CH5. 확률과정

확률 과정

- **확률과정(stochastic process)이란**
  - 확률법칙에 의해 생성되는 일련의 통계적인 현상
  - 확률공간에서 정의되는 확률변수들의 모임

$$\{Z_t, t \in T\}$$

  - 연속형 확률과정 / 이산형 확률과정
  - 시계열은 T가 시간의 부분집합인 확률과정
  - 실현값(realization) 또는 표본통로(sample path) : 관측된 확률과정

정상 확률 과정

- **정상성(stationarity)**

정상 (stationary) = 시간 불변 (time-invariant)

- 시계열의 확률적인 성질들이 시간의 흐름에 따라 불변
- 평균, 분산, 자기공분산 등에 체계적인 변화가 없음
- 주기적인 변화가 없음

※ 대부분의 시계열의 이론들이 정상성을 가정하고 전개되어 있기 때문에 정상적이 아닌 시계열은 정상시계열로 변환해야함

정상 확률 과정

- **강정상성(strict stationarity)**

  - Strong stationarity

  - 결합확률분포가 시간에 따라 변하지 않음

  - [정의] $\{Z_t, t = 1,2,\dots\}$ 가 강정상성을 갖는다 $\overset{\text{def}}{=}$

    임의의 자연수 $t_1, \dots, t_k, h$ 와 실수 $z_1, \dots, z_k$ 에 대하여 다음 등식이 성립

$$\Pr\left(Z_{t_1} \le z_1, \dots, Z_{t_k} \le z_k\right) = \Pr\left(Z_{t_1+h} \le z_1, \dots, Z_{t_k+h} \le z_k\right)$$

정상 확률 과정

- **약정상성(weakly stationarity)**

  - Weak stationarity, second-order stationarity, covariance stationarity
  - 평균, 분산, 자기공분산이 시간에 따라 변하지 않음
  - [정의] $\{Z_t, t = 1,2,...\}$ 가 정상성을 갖는다 $\stackrel{\text{def}}{=}$

    모든 t에 대하여 $E|Z_t|^2 < \infty$ 이고,

    (1) 평균 일정 : $E(Z_t) = \mu, \ \forall t$

    (2) 분산 일정 : $Var(Z_t) = \sigma^2, \ \forall t$

    (3) 자기공분산이 시차 h에만 의존 : $Cov(Z_t, Z_{t+h}) = \gamma_h, \ \forall t$

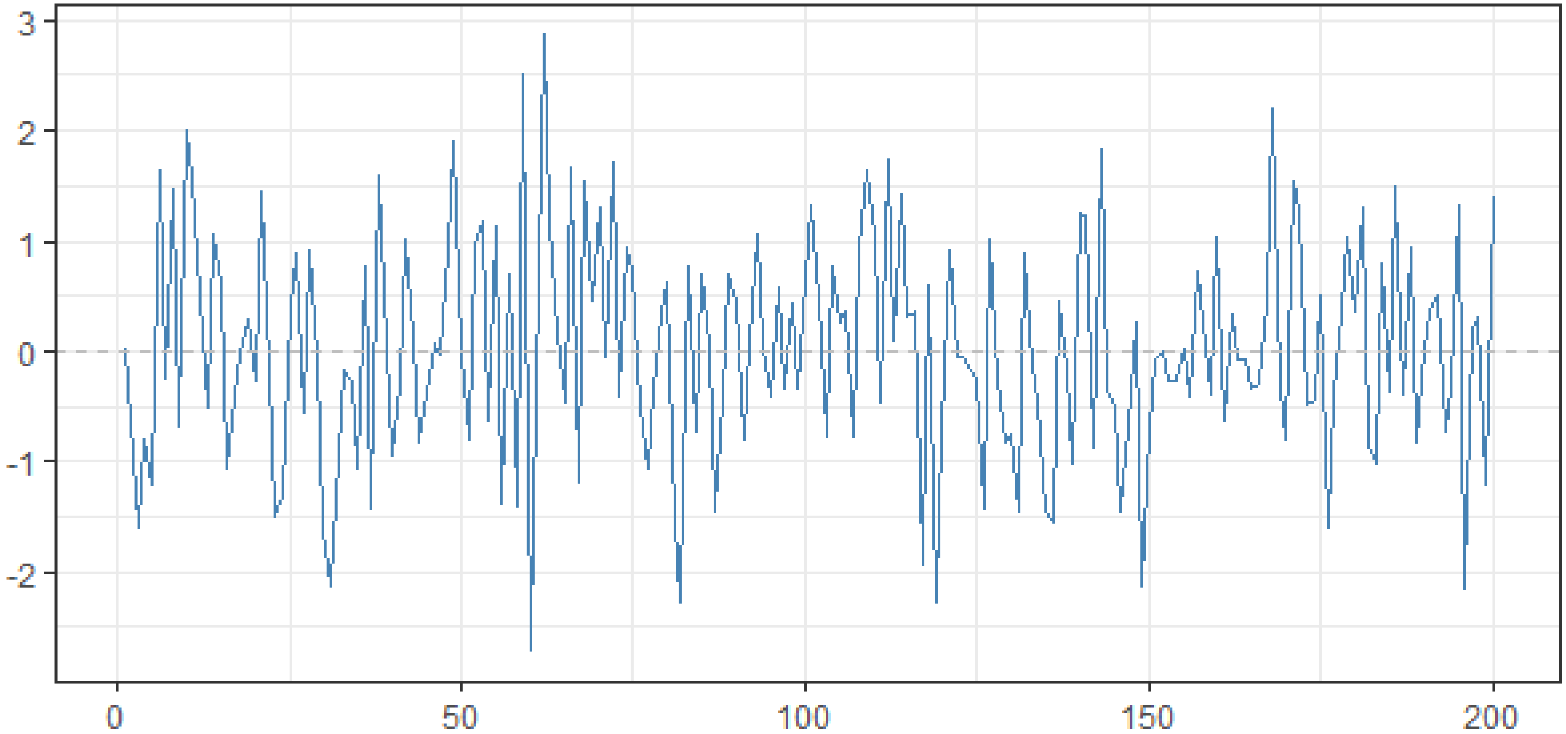
정상 확률 과정

- **강정상과 약정상성의 관계**

  - 일반적으로 두 정상성은 포함관계가 없음
  - 분산이 존재하는 강정상 시계열 => 약정상 시계열
  - Gaussian process인 경우(결합확률분포가 다변량 정규분포인 경우)에는 두 정상성의 개념은 일치

정상 확률 과정

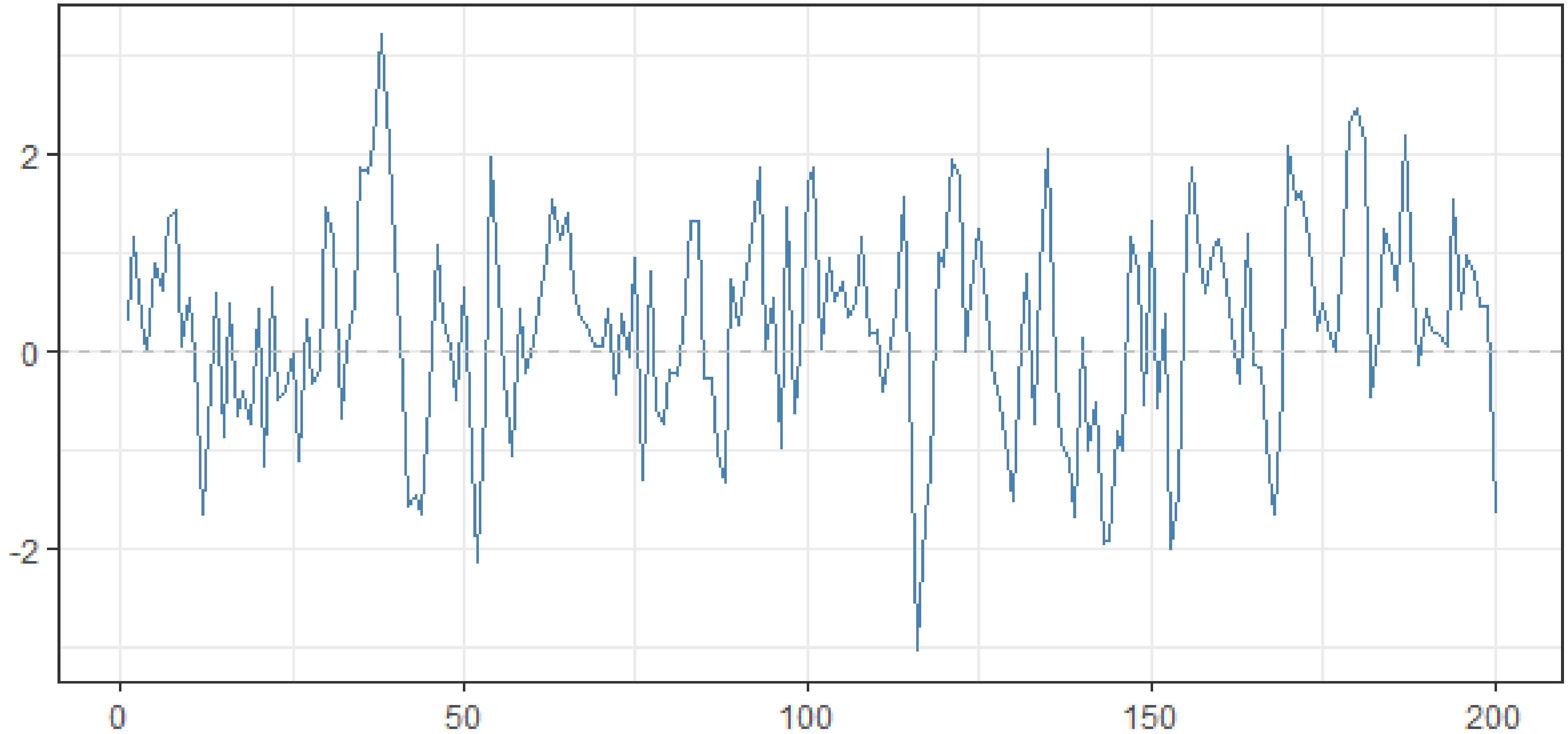
- **정상 시계열 : N(0,1) iid**



정상 확률 과정

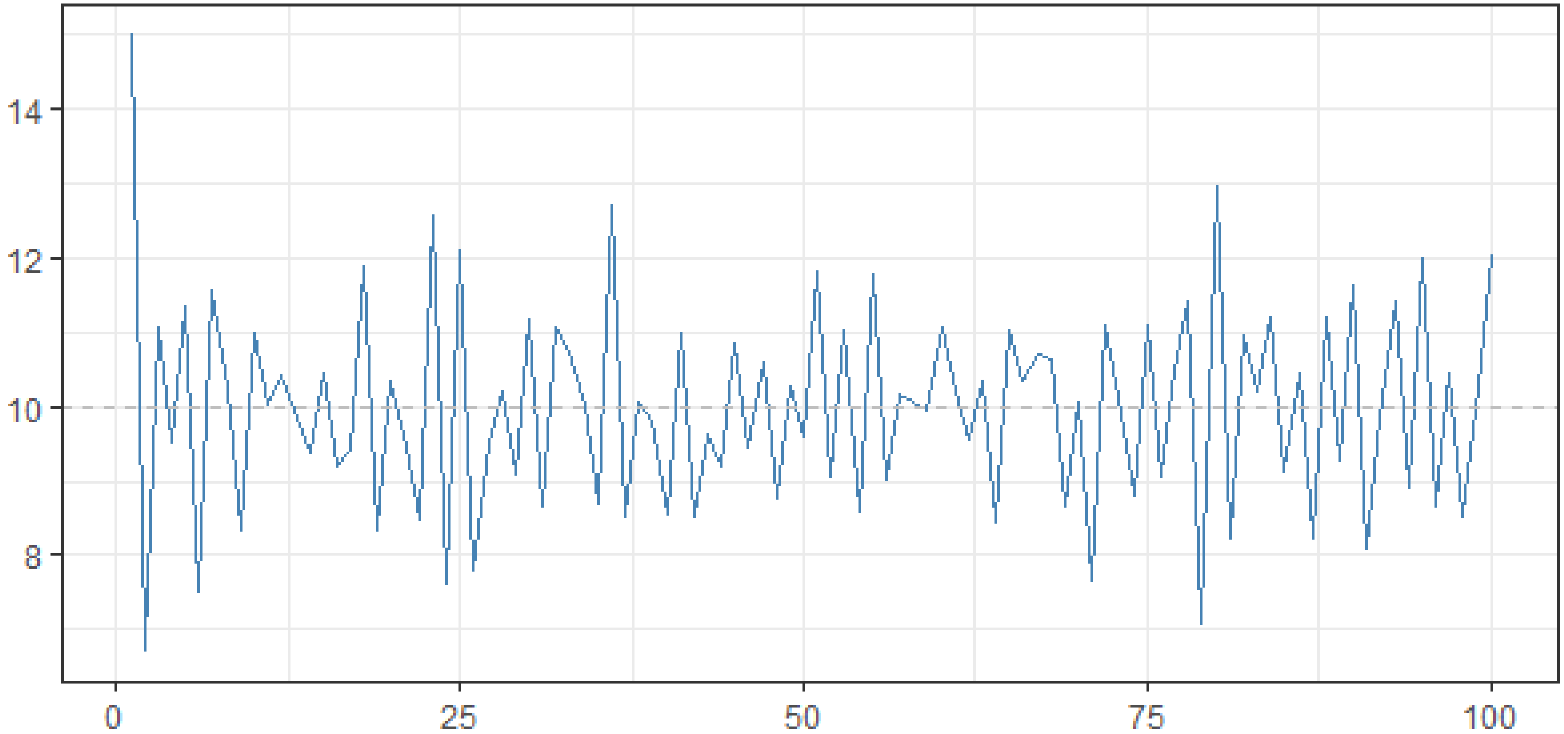
- **정상 시계열 :** $Z_t = 0.5Z_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0,1)$ i.i.d

  - 추세 없음(시계열의 평균이 시간 축에 평행)
  - 시계열의 진폭(변동)이 시간의 흐름에 따라 일정



정상 확률 과정

- **정상 시계열** : $Z_t = 17 - 0.7Z_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0,1)$ i.i.d
  - 추세 없음(시계열의 평균이 시간 축에 평행)
  - 시계열의 진폭(변동)이 시간의 흐름에 따라 일정



백색잡음 (White Noise)

- **IID 확률 변수**
  - $Z_t \sim_{iid} (\mu, \sigma^2) \Rightarrow \{Z_t\}$: 정상 시계열
- **백색잡음과정 (white noise process)**
  - 표기 : $\{Z_t\} \sim WN(0, \sigma^2)$
  - 정의 : 다음의 3가지 조건을 만족하는 확률 과정

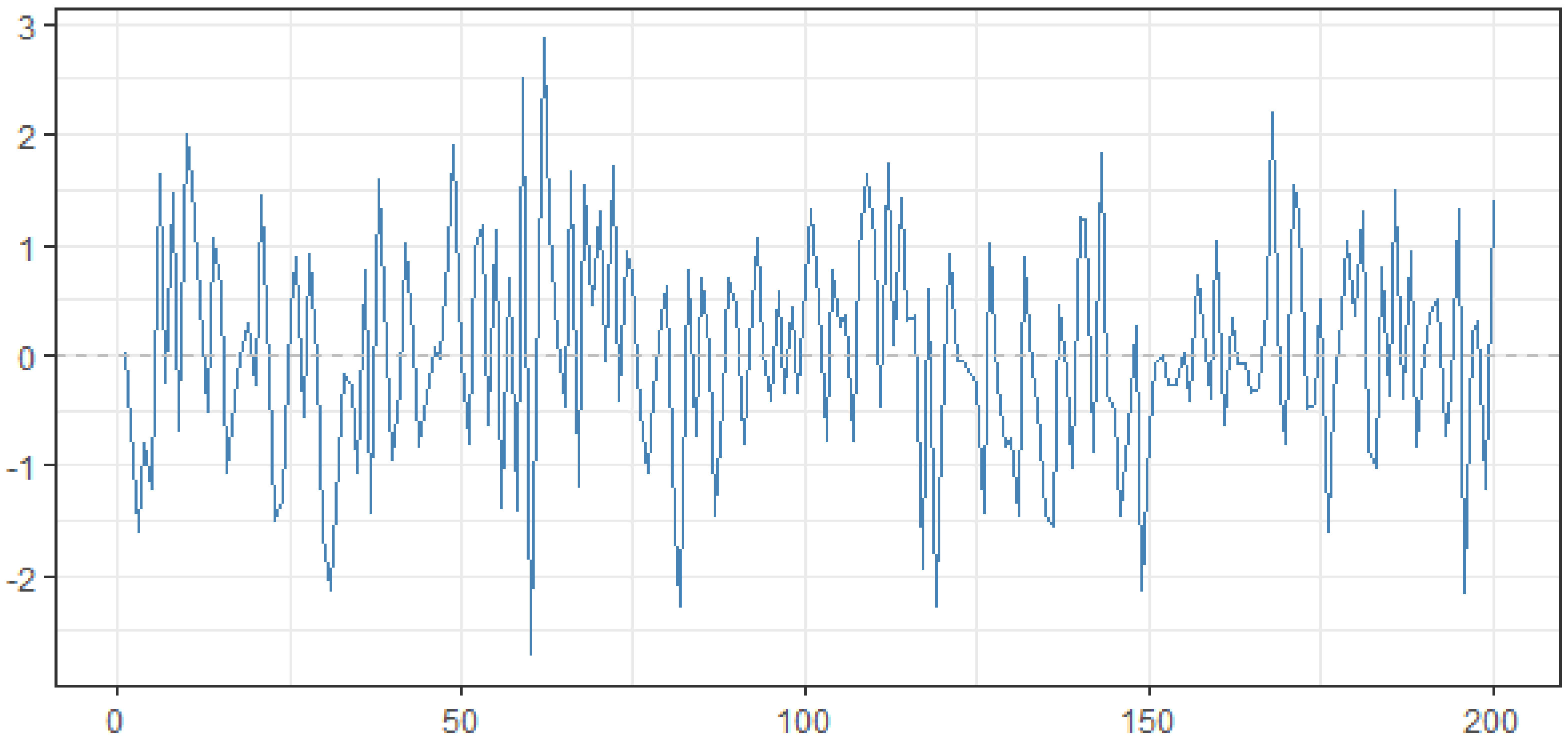
    (1) $E(Z_t) = 0, \ \forall t$

    (2) $Var(Z_t) = \sigma^2, \ \forall t$

    (3) $Cov(Z_t, Z_{t+h}) = 0, \ \forall t, h \neq 0$
  - $X_t \sim_{iid} (0, \sigma^2) \Rightarrow \{X_t\} \sim WN(0, \sigma^2)$
  - 백색잡음 $\{Z_t\}$ 는 stationary process

백색잡음 (White Noise)

- **백색잡음 N(0,1) iid**



절편이 없는 확률 보행 과정

- **확률보행과정** (random walk process)

  - $\{Z_t\}$ ~ randoem walk

  $$Z_t = Z_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN(0, \sigma^2), Z_0 = 0$$

  - $Z_t = Z_{t-1} + \varepsilon_t = Z_{t-2} + \varepsilon_{t-1} + \varepsilon_t = Z_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \cdots = \sum_{i=1}^{t} \varepsilon_i$

    (1) 평균 : $E(Z_t) = 0, \ \forall t$

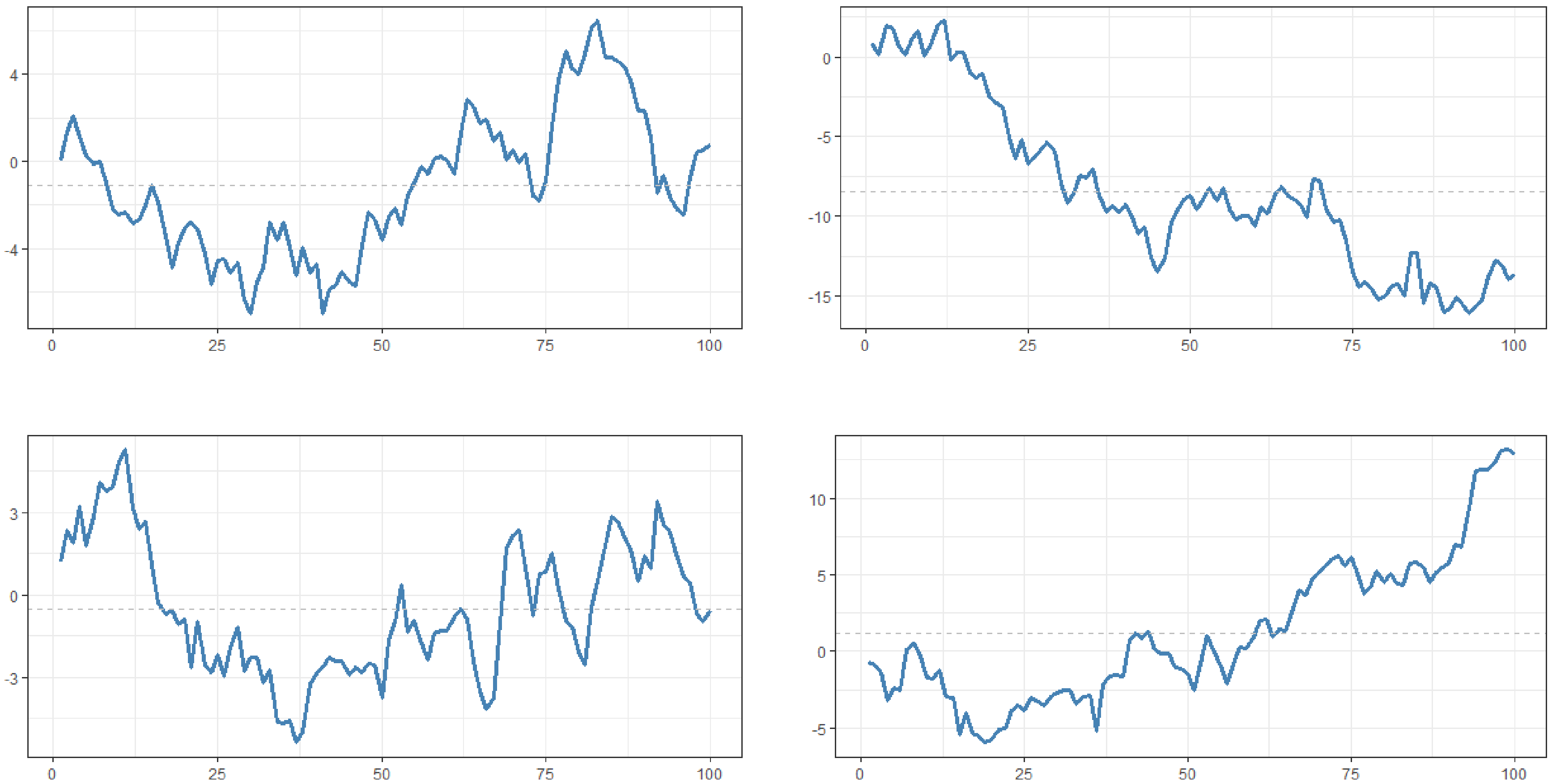
    (2) 분산 : $Var(Z_t) = t\sigma^2,$

    (3) 자기공분산 : $Cov(Z_t, Z_{t+h}) = t\sigma^2$

  - 확률보행과정 $\{Z_t\}$ 는 정상확률 과정이다 (O / X)

절편이 없는 확률 보행 과정

- **확률보행과정** (random walk process)



절편이 있는 확률 보행 과정

- **절편이 있는 확률보행과정** (random walk process with drift)

  - $\{Z_t\}$ ~ randoem walk with drift $\delta(\neq 0)$

$$Z_t = \delta + Z_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN(0, \sigma^2), Z_0 = 0$$

  - $Z_t = \delta + Z_{t-1} + \varepsilon_t = \delta + \delta + Z_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \cdots = t\delta + \sum_{i=1}^{t} \varepsilon_i$

    (1) 평균 : $E(Z_t) = t\delta$

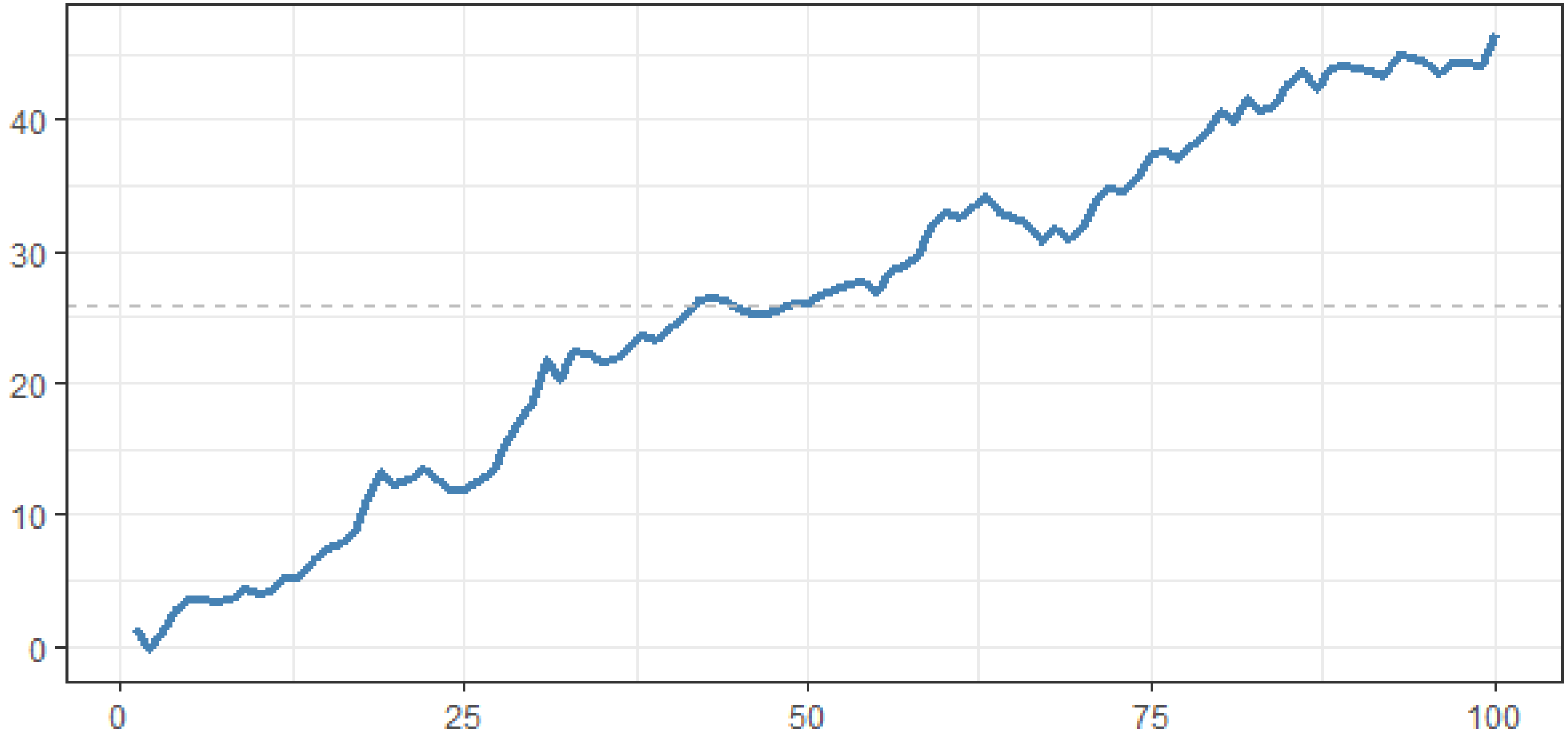
    (2) 분산 : $Var(Z_t) = t\sigma^2$

    (3) 자기공분산 : $Cov(Z_t, Z_{t+h}) = t\sigma^2$

  - 절편이 있는 확률보행과정 $\{Z_t\}$ 는 대표적인 비정상 (non-stationary) 확률 과정

절편이 있는 확률 보행 과정

- **절편이 있는 확률보행과정** (random walk process with drift)



이동평균 과정(moving average process)

- **1차 - 이동평균 과정 (MA(1))**

$$Z_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1}, \qquad \varepsilon_t \sim WN(0, \sigma^2)$$

  - 평균 : $E(Z_t) = \mu$
  - 분산 : $Var(Z_t) = (1+\theta^2)\sigma^2$
  - 자기공분산 : $Cov(Z_t, Z_{t+h}) = \begin{cases} -\theta\sigma^2, & h = 1 \\ 0, & h > 1 \end{cases}$
  - $\{Z_t\}$: stationary process

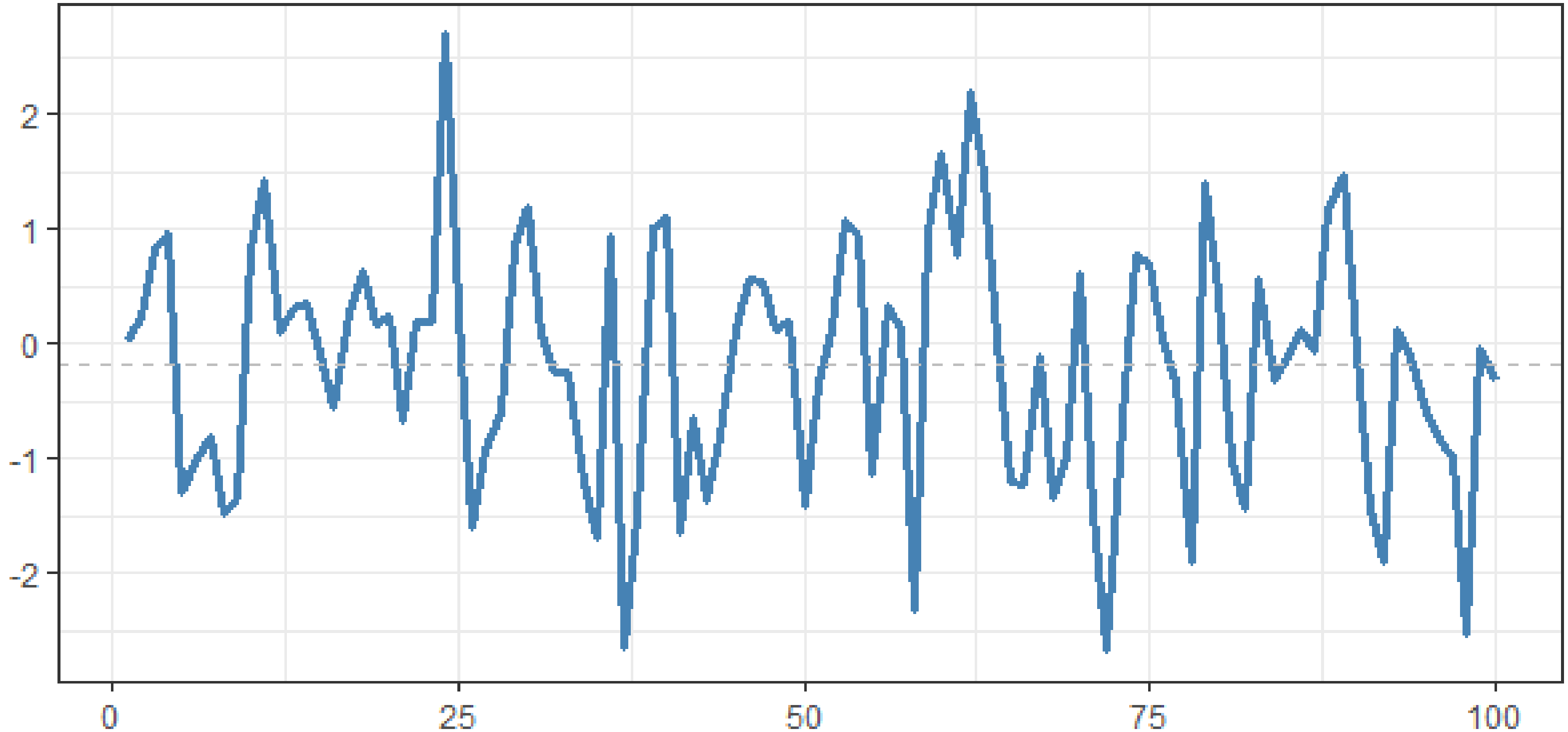
- **q차 - 이동평균 과정 (MA(q))**

$$Z_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}, \qquad \varepsilon_t \sim WN(0, \sigma^2)$$

  - $\{Z_t\}$: stationary process
  - 평균, 분산, 자기공분산 = ?

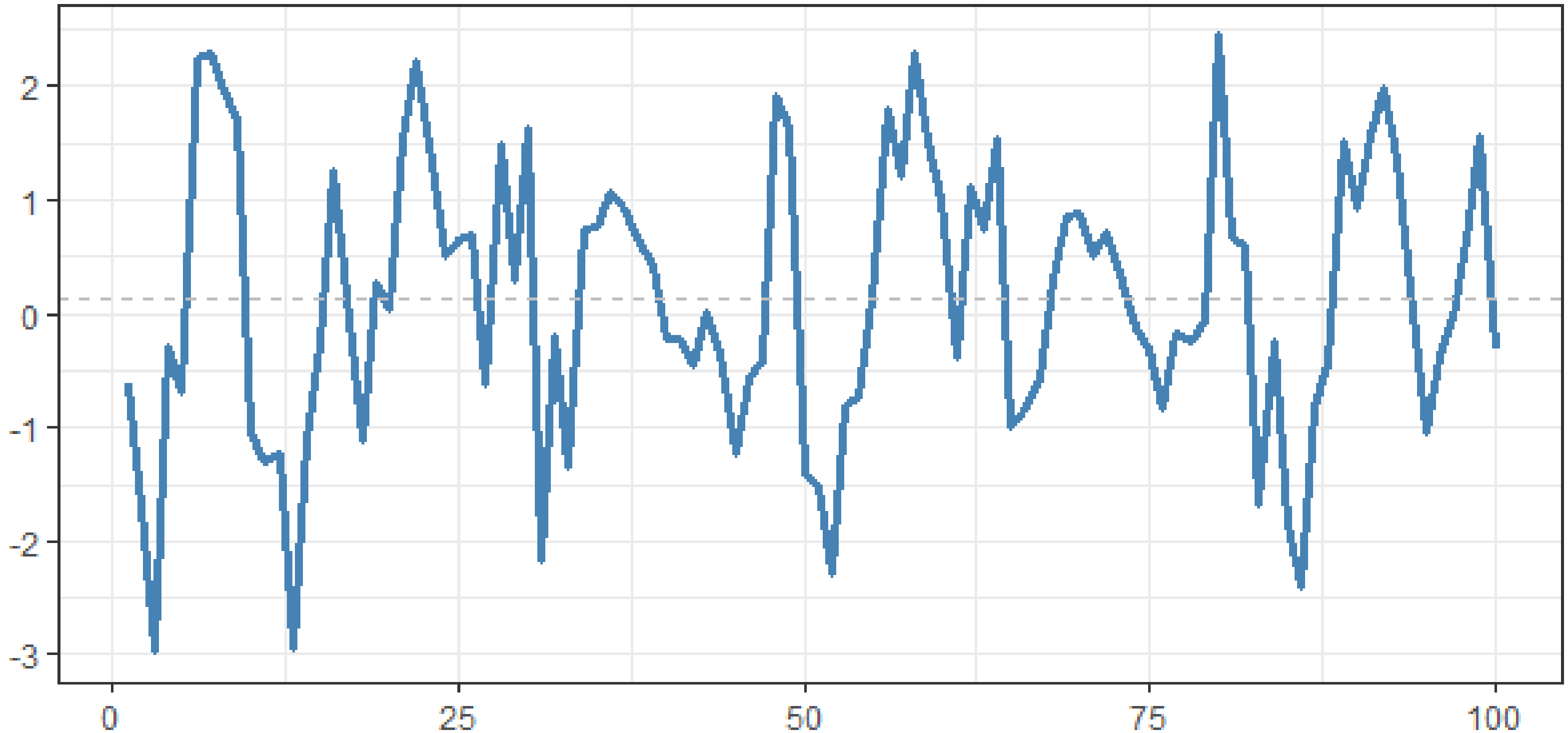
이동평균 과정(moving average process)

- **MA(1) :** $Z_t = \mu + \varepsilon_t - 0.5\varepsilon_{t-1}, \varepsilon_t \sim WN(0,1)$



이동평균 과정(moving average process)

- **MA(2) :** $Z_t = \mu + \varepsilon_t - 0.5\varepsilon_{t-1} - 0.3\varepsilon_{t-2}, \varepsilon_t \sim_{iid} N(0,1)$



선형 과정(linear process)

- **선형과정**

$$Z_t = \mu + \varepsilon_t \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \cdots = \mu + \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j}\,, \qquad \varepsilon_t \sim \mathrm{WN}(0, \sigma^2), \varphi_0 = 1$$

- 평균 : $E(Z_t) = \mu$
- 분산 : $Var(Z_t) = \sigma^2 \sum_{j=0}^{\infty} \varphi_j^2$
- 자기공분산 : $Cov(Z_t, Z_{t+h}) = \sigma^2 \sum_{j=0}^{\infty} \varphi_j \varphi_{j+|h|}$
- 정상성 조건 :

$$\sum_{j=0}^{\infty} \varphi_j^2 < \infty \Rightarrow \{Z_t\}$$: stationary process

- $\mathrm{MA}(\infty)$ process, 무한이동평균과정(infinite moving average process)

후진 연산자

- **후진연산자 (Backshift Operator)**

  - $BZ_t = Z_{t-1},\ B^j Z_t = Z_{t-j}$
  - 선형 과정

$$Z_t = \mu + \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j} = \mu + \sum_{j=0}^{\infty} \varphi_j B^j \varepsilon_t = \mu + \Psi(B)\varepsilon_t$$

  - MA(1)

$$Z_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1} = \mu + \varepsilon_t - \theta B\varepsilon_t = \mu + (1 - \theta B)\varepsilon_t$$

  - MA(q) : $\theta_0 = 1$

$$Z_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}$$

$$= \mu + \left(1 - \theta_1 B - \cdots - \theta_q B^q\right)\varepsilon_t = \mu + \sum_{j=0}^{q} \theta_j B^j \varepsilon_t$$

자기회귀 과정(autoregressive process)

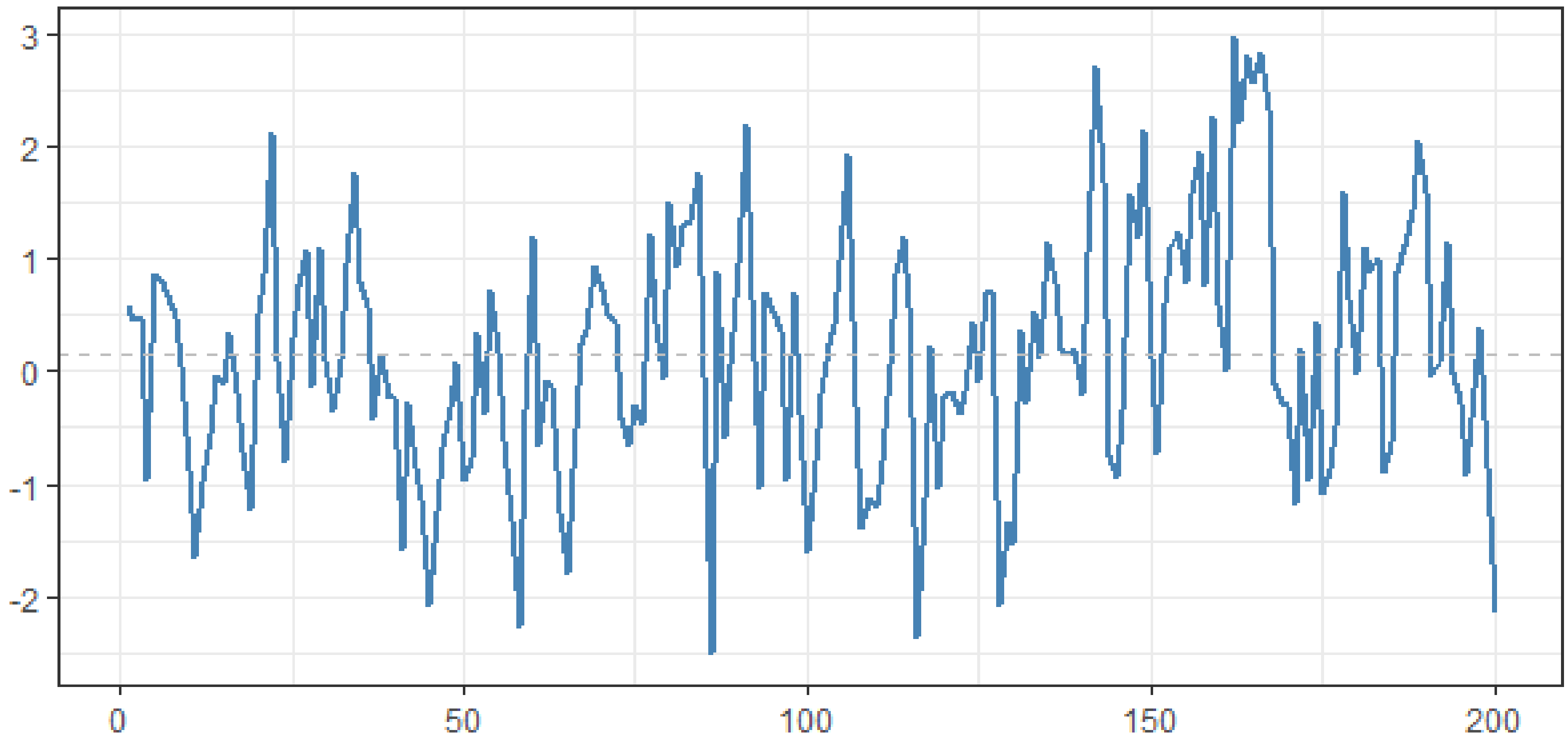
- **1차 – 자기회귀 과정 (AR(1) process)**

$$Z_t - \mu = \varphi\,(Z_{t-1} - \mu) + \varepsilon_t, \qquad \varepsilon_t \sim_{iid} (0, \sigma^2)$$

- $|\varphi| < 1$ 인 경우
  - $Z_t = \mu + \sum_{j=0}^{\infty} \varphi^j \varepsilon_{t-j}$: stationary linear process
  - 평균 : $E(Z_t) = \mu$
  - 분산 : $Var(Z_t) = \sigma^2 \sum_{j=0}^{\infty} \varphi^{2j} = \frac{\sigma^2}{1-\varphi^2}$
  - 자기공분산 : $Cov(Z_t, Z_{t+h}) = \varphi^h \frac{\sigma^2}{1-\varphi^2}$
- $|\varphi| = 1$ 인 경우 : random walk ⇒ non-stationary
- 정상성 조건 : $|\varphi| < 1 \Rightarrow \{Z_t\}$: stationary process

자기회귀 과정(autoregressive process)

- **AR(1) :** $Z_t = 0.5\,Z_{t-1} + \varepsilon_t,\ \ \varepsilon_t \sim_{iid} N(0,1)$



자기공분산 함수

- **자기공분산함수 (autocovariance function, ACF)**

$$Cov(Z_t, Z_{t+h}) = E\{(Z_t - E(Z_t))(Z_{t+h} - E(Z_{t+h})\}$$

- **정상시계열의 ACF**
  - $\gamma_h = Cov(Z_t, Z_{t+h})$: t와 무관하고, h에만 의존
  - 성질
    - $\gamma_0 = Var(Z_t), \forall t$
    - $|\gamma_h| \leq \gamma_0, \ \forall h$
    - $\gamma_{-h} = \gamma_h, \ \forall h = 1,2, \dots$

- **표본 자기공분산 (sample autocovariance)**

$$\hat{\gamma}_h = \frac{1}{n}\sum_{t=1}^{n-h}(Z_t - \bar{Z})(Z_{t+h} - \bar{Z}), h = 0,1,2, \dots$$

자기상관함수

- **자기상관함수 (autocorrelation function, ACF)**

$$\rho_h = Corr(Z_t, Z_{t+h})$$

- **정상시계열의 ACF**

  - $\rho_h = Corr(Z_t, Z_{t+h}) = \frac{Cov(Z_t, Z_{t+h})}{sd(Z_t)sd(Z_{t+h})} = \frac{\gamma_h}{\gamma_0}, \quad \forall t$
  - 성질
    - $\rho_0 = 1$
    - $|\rho_h| \leq 1, \ \forall h$
    - $\rho_{-h} = \rho_h, \ \forall h = 1,2, \ldots$

- **표본 자기상관함수 (sample autocorrelation function : SACF)**

$$\hat{\rho}_h = \hat{\gamma}_h / \hat{\gamma}_0$$

자기상관함수

- **표본 자기상관함수 (sample autocorrelation function : SACF)**

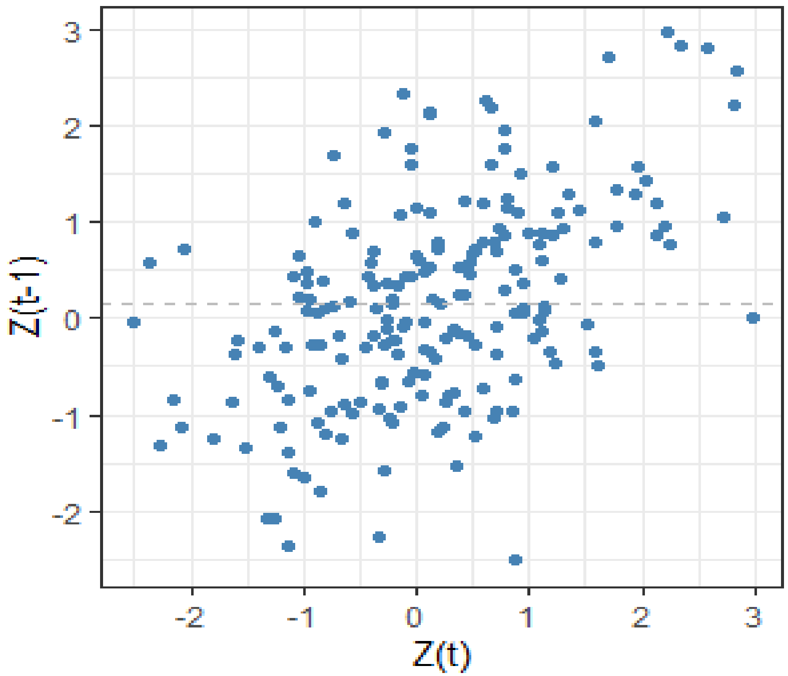
$$\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-h}(Z_t - \bar{Z})(Z_{t+h} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2}, \qquad h = 1,2, \ldots$$

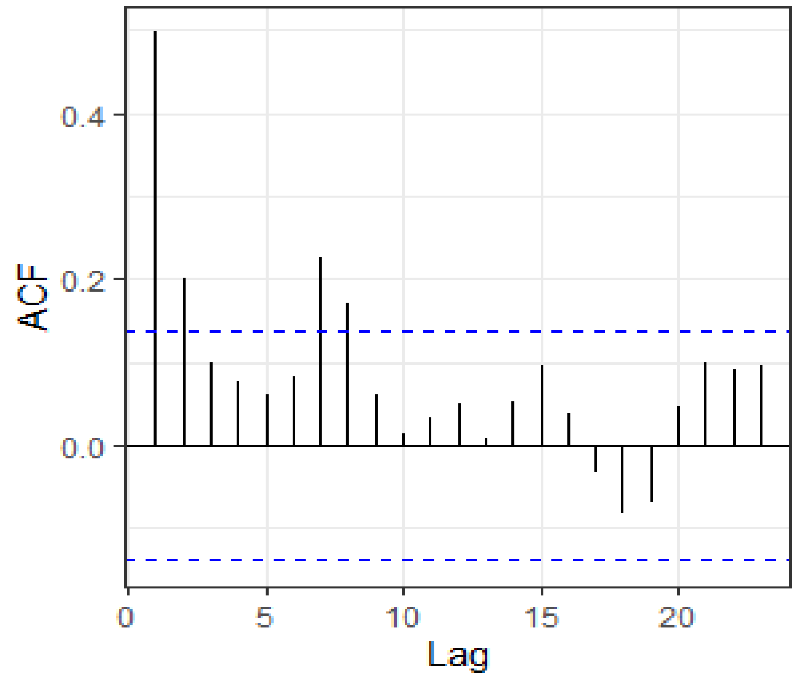
- 성질
  - $E(\hat{\rho}_h) \cong \rho_h$
  - MA(q) process (Bartlett, 1946) : $\hat{\rho}_h \approx N\left(\rho, \frac{1}{n}(1 + 2\sum_{i=1}^{q} \rho_i^2)\right)$
  - White noise : $\hat{\rho}_h \approx N\left(0, \frac{1}{n}\right)$

  : $\Rightarrow H_0: \rho_h = 0 \; vs. H_1: \rho_h \neq 0$ 에 대한 검정에서 $|\hat{\rho}_h| > \frac{2}{\sqrt{n}}$ 이면 H_0 기각되고, 시차 h에서 자기상관관계가 존재한다고 할 수 있음

자기상관함수

- **표본 상관 도표 :**
  - 표본자기상관계수를 그린 그림
  - 시계열 모형 식별에 사용
- 예) $Z_t = 0.5\, Z_{t-1} + \varepsilon_t,\ \ \varepsilon_t \sim_{iid} N(0,1)$





부분자기상관함수

- **부분자기상관함수 (partial autocorrelation function : PACF)**
  - $\varphi_{kk} = Corr(Z_t^*, Z_{t+k}^*),\ k = 1,2,3, \ldots$
    - $Z_t^* = Z_t - E(Z_t | Z_{t+1}, \ldots, Z_{t+k-1})$

      : Z_t를 $Z_{t+1}, \ldots, Z_{t+k-1}$에 회귀분석을 시행한 후의 잔차
    - $Z_{t+k}^* = Z_{t+k} - E(Z_{t+k} | Z_{t+1}, \ldots, Z_{t+k-1})$

      : Z_{t+k}를 $Z_{t+1}, \ldots, Z_{t+k-1}$에 회귀분석을 시행한 후의 잔차
  - 의미 : Z_t와 Z_{t+k}에서 $Z_{t+1}, \ldots, Z_{t+k-1}$의 효과를 제거한 후의 상관계수
- **표본 부분자기상관함수 (SPACF)**
  - ARMA(p.q) 모형 적합시 모형의 차수 결정하는 데 이용

부분자기상관함수

- **부분자기상관함수 구하는 방법 (1/2)**
  - 종속변수 Z_{t+k}를 $Z_{t+k-1}, Z_{t+k-2},, \ldots, Z_t$ 로 회귀분석

$$Z_{t+k} = \phi_{k1} Z_{t+k-1} + \phi_{k2} Z_{t+k-2} + \cdots + \phi_{kk} Z_t$$

  - 양변에 $Z_{t+k-j}(j \geq 1)$ 을 곱한 후 기대값을 취하면

$$\gamma_j = \phi_{k1}\gamma_{j-1} + \phi_{k2}\gamma_{j-2} + \cdots + \phi_{kk}\gamma_{j-k}$$

  - 양변을 γ_0로 나누면

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{kk}\rho_{j-k}$$

부분자기상관함수

▪ 부분자기상관함수 구하는 방법 (2/2)

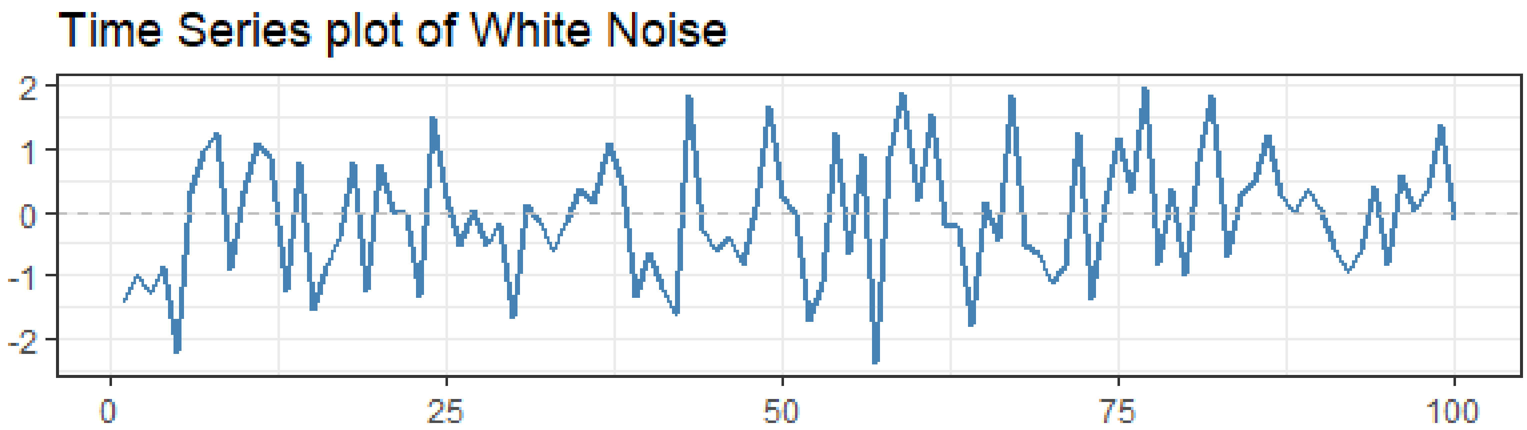
- $j = 1, \dots, k$ 에 대해 다음의 연립 방정식 생성

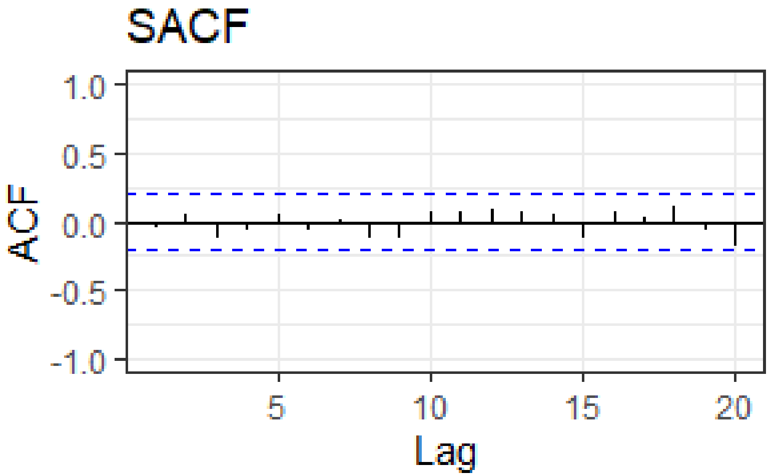
$$\rho_1 = \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1}$$
$$\rho_2 = \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \cdots + \phi_{kk}\rho_{k-2}$$
$$\vdots$$
$$\rho_k = \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}\rho_0$$

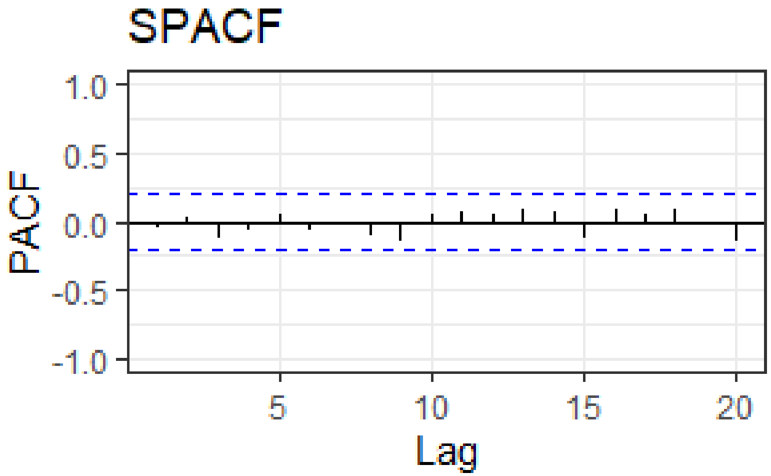
- 연립방정식의 해를 구함
  - $\phi_{11} = \rho_1$
  - $\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$
  - $\vdots$

백색잡음의 SACF와 SPACF

- **표본 상관 도표 : WN process**

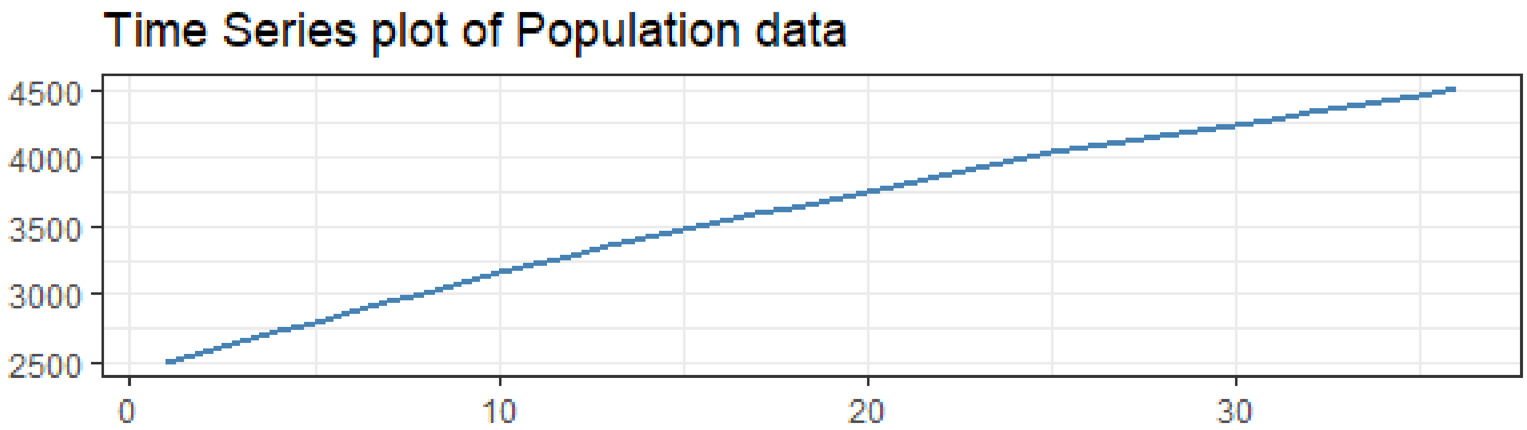


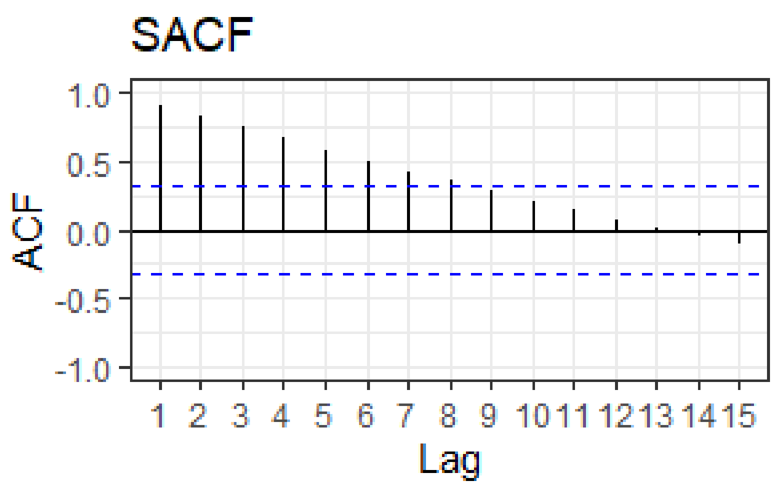


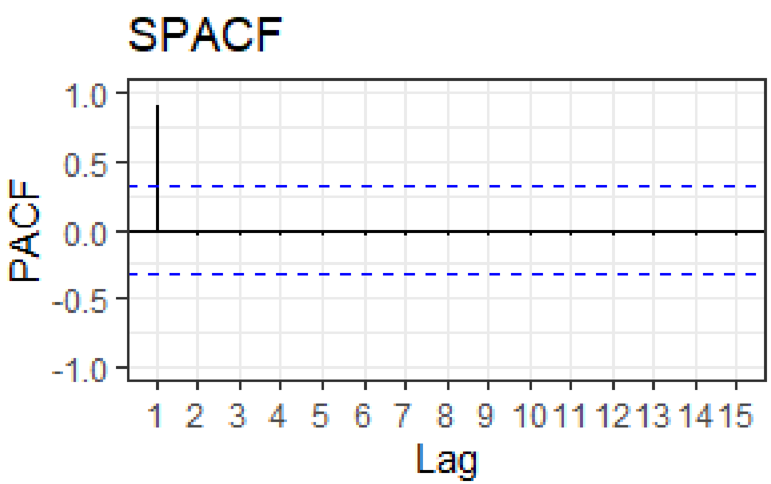


SACF와 SPACF

- **표본 상관 도표 : Population data**

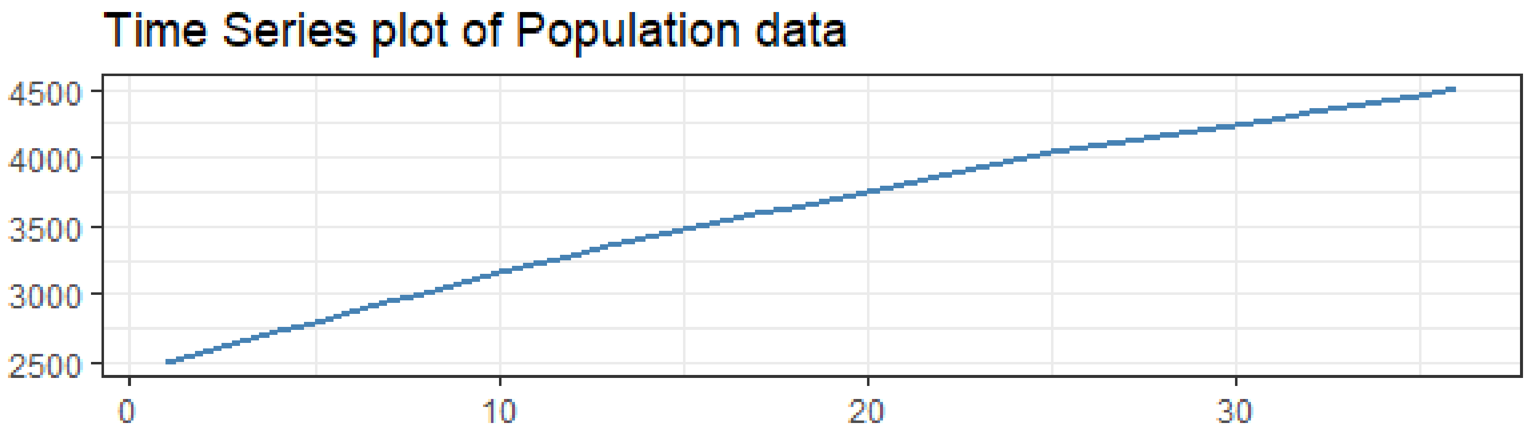


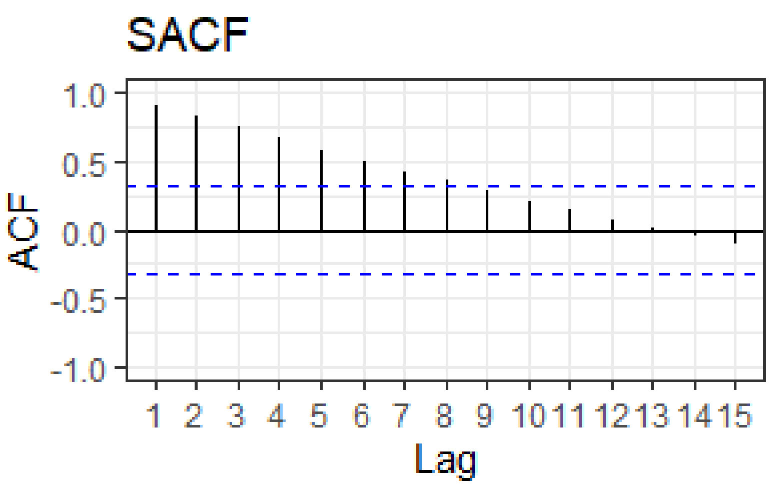


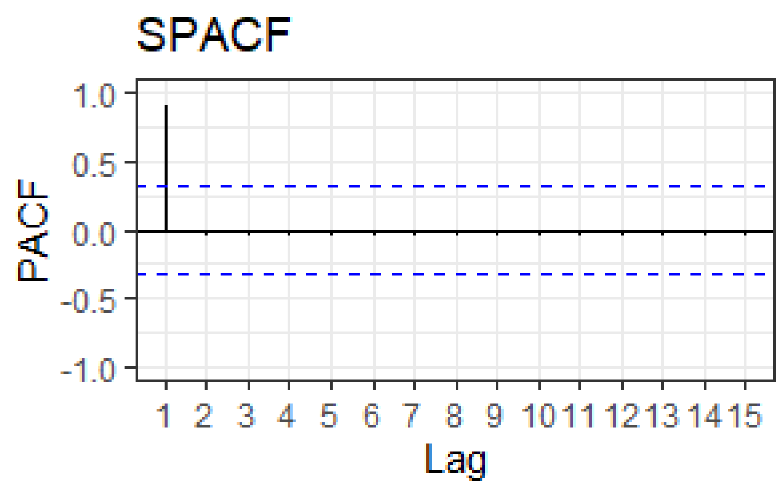


SACF와 SPACF

- **표본 상관 도표 : Population data**

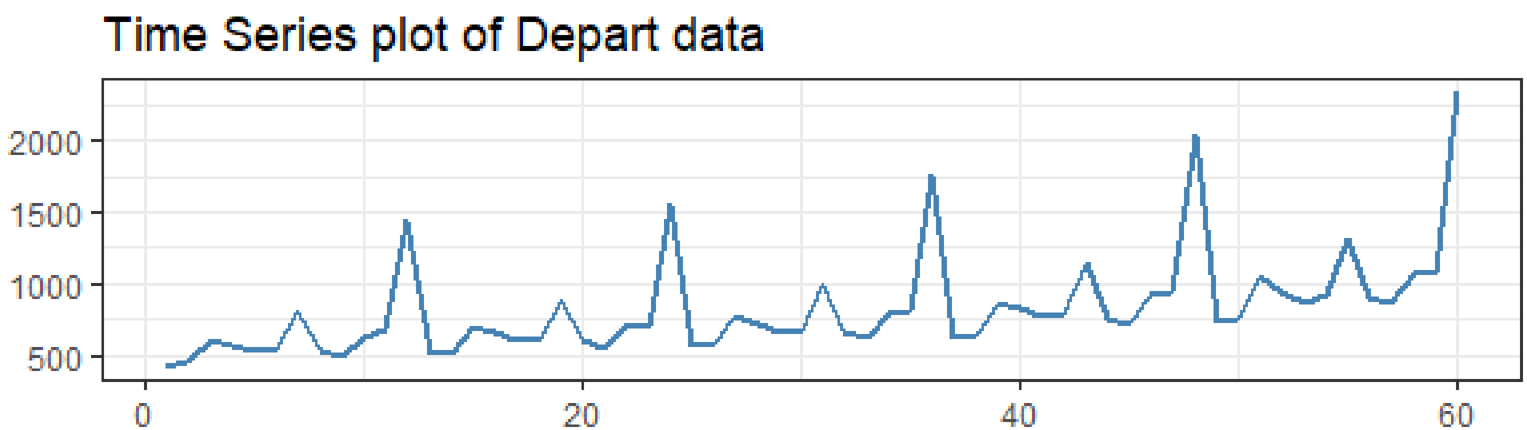


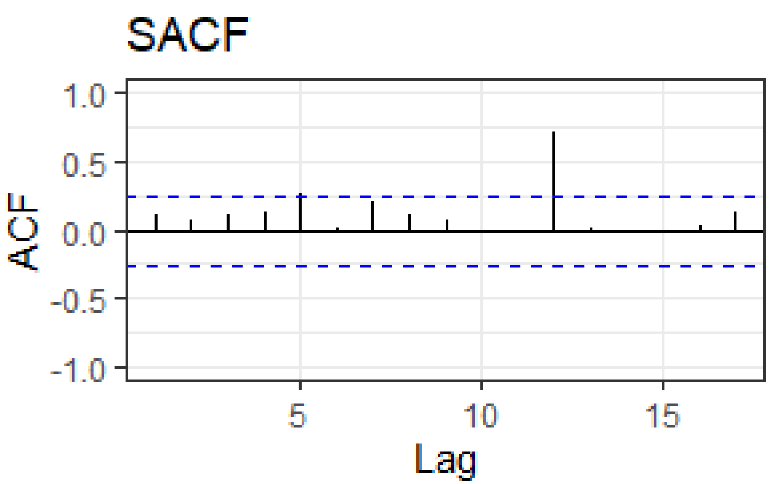


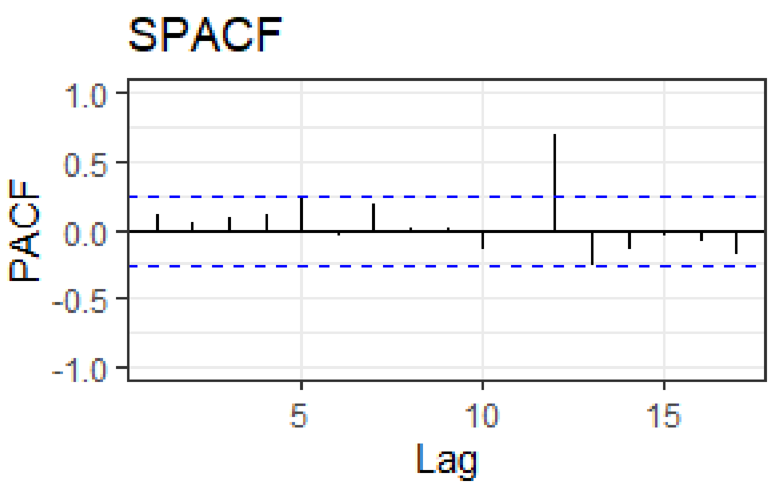


SACF와 SPACF

- **표본 상관 도표 : 백화점 매출**

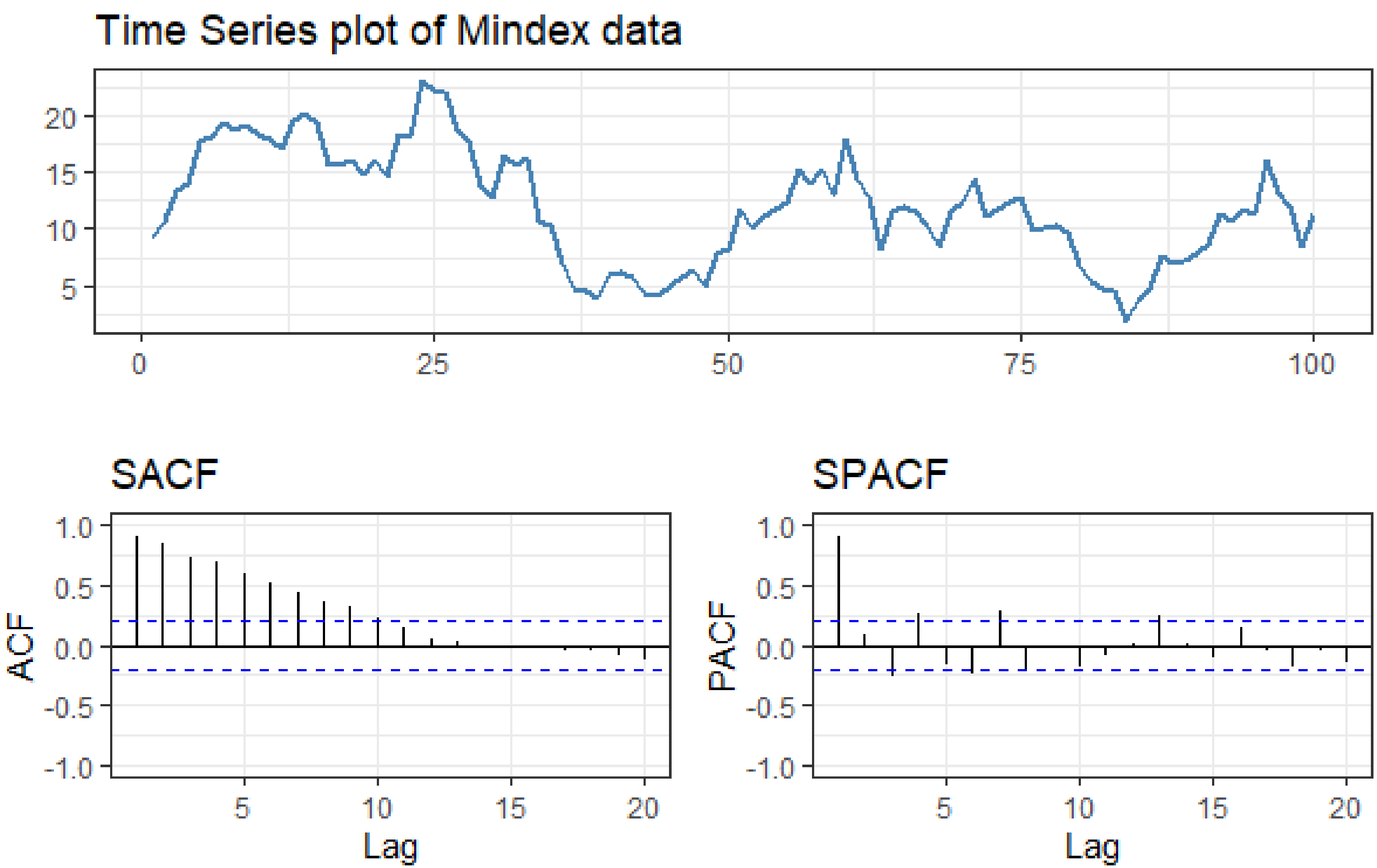






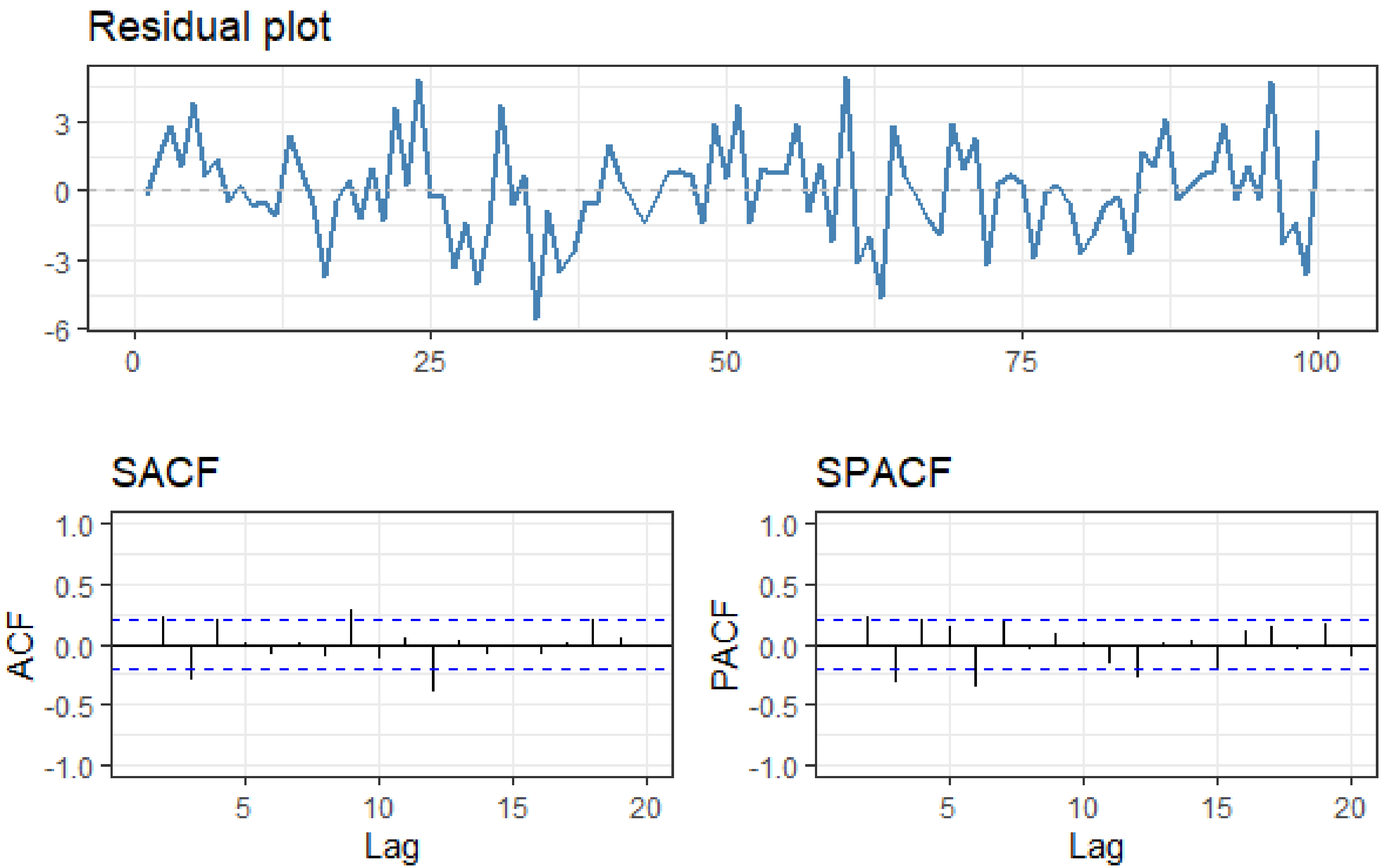
SACF와 SPACF

- **표본 상관 도표 : 중간재 출하지수**



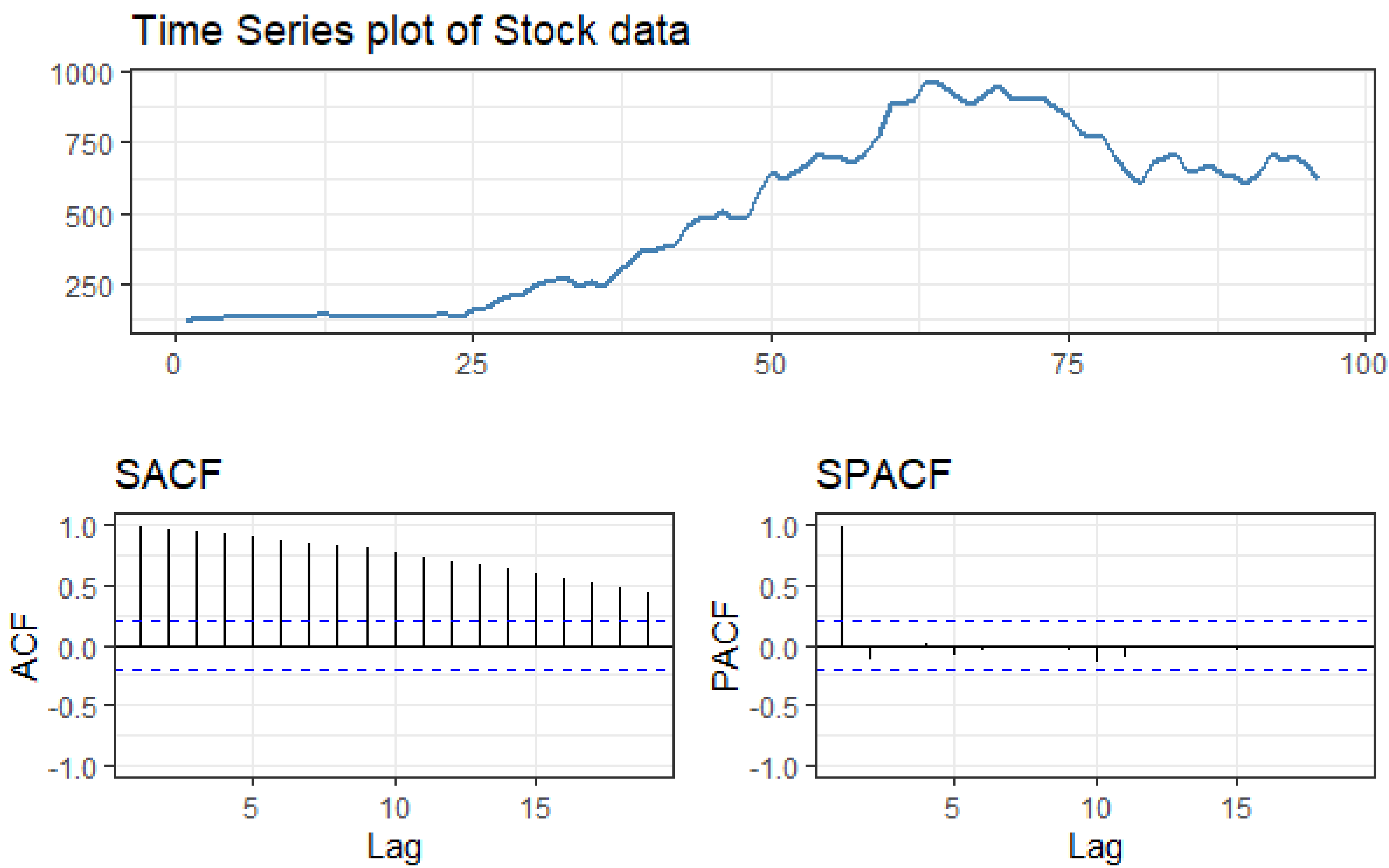
SACF와 SPACF

- **표본 상관 도표 : 중간재 출하지수의 지수평활 후 잔차**



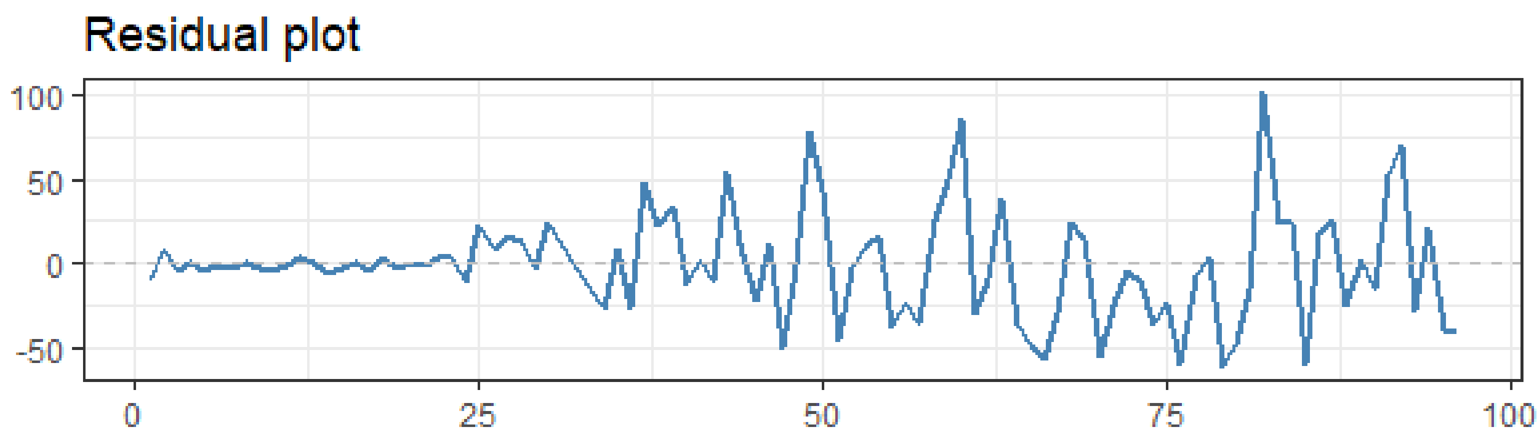
SACF와 SPACF

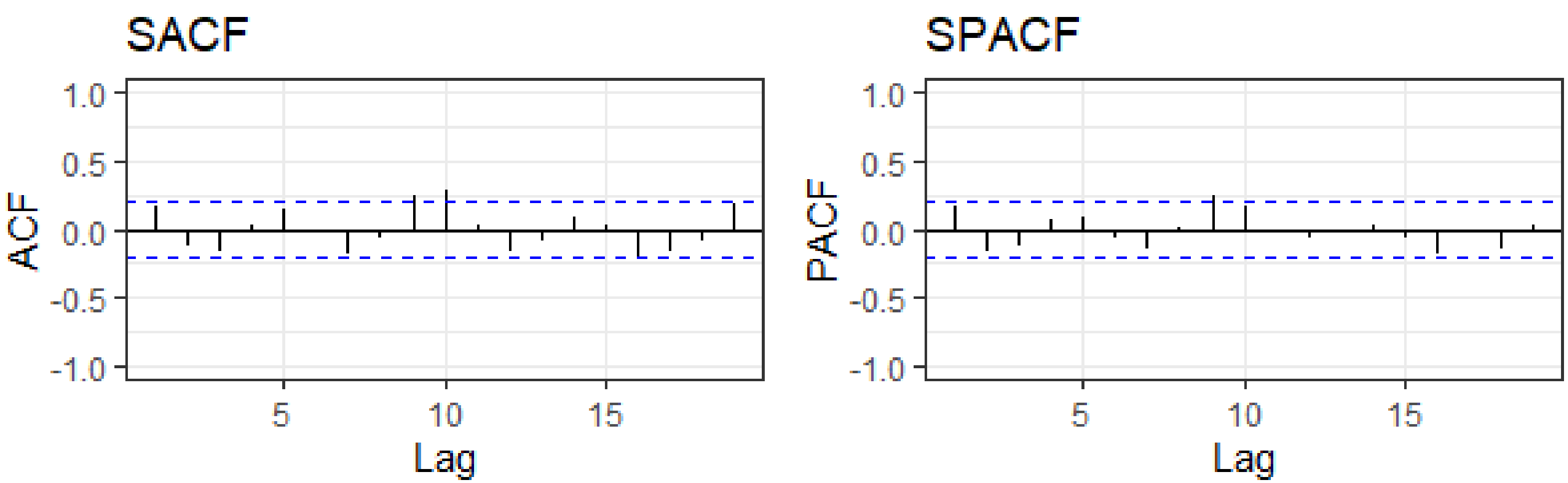
- 표본 상관 도표 : 주식가격 데이터



SACF와 SPACF

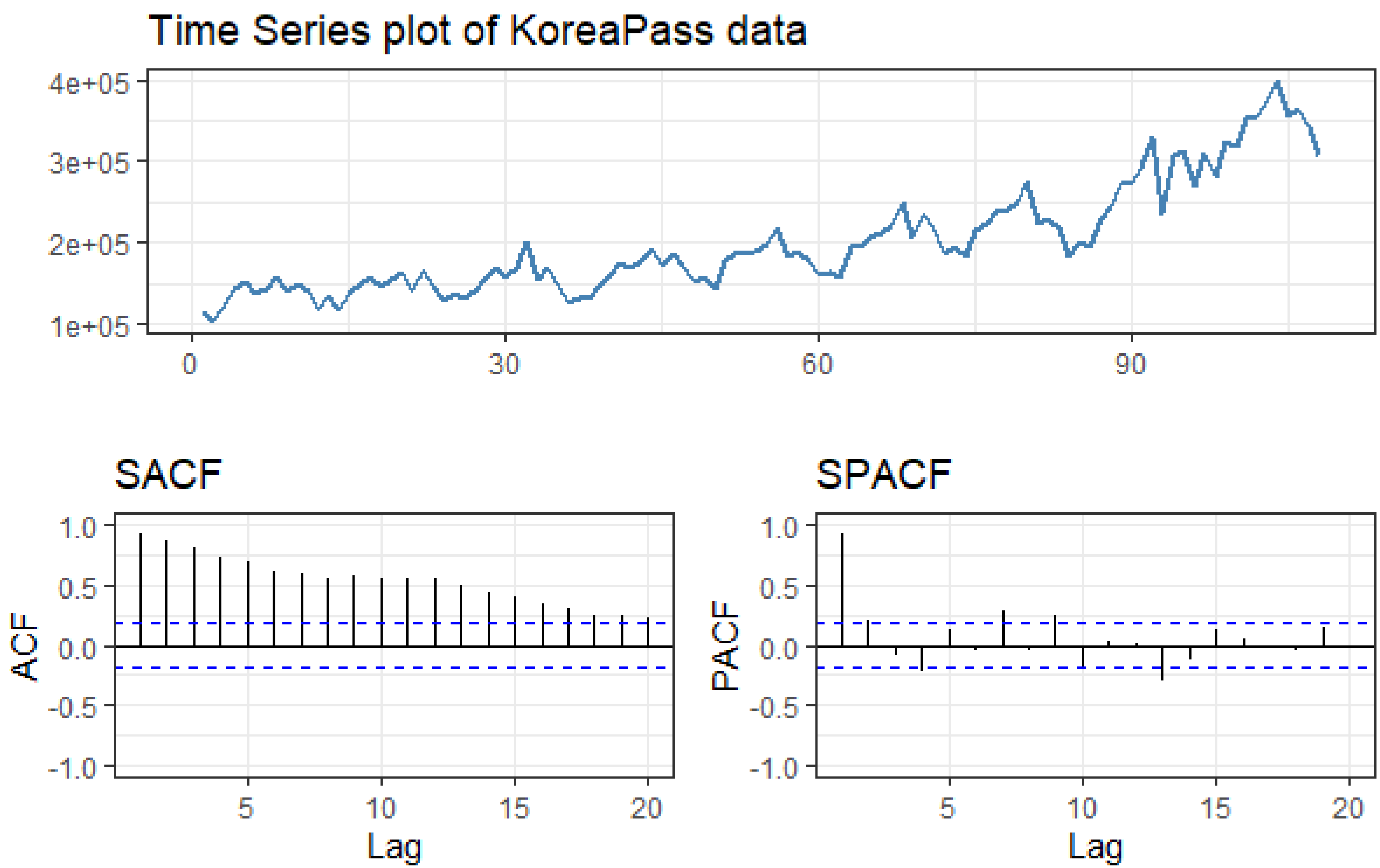
- **표본 상관 도표 : 주식가격의 이중지수평활 후 잔차**





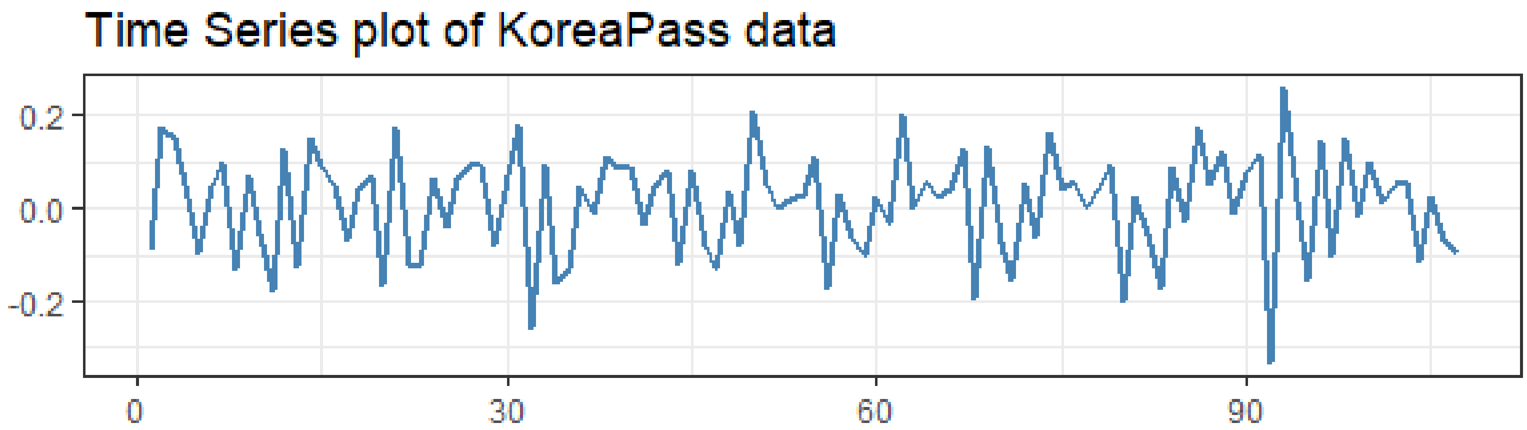
SACF와 SPACF

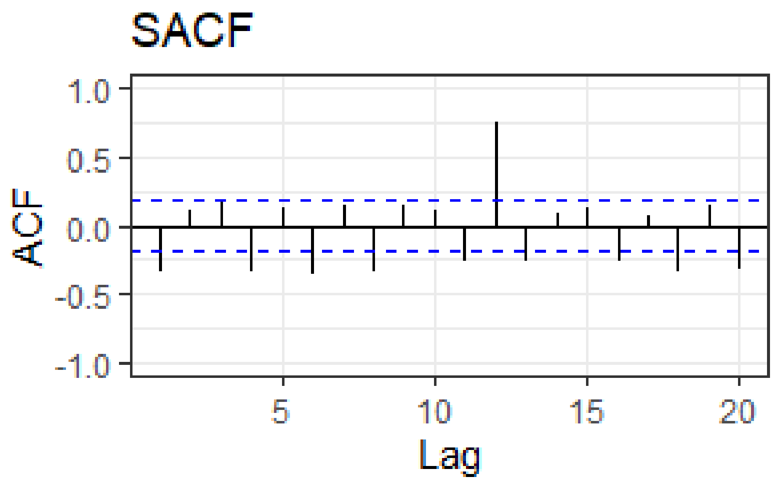
- **표본 상관 도표 : 비행기 승객수**

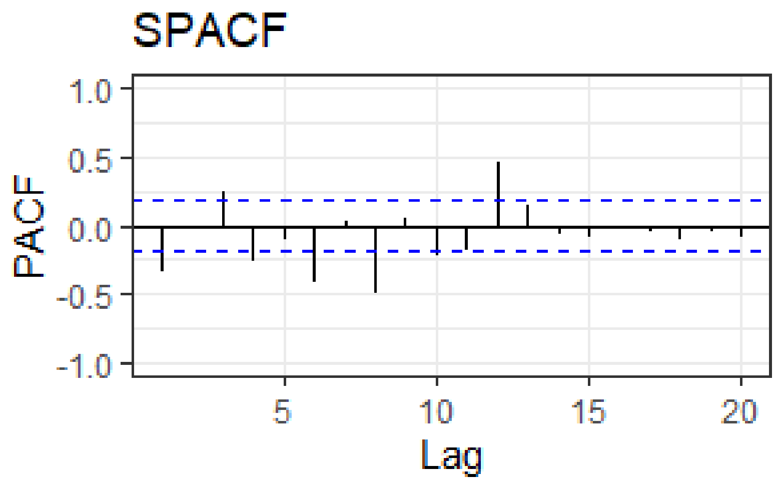


SACF와 SPACF

- **표본 상관 도표 : 비행기 승객수 –** 로그변환 후 차분 (이분산성, 추세 제거)







End of Document