (e.g., LSTM, GRU, XGBoost), might be worth exploring if your data is highly complex or nonlinear.		