BIG DATA ANALYTICS

WEEK-16 | Application 3 - Forecasting and Anomaly detection

Yonsei University
Jungwon Seo

Time Series Analysis Applications

- Forecast
- Anomaly Detection

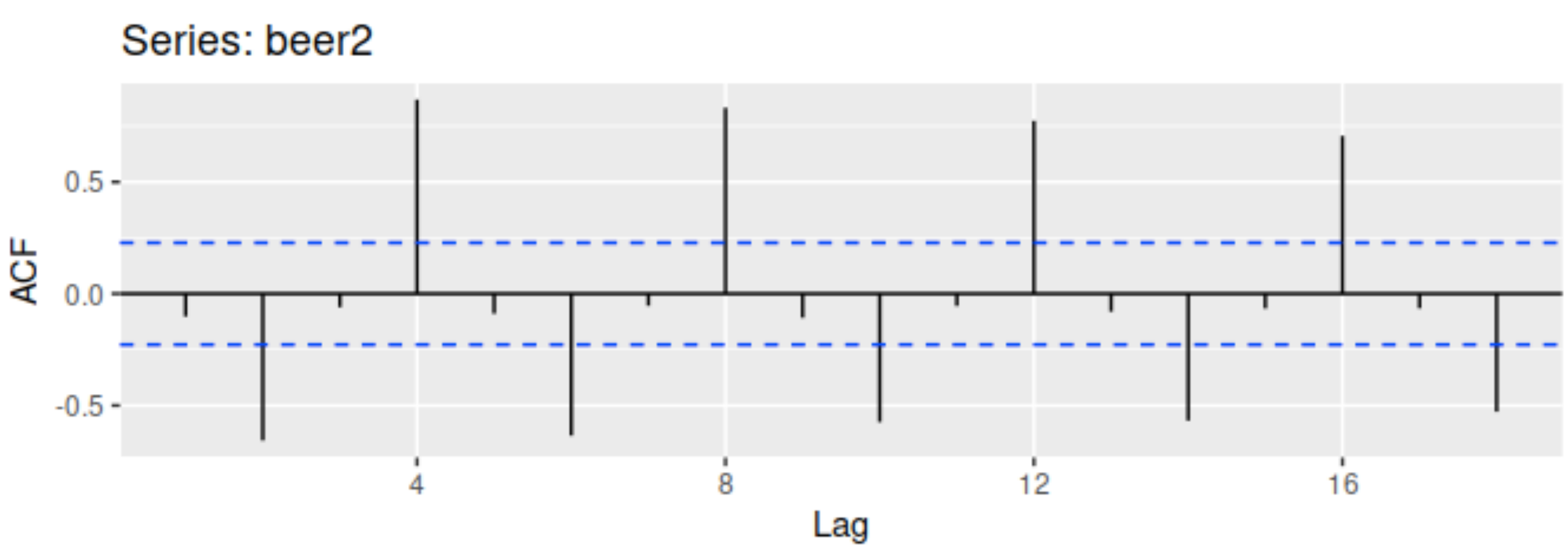
Autocorrelation

- 자기상관
  - 시계열의 시차 값(lagged values) 사이의 선형 관계
  - r1은 yt와 y_{t-1} 사이의 관계, r2는 yt와 y_{t-2} 사이의 관계

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

  -

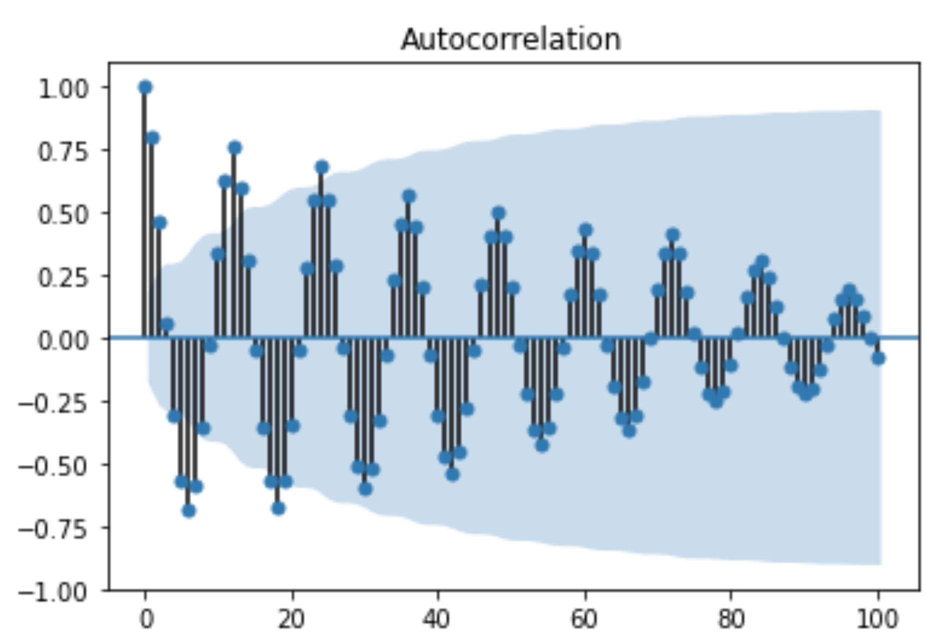
r1	r2	r3	r4	r5	r6	r7	r8	r9
-0.102	-0.657	-0.060	0.869	-0.089	-0.635	-0.054	0.832	-0.108

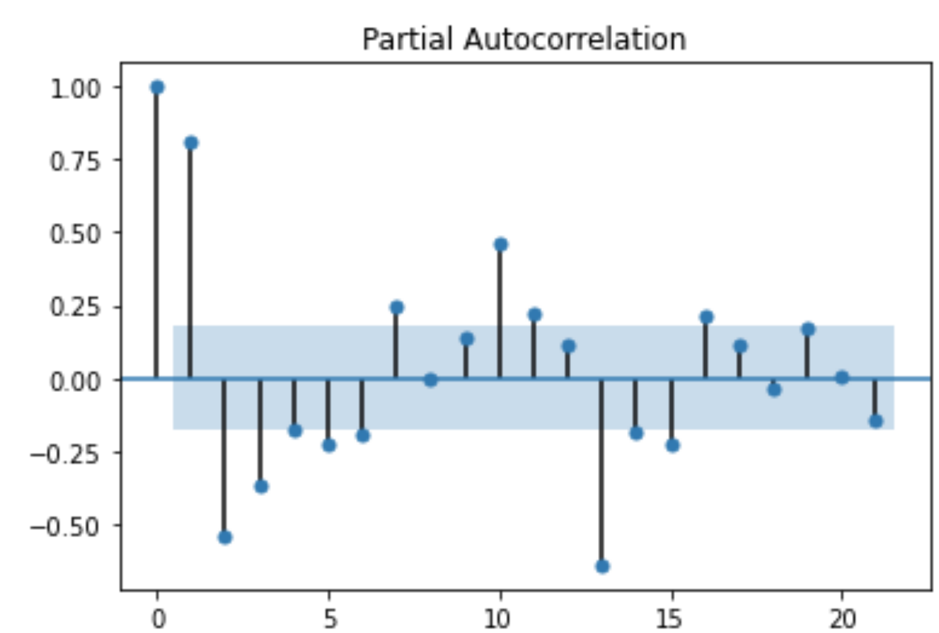


Autocorrelation

- ACF: Autocorrelation Function
  - 이전 시점을 모두 고려하여 두 시점의 상관관계를 구한다
  - t3와 t4의 상관관계를 구할때, t1,t2도 같이 고려
  - MA의 차수를 결정할 때 사용

- PACF: Partial Autocorrelation Function
  - 이전 시점은 고려하지 않고 두시점의 상관관계를 구한다
  - t4는 t3에 의해서만 영향을 받는다고 가정
  - AR의 차수를 결정 할 때 사용

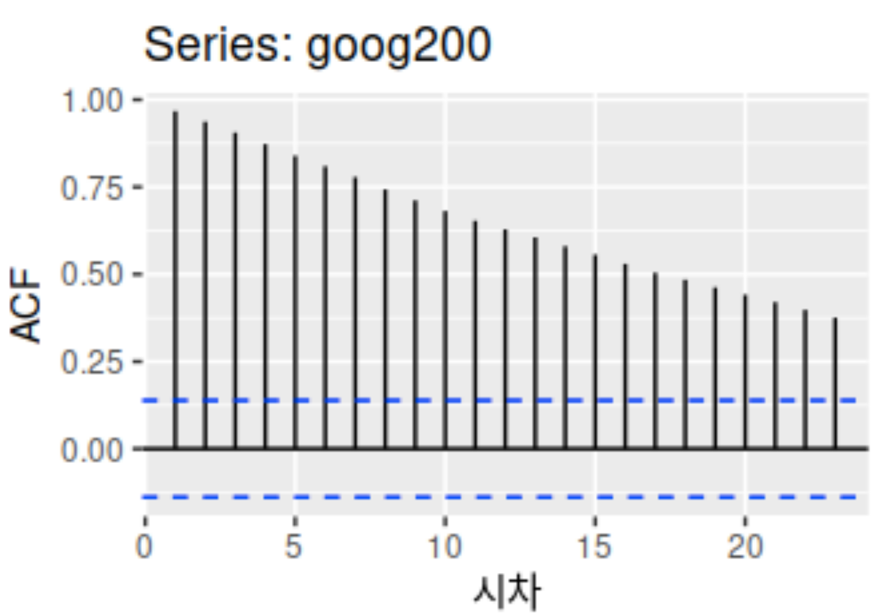


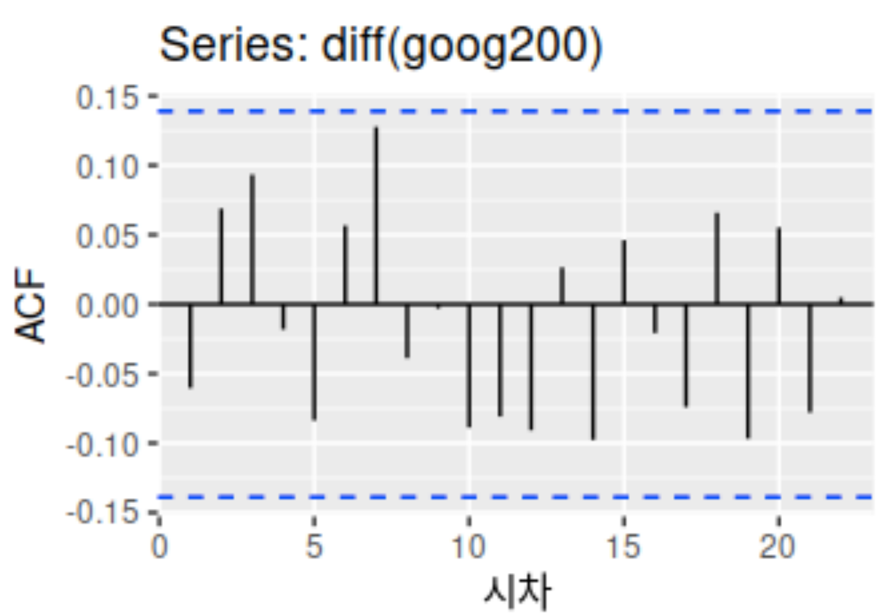


Stationarity

- 정상성의 조건
  - 평균이 일정 (No Trend)
  - 표준편차가 일정
  - No Seasonality

- 어떻게 정상성을 만족하게 만들 수 있을까?
  - 차분(Differencing)을 활용

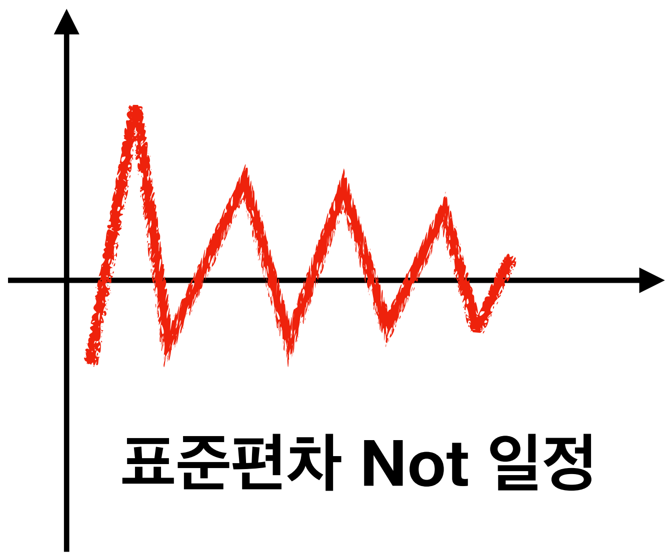




- Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 2020.11.11.

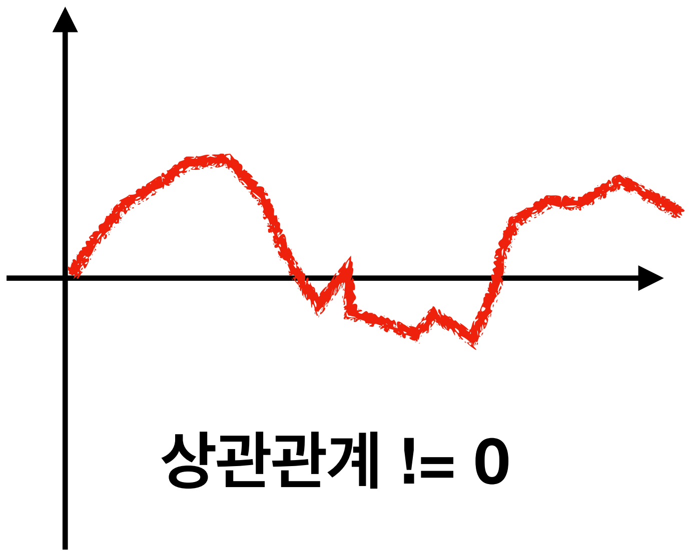
White Noise

- 백색잡음의 조건
  - 평균 = 0
  - 표준편차가 일정
  - 일정 구간간의 상관관계 0



표준편차 Not 일정

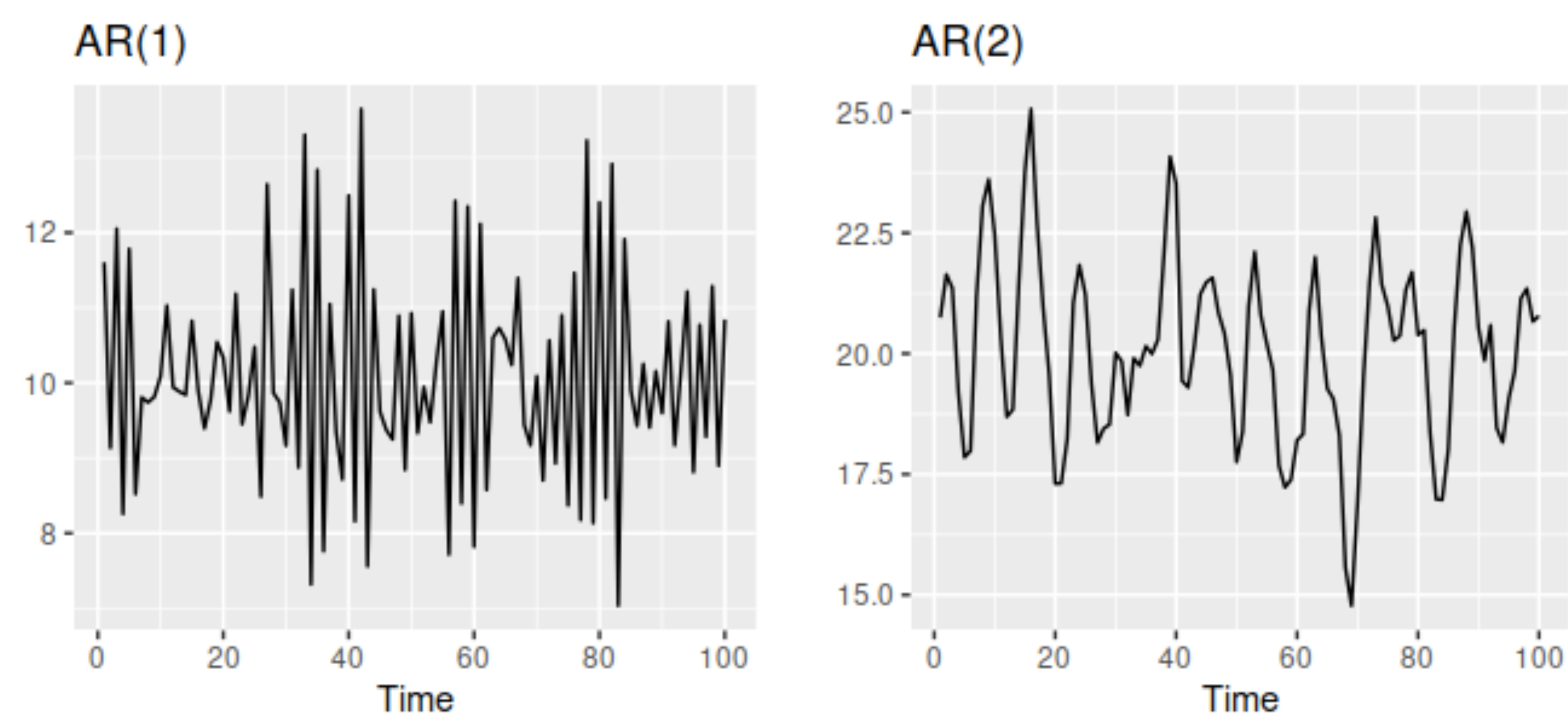
- 왜 중요할까?
  - 시계열 데이터 = 신호 + 잡음
  - 잡은은 예측을 할 수 없기 때문에, 만약에 신호를 확실하게 포착 했다면, 잡음과 무관하게 시계열 데이터 분석을 완료 할 수 있음
    (잡음 까지 맞추려고 할 필요 X)



상관관계 != 0

Autoregressive Model

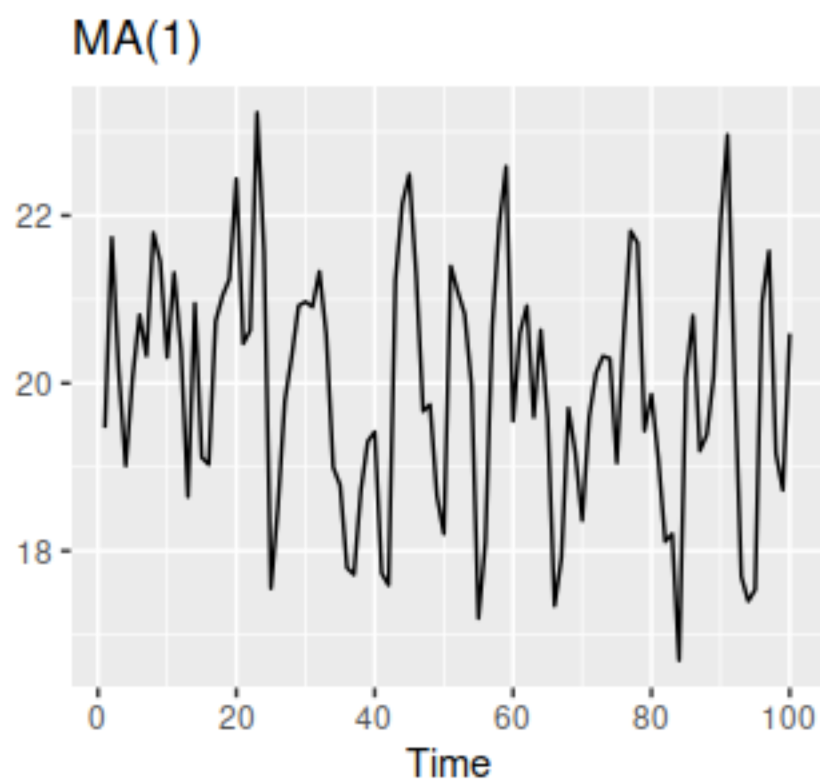
- 자기회귀 모델에서는, 변수의 과거 값의 선형 조합을 이용하여 관심 있는 변수를 예측
  - $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$
-

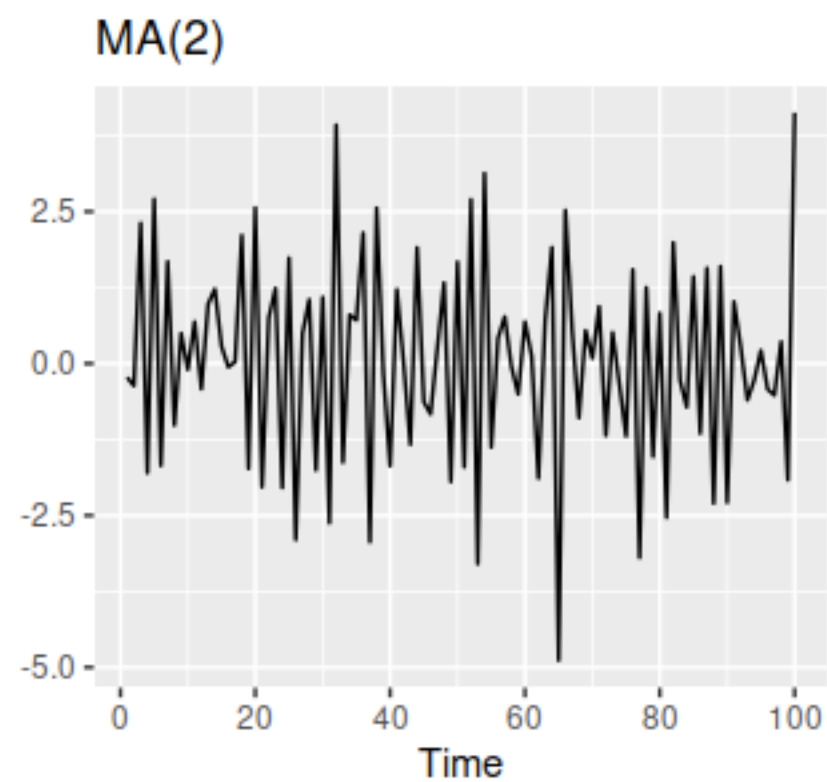


- Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 2020.11.11.

Moving Average Model

- 이동 평균 모델은 회귀처럼 보이는 모델에서 과거 예측 오차 (forecast error)을 이용합니다.
  - 이전 강의에서 했던, 30일 이동평균은 이동 평균 **평활**
    - 과거의 주기-추세를 측정할 때 사용
    - MA Model은 예측 할때 사용
  - $y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$





- Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 2020.11.11.

ARMA

- Auto
  Regressive
  Moving
  Average

- 과거의 값으로 미래를 예측(AR), 과거에 발생하는 오류로 미래의 오류를 예측 (MA)

- ARMA(1,1): $\ell_t = \beta_0 + \beta_1 \ell_{t-1} + \phi_1 \epsilon_{t-1} + \epsilon_t$

- ACF와 PACF를 활용해 찾는다

ARIMA

- 이동평균을 누적한 자기회귀
    - Auto
      Regressive
      Integrated(누적)
      Moving
      Average
    - $y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$
- Trend가 존재하는 데이터에 대응하기 위해

SARIMA

- 계절성 ARIMA 모델 : 계절성이 있는 데이터에 대응하기위해
  - Seasonality
    Auto
    Regressive
    Integrated
    Moving
    Average

- ARIMA$\underbrace{(p,d,q)}^{\text{비계절성 부분}}$ $\underbrace{(P,D,Q)m}^{\text{계절성 부분}}$
  - m = seasonal factor
  - p: AR 차수, d: 1차 차분이 포함된 정도, q: MA 차수

- ARIMA(1,1,1)(1,1,1)4
  - $(1-\phi_1 B)\ (1-\Phi_1 B^4)(1-B)(1-B^4)y_t = (1+\theta_1 B)\ (1+\Theta_1 B^4)\varepsilon_t$

References

- Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 2020.11.11.

E.O.D