Forecasting Methods

예측방법론
Forecasting Methods

04 강

시계열의 자기상관

통계·데이터과학과 **이긍희** 교수

한국방송통신대학교

Forecasting Methods

목차

CONTENTS

chapter 01

자기상관

1 시계열의 자기상관

- 시계열의 시간에 따른 의존구조 : 자기공분산 또는 자기상관으로 파악

- 자기공분산

$$\gamma(s, t) = Cov(Y_s, Y_t)$$

- 자기상관함수 :

$$\rho(s, t) = \frac{Cov(Y_s, Y_t)}{\sqrt{Var(Y_s)Var(Y_t)}} = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}$$

2 자기공분산함수와 자기상관함수의 성질

성질	자기공분산함수	자기상관함수
① 대칭성	$\gamma(s,t)=\gamma(t,s)$	$\rho(s,t)=\rho(t,s)$
② $s=t$	$\gamma(t,t)=Var(Y_t)$	$\rho(t,t)=1$
③ 자기상관 없음	$\gamma(s,t)=0$	$\rho(s,t)=0$
④ 양의 자기상관	$\gamma(s,t)>0$	$0<\rho(s,t)\le 1$
⑤ 음의 자기상관	$\gamma(s,t)<0$	$-1\le\rho(s,t)<0$

3 여러가지 시계열의 자기상관

- 여러가지 시계열

» 백색잡음계열 : $Y_t = \varepsilon_t$

» 이동평균계열 : $Y_t = \theta_1 \varepsilon_{t-1} + \varepsilon_t$

» 확률보행계열 : $Y_t = \phi_0 + Y_{t-1} + \varepsilon_t$

3 여러가지 시계열의 자기상관

- **백색잡음계열의 기댓값, 분산, 자기공분산, 자기상관함수**

$$\mu_t = E(\varepsilon_t) = 0$$

$$\sigma_t^2 = Var(\varepsilon_t) = \sigma_w^2$$

$$\gamma(s,t) = \begin{cases} \sigma_w^2, & s = t \\ 0, & s \neq t \end{cases}$$

$$\rho(s,t) = \begin{cases} 1, & s = t \\ 0, & s \neq t \end{cases}$$

3 여러가지 시계열의 자기상관

- 이동평균계열의 기댓값과 분산

$$E(Y_t) = E(\theta_1 \varepsilon_{t-1} + \varepsilon_t)$$

$$= \theta_1 E(\varepsilon_{t-1}) + E(\varepsilon_t) = 0$$

$$Var(Y_t) = Var(\theta_1 \varepsilon_{t-1} + \varepsilon_t) = \theta_1^2 Var(\varepsilon_{t-1}) + Var(\varepsilon_t)$$

$$= (\theta_1^2 + 1)\sigma_w^2$$

3 여러가지 시계열의 자기상관

- **이동평균계열의 자기공분산**

$$\gamma(s,t) = Cov(\theta_1\varepsilon_{s-1} + \varepsilon_s,\ \theta_1\varepsilon_{t-1} + \varepsilon_t)$$

$$= E(\theta_1^2\varepsilon_{t-1}\varepsilon_{s-1} + \theta_1\varepsilon_{t-1}\varepsilon_s + \theta_1\varepsilon_t\varepsilon_{s-1} + \varepsilon_t\varepsilon_s)$$

$$= \theta_1 E(\varepsilon_{t-1}\varepsilon_s) + \theta_1 E(\varepsilon_t\varepsilon_{s-1})$$

$$= \begin{cases} \theta_1\sigma_w^2, & s = t \pm 1 \\ 0, & s = \text{그 밖에} \end{cases}$$

3 여러가지 시계열의 자기상관

- 이동평균계열의 자기상관함수

$$\rho(t,s) = \frac{\gamma(t,s)}{\sqrt{\gamma(s,s)\,\gamma(t,t)}}$$

$$= \begin{cases} \dfrac{\theta_1}{1+\theta_1^2}, & s = t \pm 1 \\ 0, & s = \text{그 밖에} \end{cases}$$

3 여러가지 시계열의 자기상관

- **확률보행계열의 표현**

$$Y_1 = \phi_0 + \varepsilon_1$$

$$Y_2 = \phi_0 + (\phi_0 + \varepsilon_1) + \varepsilon_2 = 2\phi_0 + \varepsilon_1 + \varepsilon_2$$

$$\vdots$$

$$Y_t = \phi_0 t + \sum_{i=1}^{t} \varepsilon_i$$

3 여러가지 시계열의 자기상관

- 확률보행계열의 기댓값과 분산

$$E(Y_t) = E\left(\phi_0 t + \sum_{i=1}^{t} \varepsilon_i\right)$$

$$= \phi_0 t + \sum_{i}^{t} E(\varepsilon_i) = \phi_0 t$$

$$Var(Y_t) = Var\left(\phi_0 t + \sum_{i}^{t} \varepsilon_i\right)$$

$$= \sum_{i=1}^{t} Var(\varepsilon_i) = \sigma_w^2 t$$

3 여러가지 시계열의 자기상관

- **확률보행계열의 공분산**

$$\gamma(h) = \gamma(t, t+h) = Cov(Y_t, Y_{t+h})$$

$$= Cov\left(\phi_0 t + \sum_{i=1}^{t} \varepsilon_i,\ \phi_0 (t+h) + \sum_{j=1}^{t+h} \varepsilon_j\right)$$

$$= \sum_{i=1}^{t+h} \sum_{j=1}^{t} Cov(\varepsilon_i, \varepsilon_j) = t\sigma_w^2$$

3 여러가지 시계열의 자기상관

- **확률보행계열의 자기상관함수**

$$\rho(h) = \frac{\gamma(h)}{\sqrt{Var(Y_t)\,Var(Y_{t+h})}}$$

$$= \frac{t}{\sqrt{t(t+h)}} = \frac{\sqrt{t}}{\sqrt{t+h}}$$

chapter 02

자기상관계수와 부분자기상관계수

Forecasting Methods

1 상관계수

- 표본상관계수 r : 두 변수가 얼마나 선형적으로 밀접한지 파악할 수 있는 통계량

» **표본상관계수**

$$r = \frac{\sum_t (X_t - \overline{X})(Y_t - \overline{Y})}{\sqrt{\sum_t (X_t - \overline{X})^2 \sum_t (Y_t - \overline{Y})^2}}$$

(단, $\overline{X}$, $\overline{Y}$는 X_t, Y_t의 표본평균)

2 표본자기상관계수

- **표본자기상관계수 $\hat{\rho}(h)$: 두 변수 중 한 변수를 시차변수로 두고 구한 표본상관계수**

» 이론적 자기상관계수 $\rho(h)$의 추정량

$$\hat{\rho}(h) = \frac{\sum_t (Y_t - \overline{Y})(Y_{t-h} - \overline{Y})}{\sum_t (Y_t - \overline{Y})^2} = \frac{\sum_t (Y_t - \overline{Y})(Y_{t+h} - \overline{Y})}{\sum_t (Y_t - \overline{Y})^2}$$

2 표본자기상관계수

- 시계열 $Y_1, Y_2, \cdots, Y_n$ 서로 독립, 동일한 분포

  → $\hat{\rho}(h)$: 점근적으로 $N(0, \frac{1}{n})$

- 귀무가설 $H_0 : \rho(h) = 0$ 검정

  → $|\hat{\rho}(h)| > 1.96/\sqrt{n}$: 유의수준 5%에서 H_0 기각

  → Y_t가 시차 h에서 유의한 자기상관관계 있음

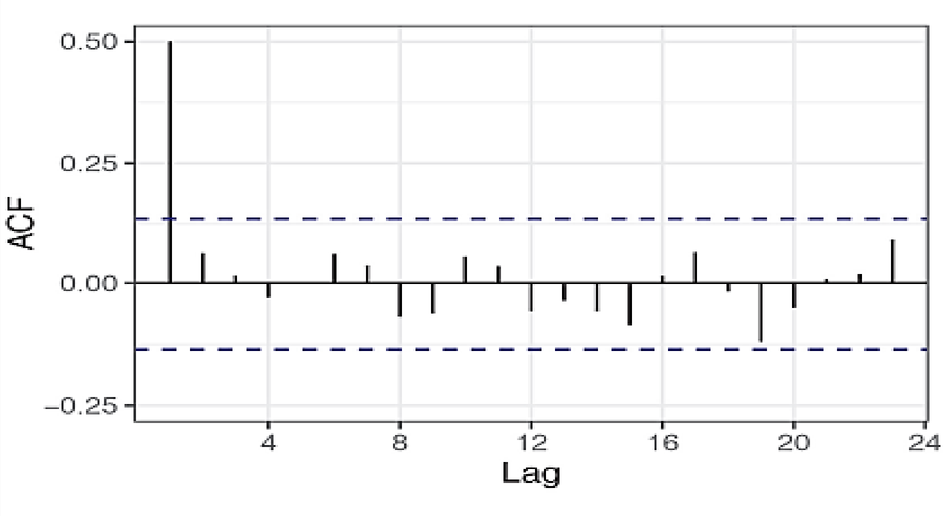
2 표본자기상관계수

- **상관도표(correlogram)**

  → x **축**은 시차 h, y**축은** $\hat{\rho}(h)$**인 도표**

- **상관도표의 점선 :** $H_0 : \rho(h) = 0$ **검정**

  → **점선은** 5% **유의수준 기각역**



3 표본부분자기상관계수

- **부분자기상관계수 : Y_t와 Y_{t+h}의 부분자기상관계수**
  → Y_t와 Y_{t+h} 사이 기간 $Y_{t+1}, \cdots, Y_{t+h-1}$의
  영향력 제거 후 구한 자기상관계수

$$\phi(h) = Corr(Y_t, Y_{t+h} \mid Y_{t+1}, \cdots, Y_{t+h-1})$$
$$= \rho(h \mid Y_{t+1}, \cdots, Y_{t+h-1})$$

3 표본부분자기상관계수

- Y_t와 Y_{t+h}간 표본부분자기상관계수

$$Z_{t+h} = Y_{t+h} - \hat{Y}_{t+h}$$
$$= Y_{t+h} - (\beta_1 Y_{t+h-1} + \cdots + \beta_{h-1} Y_{t+1})$$

$$\hat{\phi}(h) = \frac{\sum (Z_t - \bar{Z})(Z_{t+h} - \bar{Z})}{\sum (Z_t - \bar{Z})^2}$$

3 표본부분자기상관계수

- **시계열 $Z_1, Z_2, \cdots, Z_n$ 서로 독립, 동일한 분포**

  $\rightarrow \hat{\phi}(h)$: **점근적으로** $N(0, \frac{1}{n})$

- **귀무가설 $H_0 : \phi(h) = 0$ 검정**

  $\rightarrow |\hat{\phi}(h)| > 1.96/\sqrt{n}$ **: 유의수준 5%에서 H_0 기각**

  $\rightarrow Z_t$**가 시차 h에서 유의한 부분자기상관관계 있음**

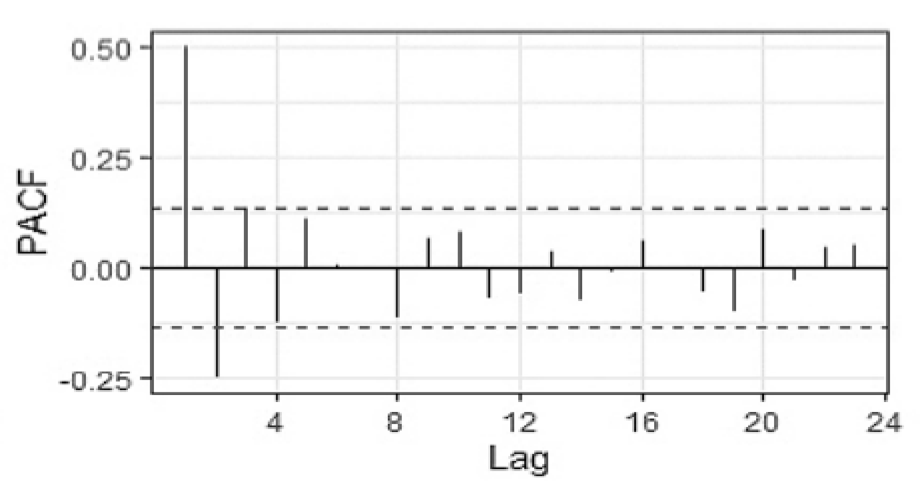
3 표본부분자기상관계수

- **부분상관도표(partial correlogram)**

  → x축은 시차 h, y축은 $\hat{\phi}(h)$인 도표

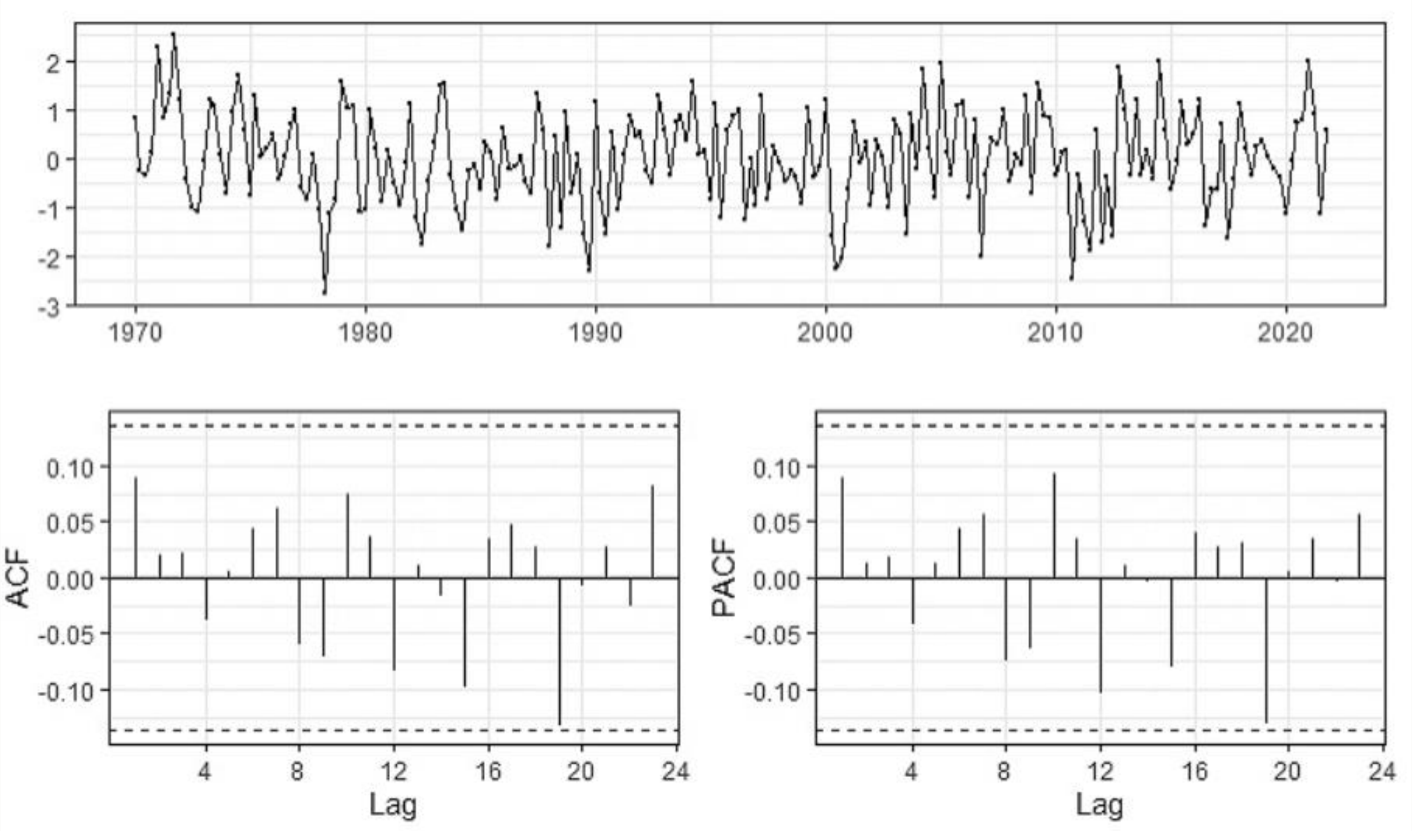
- **상관도표의 점선 :** $H_0 : \phi(h) = 0$ **검정**

  → **점선은** 5% **유의수준 기각역**



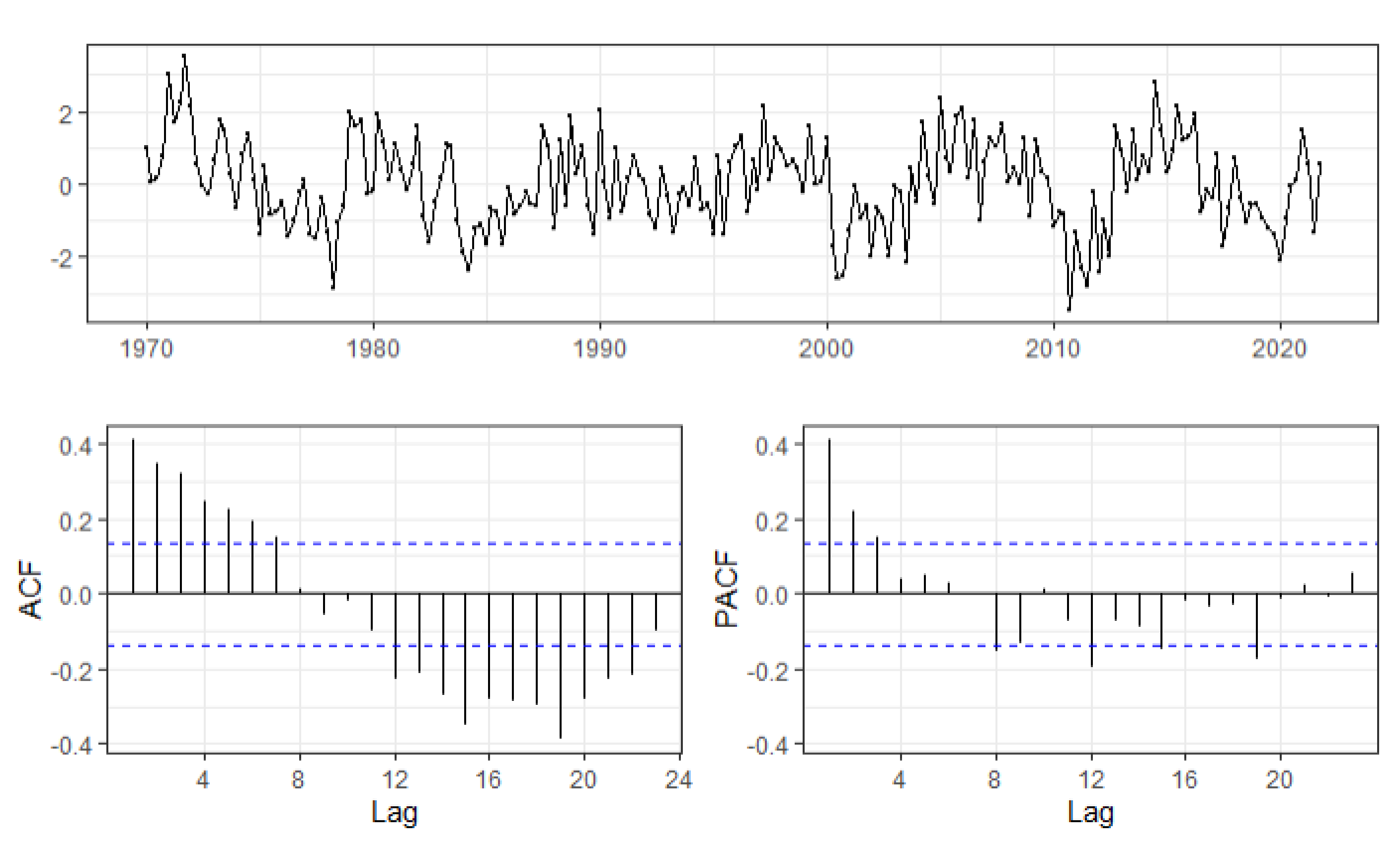
4 여러 시계열의 상관도표와 부분상관도표

- **백색잡음계열**



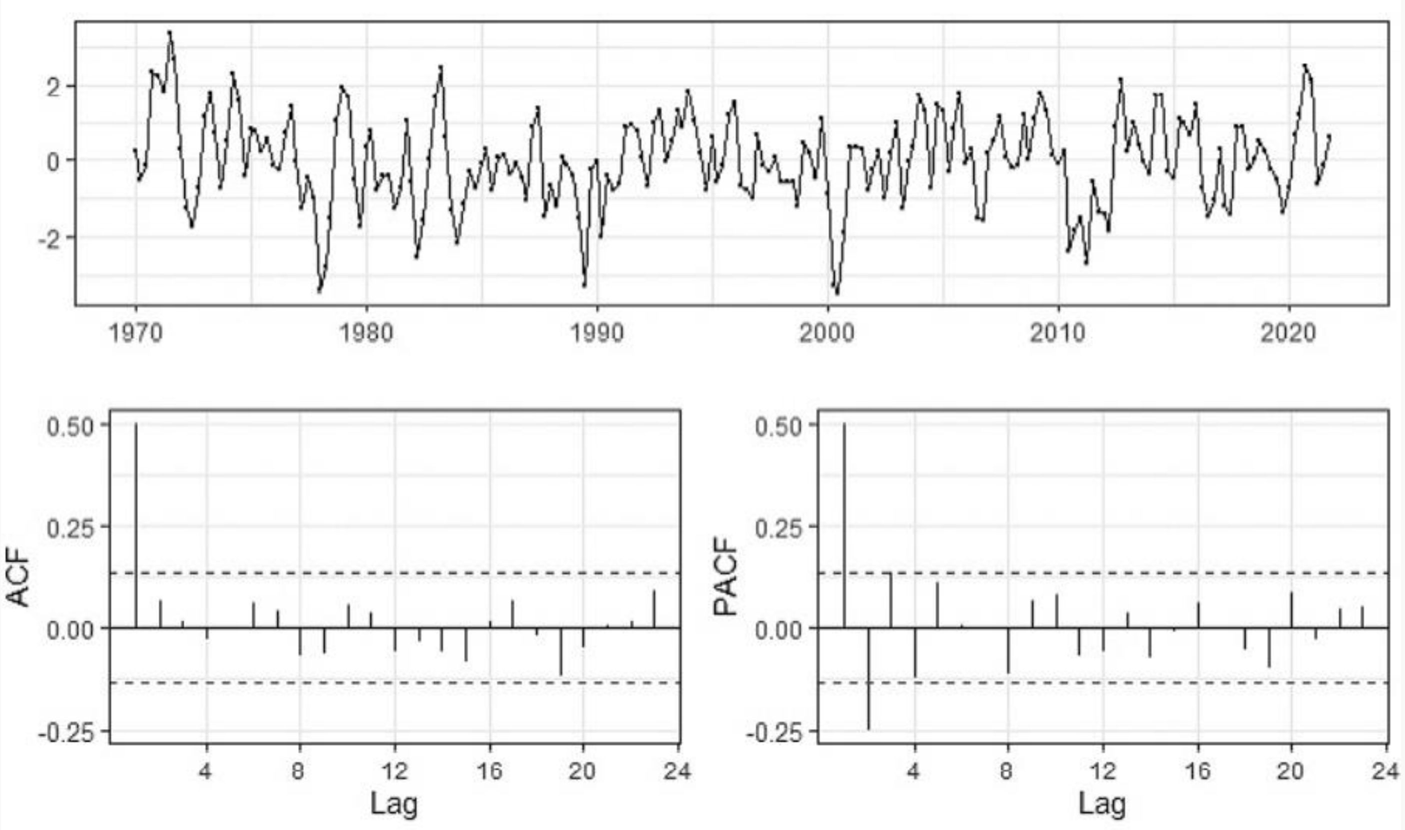
4 여러 시계열의 상관도표와 부분상관도표

- sin 계열



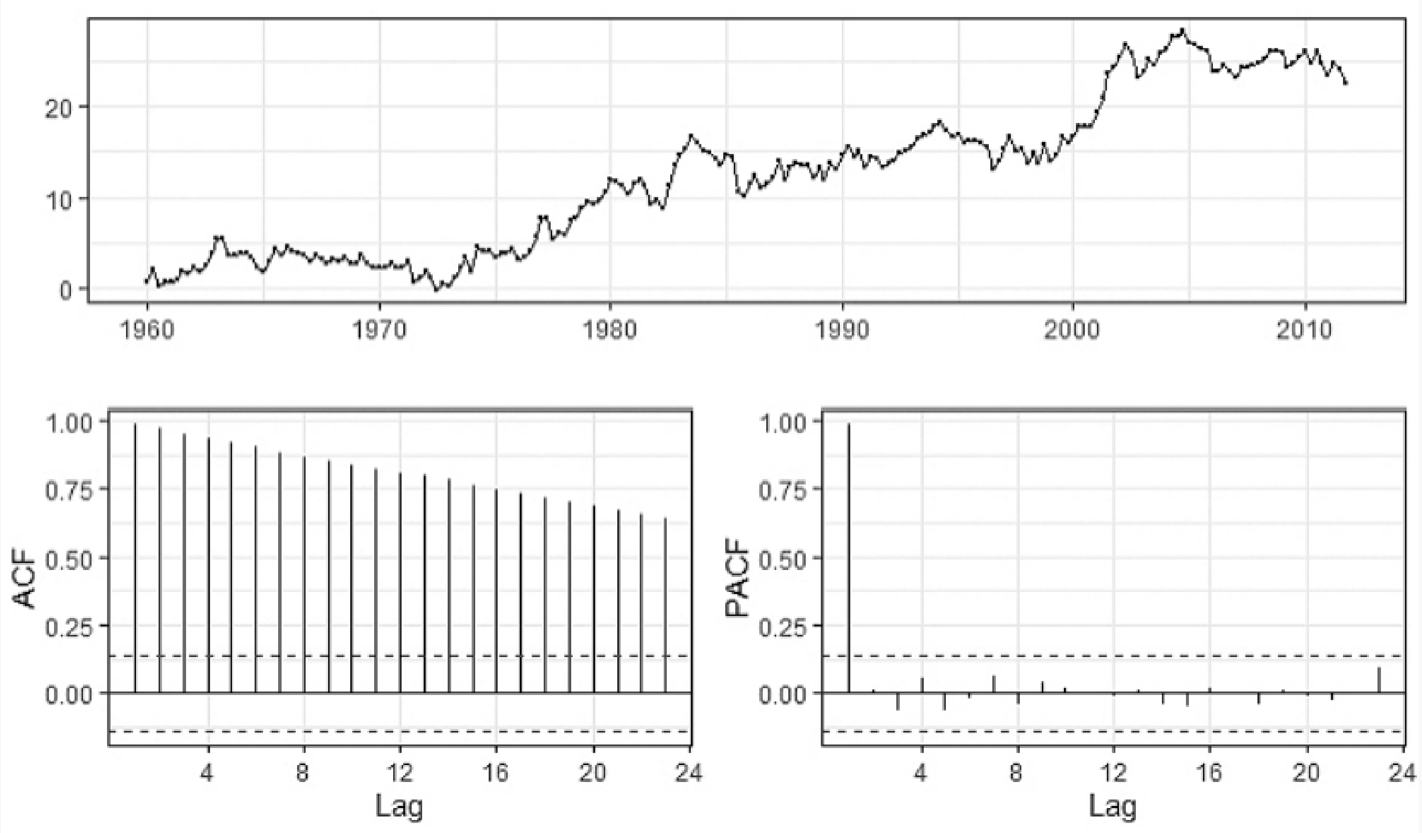
4 여러 시계열의 상관도표와 부분상관도표

- **이동평균계열**



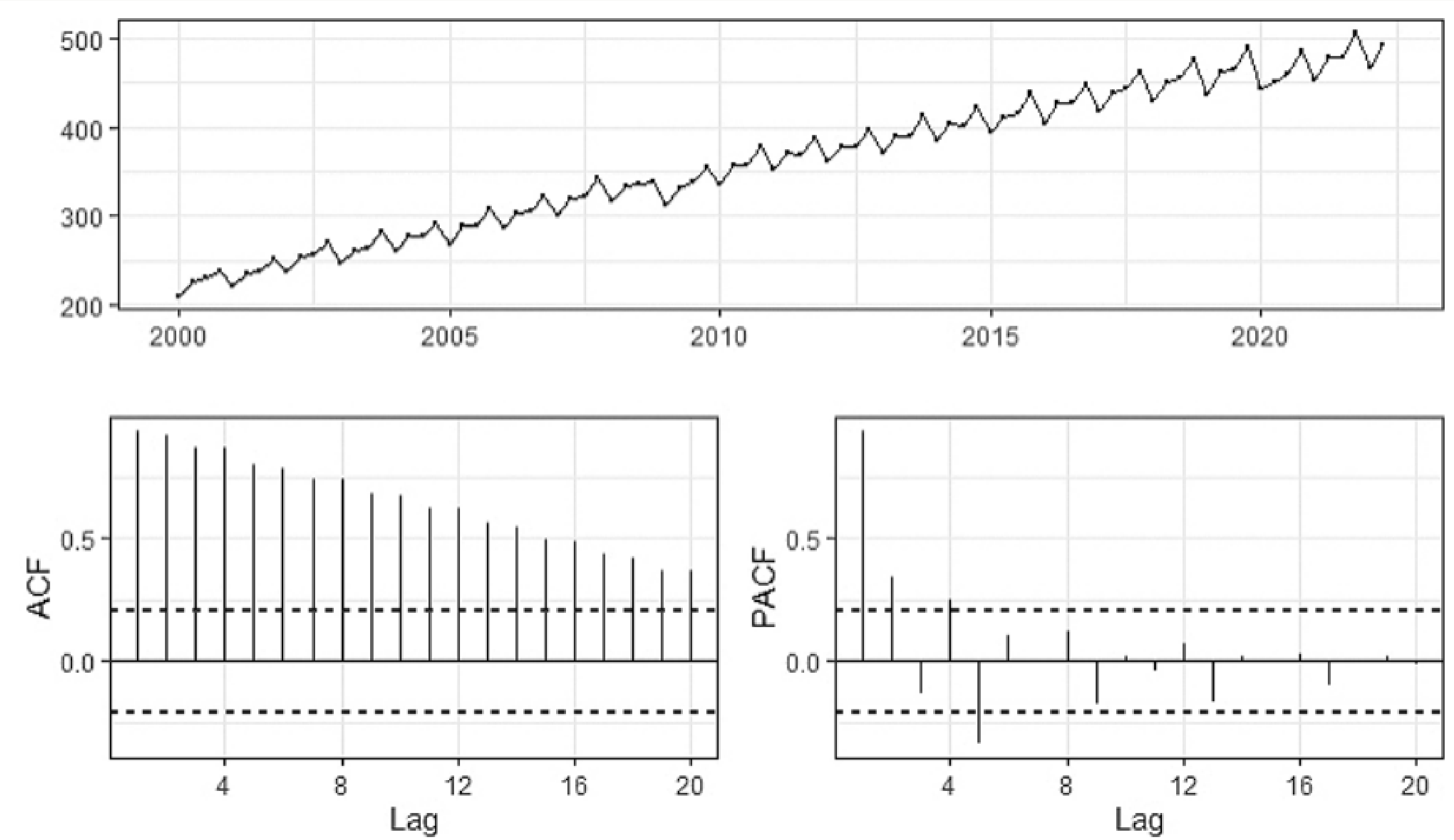
4 여러 시계열의 상관도표와 부분상관도표

- **확률보행계열**



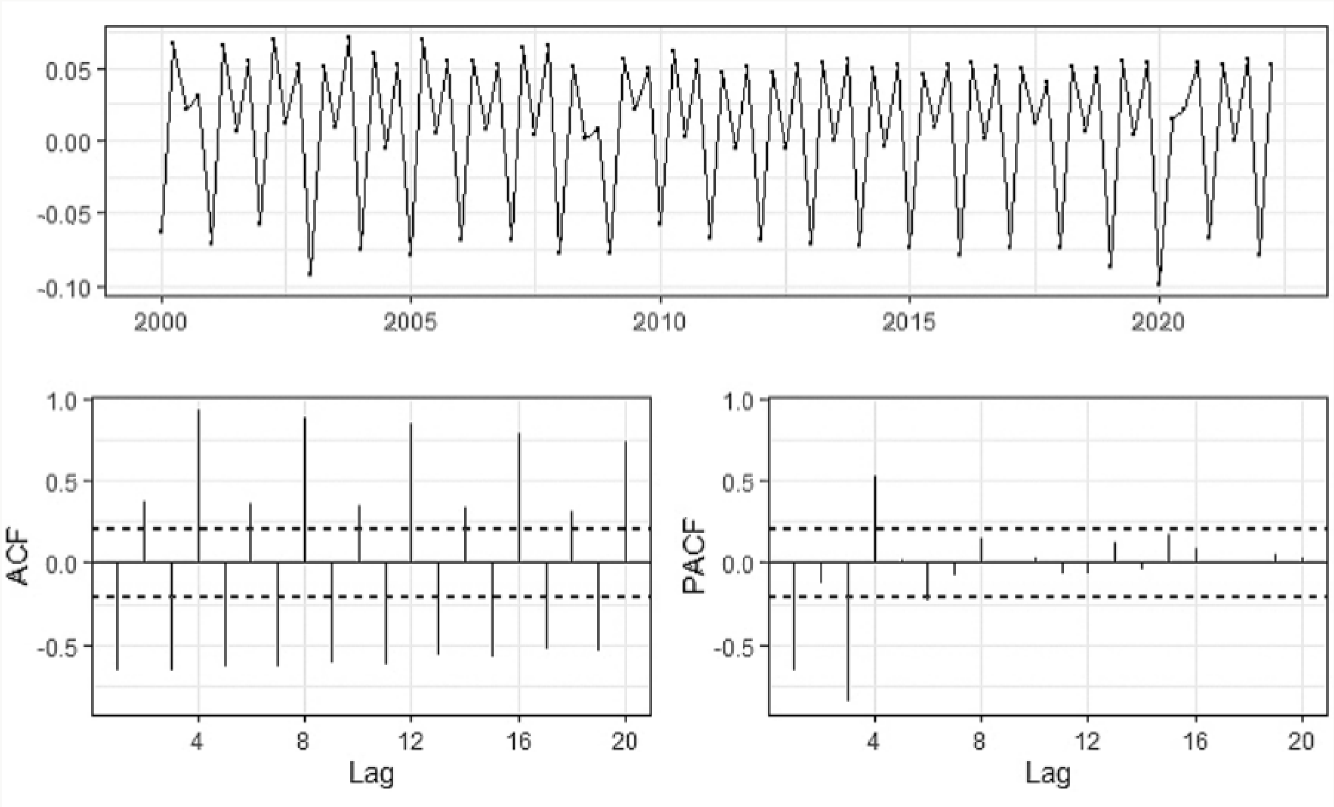
4 여러 시계열의 상관도표와 부분상관도표

- GDP 원계열



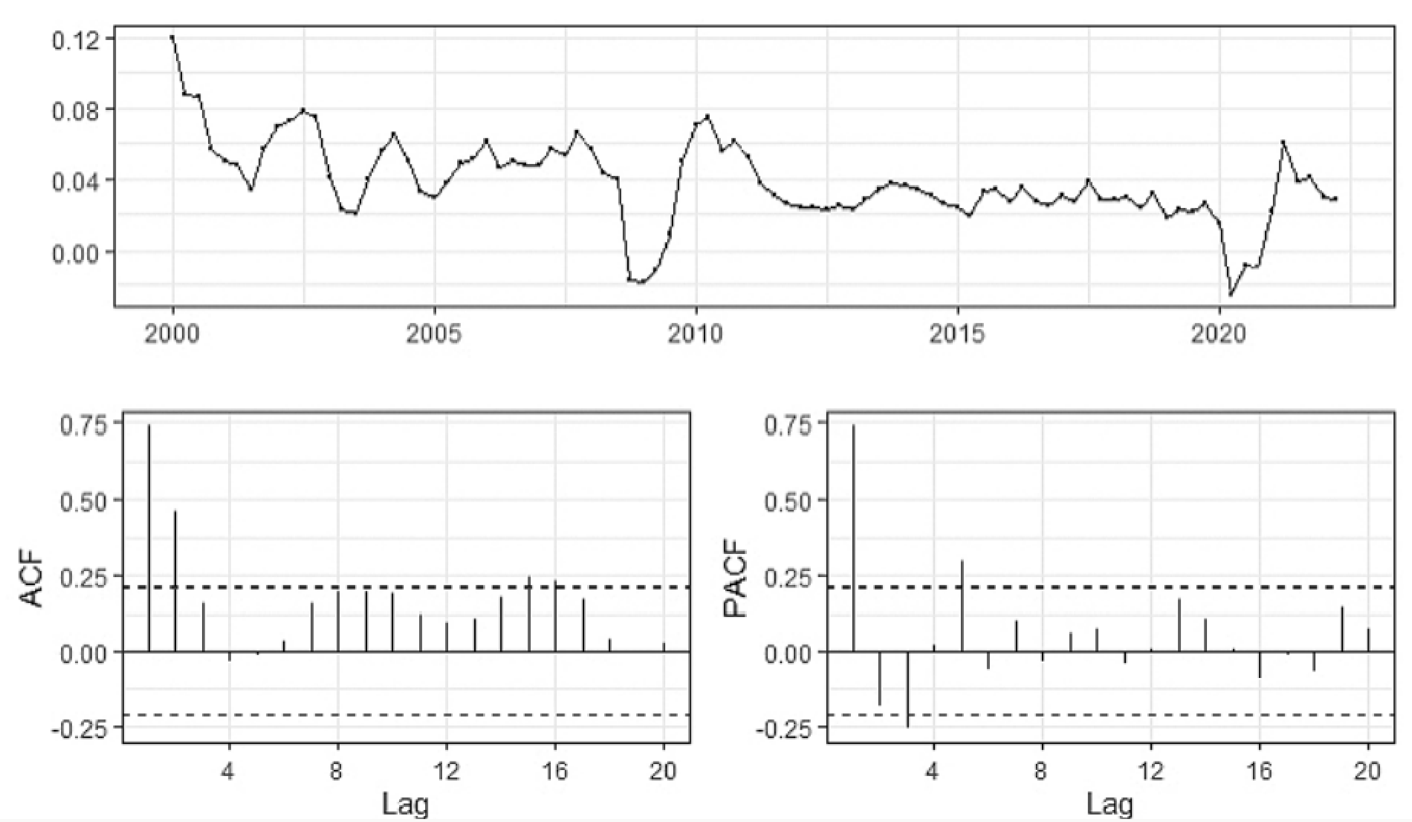
4 여러 시계열의 상관도표와 부분상관도표

- GDP 원계열의 로그 차분 계열



4 여러 시계열의 상관도표와 부분상관도표

- GDP 원계열 의 로그 4차 차분



4 륭과 박스의 검정

- 시계열이 다변량 정규분포, 모든 자기상관계수가 0
  → 시계열은 백색잡음계열
- 륭과 박스(Ljung and Box)의 검정(portmanteau 검정)
  → 일정 시차까지의 자기상관계수가 동시에 모두 0인지 검정

4 륭과 박스의 검정

- **귀무가설**

$$H_0 : \rho(1) = \rho(2) = \cdots = \rho(m) = 0$$

- **검정통계량 :**

$$Q_m = n(n+2)\sum_{h=1}^{m}\frac{\hat{\rho}(h)^2}{(n-h)}$$

→ $Q_m > \chi^2_{m,\alpha}$ **또는 유의확률(p-value)가 α보다 작으면**

: 유의수준 α 에서 귀무가설 기각

4 륭과 박스의 검정

- 륭-박스 검정 결과

계열명	χ^2검정통계량값	유의확률(p-value)
백색잡음계열	4.20	0.8384
이동평균계열	56.14	2.645e-09
sin 계열	119.77	< 2.2e-16
확률보행계열	1472.10	< 2.2e-16
GDP 원계열	538.88	< 2.2e-16
GDP 로그차분계열	343.60	< 2.2e-16
GDP 로그 4차 차분계열	75.51	9.793e-13

chapter 03

안정시계열과 불안정시계열

1 안정시계열

- 시계열 : 안정(stationary)시계열과 불안정(nonstationary)시계열로 구분
- 시계열의 안정성 : 강안정성과 약안정성

» 강안정성
: 구간이 달라지더라도 매 구간별 특성이 동일한 시계열

» 약안정성
: 시계열의 평균과 분산에 시간에 따른 규칙적인 변화가 없고
자기공분산이 시차에 의존하는 시계열

1 안정시계열

- **강안정성을 가지는 시계열**

$$P(Y_{t_1} \le y_1, Y_{t_2} \le y_2, \cdots, Y_{t_k} \le y_k)$$
$$= P(Y_{t_1+h} \le y_1, Y_{t_2+h} \le y_2, \cdots, Y_{t_k+h} \le y_k)$$

(단, $k = 1, 2, 3, \cdots,\ h = 0, \pm 1, \pm 2, \cdots$)

1 안정시계열

- 약안정성을 가지는 시계열

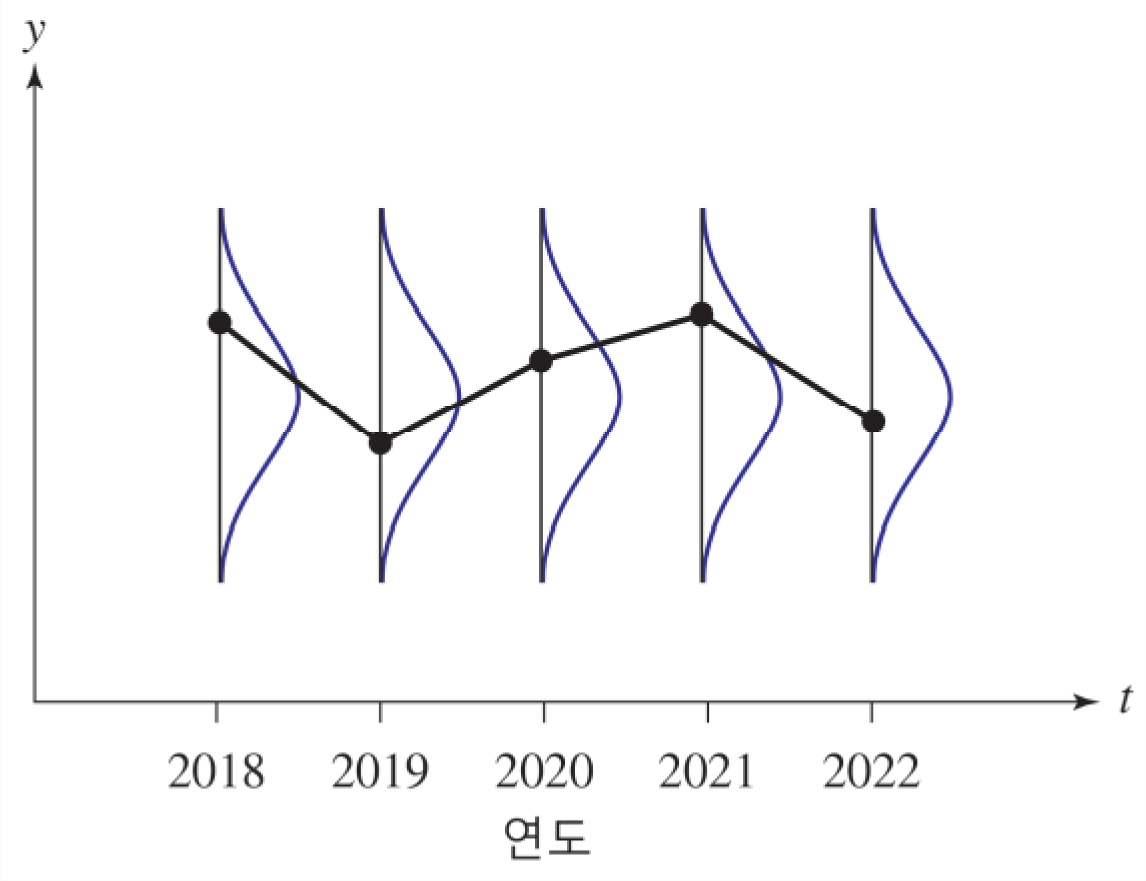
$$E(Y_t) = \mu$$

$$\gamma(s,t) = \gamma(0, |s-t|)$$

- 시계열의 안정성 : 강안정성보다는 약안정성으로 파악

1 안정시계열

- 안정시계열의 의미



2 불안정시계열

- 불안정시계열은 안정시계열이 아닌 시계열
  → 기댓값, 또는 공분산(분산)이 시간에 따라 변하는 시계열

- 경제시계열 : 대체로 추세변동과 계절변동이 뚜렷한 불안정시계열

3 시계열의 예

- 안정시계열 : 백색잡음계열과 이동평균계열
  → 기댓값이 일정, 자기공분산함수 시차에만 의존

- 불안정시계열 : 확률보행계열
  → 기댓값, 자기공분산함수이 시간에 의존

Forecasting
Methods

chapter
04

R을 이용한 실습

Forecasting Methods

다음시간안내

05 | 시계열모형(1)

Forecasting Methods

한국방송통신대학교