클린업 2주차

5팀 시계열자료분석팀

염예빈
한유진
이재현
박세령
이정우

INDEX

1

1주차 복습

정상성의 조건 정상화 과정 백색잡음

정상성을 만족하는 시계열

1. $E(y_t) = \mu < \infty$

: 평균이 시점 t에 무관하게 일정

2. $Var(y_t) = E(y_t - \mu)^2 = \gamma_0 < \infty$

: 분산이 시점 t에 무관하게 일정

3. $\mathrm{Cov}(y_t, y_{t+k}) = E(y_t - \mu)(y_{t+k} - \mu) = \gamma_k < \infty \quad \forall k$

: 자기공분산(공분산이 시점이 아닌 시차에만 의존)

→ t에 무관하고 k에 의존

정상성의 조건 · **정상화 과정** · 백색잡음

정상화 과정

시계열 그래프 그리기
↓
비정상 시계열인지 확인

→

분산안정화
(로그변환,제곱근변환, Box-cox 변환)

→

추세, 계절성을 가짐
↓
회귀/평활/차분 정상화

↓

시계열의 정상부분/비정상부분 분해

←

비정상을 유발하는 성분(추세, 계절성) 제거

←

정상시계열 획득!

정상성의 조건 정상화 과정 **백색잡음**

백색잡음

$$Y_t = \cancel{T_t} + \cancel{S_t} + I_t$$

추세 계절성 오차

$$\mathbf{Y}_t \sim WN(0, \sigma_Y^2)$$

① $E(Y_t) = 0$

② $Var(Y_t) = \sigma^2$

③ $Cov(Y_t, Y_{t+k}) = 0$

(약정상성조건에 '공분산이 0'이라는 조건이 추가됨)

백색잡음은
가장 대표적인 정상시계열

많은 확률과정들이
백색잡음으로부터
생성될 수 있음

2

ACF, PACF

모형의 필요성

$$\text{공분산 행렬 }: \Gamma = \begin{pmatrix} Cov(Y_1, Y_1) & Cov(Y_1, Y_2) & \cdots & Cov(Y_1, Y_n) \\ Cov(Y_2, Y_1) & Cov(Y_1, Y_2) & \cdots & Cov(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \\ Cov(Y_n, Y_1) & Cov(Y_n, Y_{.}) & \cdots & Cov(Y_n, Y_n) \end{pmatrix}$$

$$= \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix}$$

- 백색잡음 O → 대각선 $\gamma(0)$ 을 제외한 나머지가 모두 0, 추가적인 모델링 필요 없음
- 백색잡음 X → 공분산 $\gamma(1), \gamma(2), \ldots \gamma(n-1)$ 모두 추정해야 하는 모수 특정모형으로 모수 추정 가능

모형의 필요성

$$\Gamma = \begin{pmatrix} Cov(Y_1,Y_1) & Cov(Y_1,Y_2) & \cdots & Cov(Y_1,Y_n) \\ Cov(Y_2,Y_1) & Cov(Y_1,Y_2) & \cdots & Cov(Y_2,Y_n) \\ \vdots & \vdots & \vdots & \\ Cov(Y_n,Y_1) & Cov(Y_n,Y_2) & \cdots & Cov(Y_n,Y_n) \end{pmatrix}$$

$$= \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix}$$

모수를 추정할 특정 모형은 어떻게 결정하나요?

- 백색잡음 O → 대각선 $\gamma(0)$ 을 제외한 나머지가 모두 0, 추가적인 모델링 필요 없음.
- 백색잡음 X → 공분산 $\gamma(1), \gamma(2), \ldots, \gamma(n-1)$ 모두 추정해야 하는 모수
  특정모형으로 모수 추정 가능

Y_t의 ACF, PACF 그래프를 보고 알 수 있다!

자기공분산함수(Auto-Covariance Function)

$$\gamma(k) = \mathrm{Cov}(X_t, X_{t+k})$$
$$= E[(X_t - \mu)(X_{t+k} - \mu)]$$

자기상관계수(Auto-Correlation Funciton)

$$\rho(k) = Corr(X_t, X_{t+k}) = \frac{Cov(X_t, X_{t+k})}{\sqrt{Var(X_t)}\sqrt{Var(X_{t+k})}} = \frac{\gamma(k)}{\gamma(0)},$$

이때 $\gamma_0 = Var(X_t) = E\left[(X_t - \mu)^2\right]$

표본자기공분산함수(SACVF)

$$\widehat{\gamma_{(k)}} = \frac{1}{T}\sum_{j=1}^{T-k}(X_j - \bar{X})(X_{j+k} - \bar{X}),$$
$$k = 1, 2, 3, \ldots$$

표본자기상관함수(SACF)

$$\widehat{\rho_{(k)}} = \frac{\widehat{\gamma_{(k)}}}{\widehat{\gamma_{(0)}}}$$

$\rho(k) = \ Corr(X_t, X_{t+k})$

자기상관함수(ACF)는
시차 k에서 자기상관관계가
존재하는지 나타내는 척도

서로 다른 두 시점의
상호 연관관계를 나타냄

PACF(Partial Autocorrelation Function) : 편자기상관함수

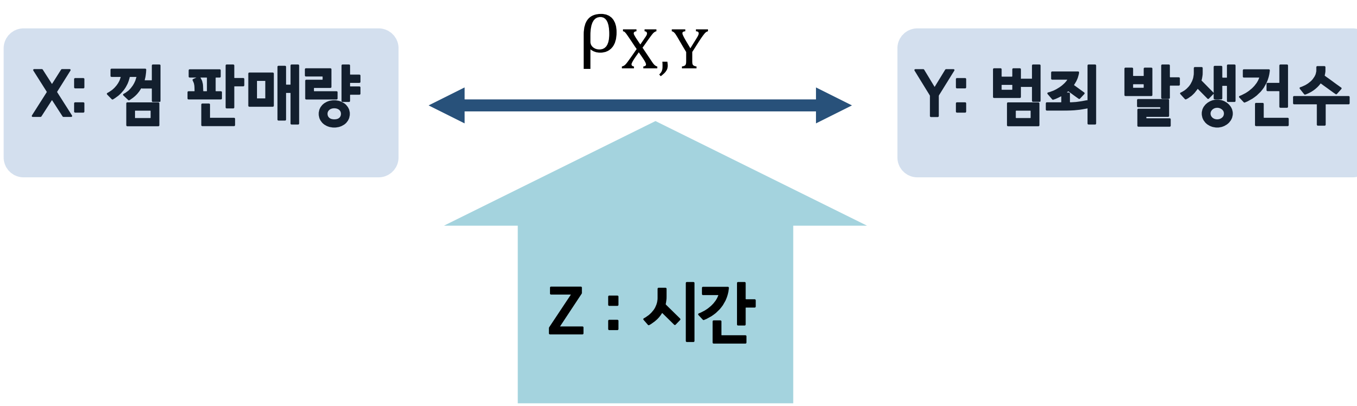
PACF

:두 변수를 제외한 **모든 변수의 영향을 제거**한 상태에서,

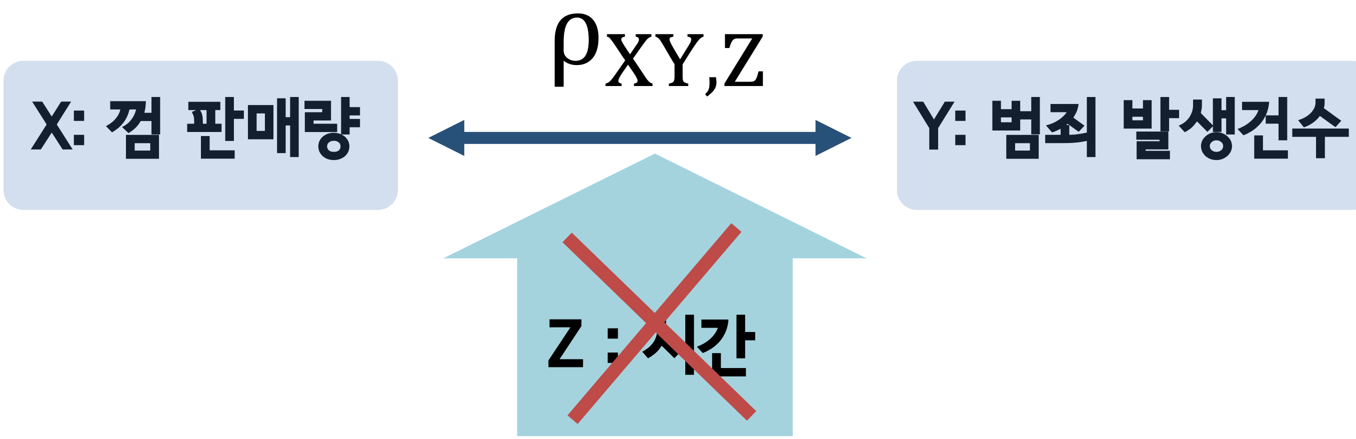
두 변수사이의 **순수한 상호연관관계**

PACF(Partial Autocorrelation Function) : 편자기상관함수

X를 껌 판매량, Y를 범죄 발생건수라고 하자. 시간에 따라 관측된 두 변수들 사이의 상관계수를 구해보면 매우 상관이 높은 것으로 나타날 것이다. 이는 두 변수사이에 밀접한 관계가 있어서라기보다는 시간이 지남에 따라 인구가 증가하면서 껌의 판매량이 늘고 범죄 발생건수도 증가했기 때문이라고 볼 수 있다



부분자기상관계수(partial correlation): PACF



X와 Y의 순수한 상관관계를 구하기 위해서는
X와 Y에서 시간의 효과를 제거한 후 상관계수를 구해야 함
→ **부분상관계수**(partial correlation coefficient)

부분자기상관계수(partial correlation): PACF

Z의 효과를 배제한 후의 X와 Y의 부분상관계수

$$\rho_{XY,Z} = \frac{E\{[X - E(X|Z)] \cdot [Y - E(Y|Z)]\}}{\sqrt{E\{[X - E(X|Z)]^2 \cdot E[Y - E(Y|Z)]^2}}$$

조건부 기댓값 $\mathrm{E}:(\mathrm{X}|\mathrm{Z})$ X를 Z에 회귀시킨 최적선형 예측값으로 X가 Z에 의해 설명되는 부분

조건부 기댓값 $\mathrm{E}(Y|\mathrm{Z})$: Y를 Z에 회귀시킨 최적선형 예측값으로 Y가 Z에 의해 설명되는 부분

$\mathrm{X}^* = \mathrm{X} - \mathrm{E}(\mathrm{X}|\mathrm{Z})$: X를 Z에 회귀시킨 후의 잔차

$\mathrm{Y}^* = \mathrm{Y} - \mathrm{E}(Y|\mathrm{Z})$: Y를 Z에 회귀시킨 후의 잔차

부분자기상관계수(partial correlation): PACF

$X^* = X - E(X|Z)$: X를 Z에 회귀시킨 후의 잔차

$Y^* = Y - E(Y|Z)$: Y를 Z에 회귀시킨 후의 잔차

X^*, Y^*

: 원래변수 X,Y가 간직하고 있던 정보 중에서 Z와 무관한 부분,

Z와는 무관한 변수 **X, Y의 순수한 상관계수**를 의미

X^*, Y^*의 상관계수 ρ_{X^*,Y^*}

: 변수Z에 관하여 수정한 후의 X와 Y의 **부분상관계수** $\rho_{XY,Z}$ 와 같아진다.

부분자기상관계수(partial correlation): PACF

PACF : ϕ_{kk}

$Z_t, Z_{t+1}, \dots, Z_{t+k-1}, Z_{t+k}$ 가 관측되었을 때,

k시차만큼 떨어진 Z_t, Z_{t+k} **의 순수한 상관관계**를 나타냄

➔ Z_t,와 Z_{t+k} 에서 $Z_{t+1}, Z_{t+2} \dots, Z_{t+k-1}$ **의 효과를 제거**한 후의 상관계수

$\widehat{\phi_{kk}}$: 표본부분자기상관함수(sample partial autocorrelation function : SPACF)

부분자기상관계수(partial correlation): PACF

<부분자기상관계수>

ϕ_{kk}

Z_t ⟷ Z_{t+k}

$Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$ **효과 제거**

$Z_t, Z_{t+1}, \dots, Z_{t+k-1}, Z_{t+k}$ 가 관측되었을 때, PACF, ϕ_{kk}

k시차만큼 떨어진 Z_t, Z_{t+k} 의 순수한 상관관계를 나타냄

➔ Z_t, 와 Z_{t+k} … 의 상관계수

$\widehat{\phi_{kk}}$: 표본부분… ion : SPACF)

3

AR, MA, ARMA

정상시계열 모형 종류 정상시계열 모형 조건

정상시계열 모형의 종류

AR

MA

ARMA

정상시계열 모형 종류

정상시계열 모형 조건

정상시계열 모형의 조건

우리는 정상시계열 모형의 조건!!
우리 중 하나라도 빠지면 정상시계열 모형을 쓸 수 없지

정상시계열 모형의 조건

1. 정상성

통계적 특성이

시점에 의존하지

않아야 함.

2. 인과성

현재 시점의

관측값을 과거시점의

오차항으로 설명 가능

3. 가역성

오차항을

과거시점의

관측값으로 설명 가능

AR 모형

자기 회귀 모형

= **A**uto **R**egressive Model

자기 자신을 과거 시점에 회귀

현재의 자료 ← 설명O ← 과거의 자료

\+

현재의 자료 ← 설명X ← 오차항

AR 모형

$$AR(P) : Z_t - \mu = \phi_1(Z_{t-1}-\mu) + \phi_2(Z_{t-2}-\mu) + \ldots + \phi_p(Z_{t-p}-\mu) + \boxed{\varepsilon_t}$$

$$= \sum_{j=1}^{p} \phi_j \boxed{(Z_{t-j}-\mu)} + \varepsilon_t$$

백색잡음

$$\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

$$\dot{Z}_t = Z_{t-j} - \mu$$

$$\dot{z}_t = \sum_{j=1}^{p} \phi_j \dot{z}_{t-j} + \varepsilon_t$$

모형의 표현 모형의 조건 ACF, PACF

AR 모형

$\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$

백색잡음

$$AR(P) : Z_t - \mu = \phi_1(Z_{t-1} - \mu) + \phi_2(Z_{t-2} - \mu) + \ldots + \phi_p(Z_{t-p} - \mu) + \varepsilon_t$$

$$= \sum_{j=1}^{p} \phi_j (Z_{t-j} - \mu) + \varepsilon_t$$

$$\dot{Z}_t = Z_{t-j} - \mu$$

현재자료를 설명 O

현재자료를 설명 X

$$\dot{z}_t = \boxed{\sum_{j=1}^{p} \phi_j \dot{z}_{t-j}} + \varepsilon_t$$

AR 모형

백색잡음

$$AR(P) : Z_t - \mu = \phi_1 (Z_{t-1} - \mu) + \phi_2 (Z_{t-2} - \mu) + \cdots + \phi_p (Z_{t-p} - \mu) + \varepsilon_t$$

$$= \sum_{j=1}^{p} \phi_j (Z_{t-j} - \mu) + \varepsilon_t$$

$$\dot{Z}_t = Z_t - \mu \qquad \dot{Z}_t = \sum_{j=1}^{p} \phi_j \dot{Z}_{t-j} + \varepsilon_t$$

후항연산자(Backshift Operator)

$$BX_t = X_{t-1}$$

1주차 때 배운 후항연산자로 간단하게 표현해 보자!!

AR 모형

$$\underline{\Phi(B)}(Z_t - \mu) = \Phi(B)\dot{Z}_t = \varepsilon_t$$

특성함수(characteristic function)

$$\boxed{\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_p B^p}$$

모형의 표현 | 모형의 조건 | ACF, PACF

AR 모형

1. 정상성

시계열의
통계적 특성이
시점에 의존하지
않아야 함.

조건 필요

2. 인과성

현재 시점의
관측값을 과거시점의
오차항으로 설명 가능

조건 필요

3. 가역성

오차항을
과거시점의
관측값으로 설명 가능

AR의 정의적 특성 때문에 만족

AR모형

1. 정상성 2. 인과성 3. 가역성

시계열의
통계적 특성이
시점에 의존하지

현재 시점의
관측값을 과거시점의

오차항을
과거시점의

조건 필요 조건 필요 AR의 정의적 특성 때문에 만족

AR모형이
정상성과 인과성을 만족하려면?

“$\Phi(B) = 0$ 의 근의 절댓값” > 1

AR 모형

"Φ(B) = 0 의 근의 절댓값" >1 일 때

1. 정상성

시계열의
통계적 특성이
시점에 의존하지
않아야 함.

만족

2. 인과성

현재 시점의
관측값을 과거시점의
오차항으로 설명 가능

만족

3. 가역성

오차항을
과거시점의
관측값으로 설명 가능

만족

모형의 표현 | 모형의 조건 | ACF, PACF

AR 모형

"Φ(B) = 0 의 근의 절댓값" =1 일 때

1. 정상성

시계열의 통계적 특성이 시점에 의존하지 않아야 함.

불만족

2. 인과성

현재 시점의 관측값을 과거시점의 오차항으로 설명 가능

논 외

3. 가역성

오차항을 과거시점의 관측값으로 설명 가능

만족

AR모형

"Φ(B) = 0 의 근의 절댓값" =1 일때

1. 정상성

시계열의 통계적 특성이 시점에 의존하지 않아야 함.

불만족

2. 인과성

관측값을 과거시점의 오차항으로 설명 가능

논 의

3. 가역성

오차항을 과거시점의 관측값으로 설명 가능

만족

Random Walk Process

대표적인 비정상성 시계열

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

AR 모형

"Φ(B) = 0 의 근의 절댓값" <1 일 때

1. 정상성

시계열의
통계적 특성이
시점에 의존하지
않아야 함.

만족

2. 인과성

현재 시점의
관측값을 과거시점의
오차항으로 설명 가능

불만족

3. 가역성

오차항을
과거시점의
관측값으로 설명 가능

만족

AR모형

“Φ(B) = 0 의 근의 절댓값” <1 일때

1. 정상성	2. 인과성	3. 가역성
시계열의 통계적 특성이 시점에 의존하지 않아야 함.	현재시점의 관측 값을 과거시점의 오차항으로 설명 가능	오차항을 과거시점의 관측값으로 설명 가능
만족	불만족	만족

> 과거시점이 아닌
> **미래시점에 의존** 하는 시계열

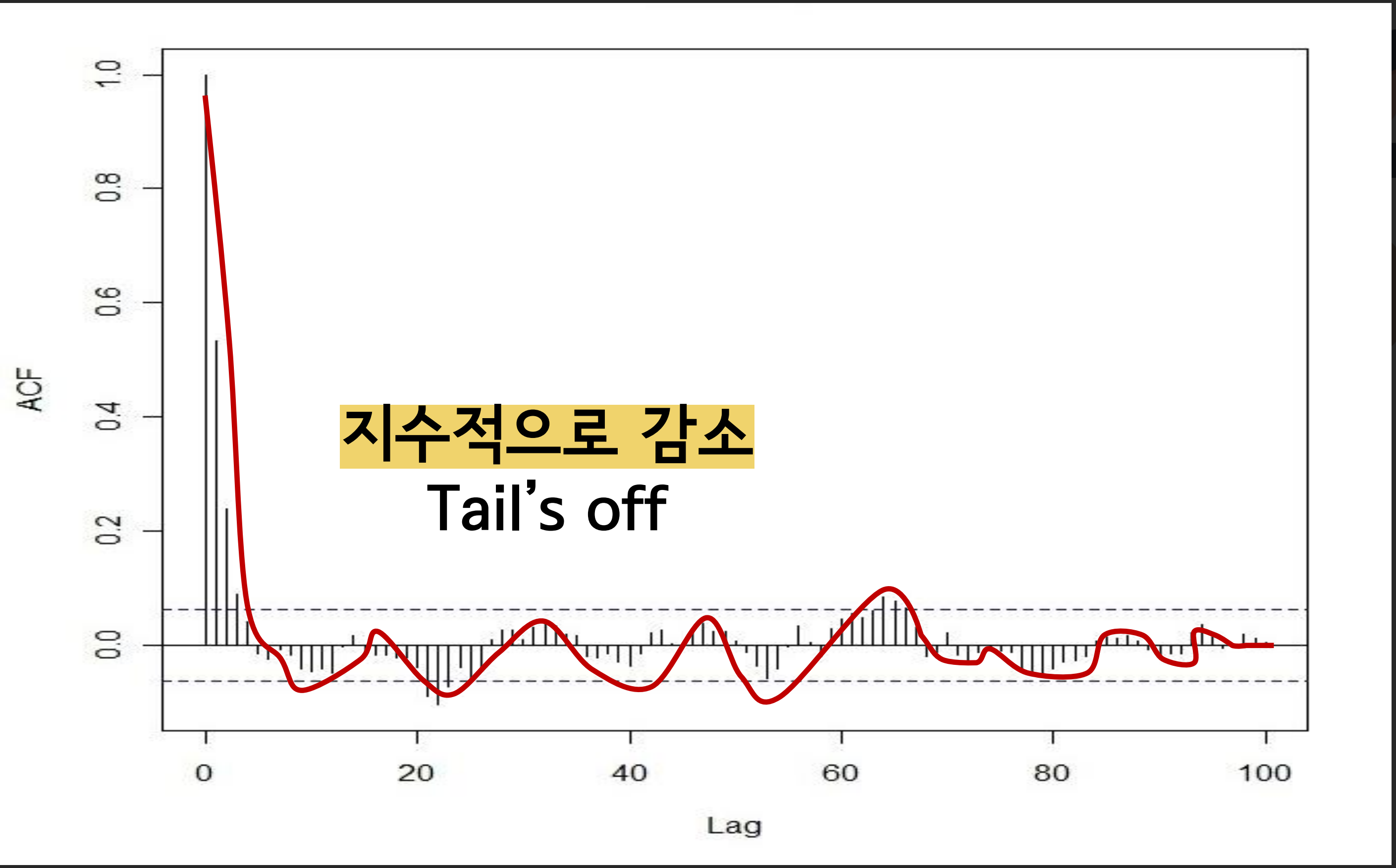
AR 모형

ACF와 PACF로 시각적으로 판단하기!

ACF	PACF
지수/사인함수 모양으로 감소	시차 p+1이후 절단된 모양
Tail's off	Cut off

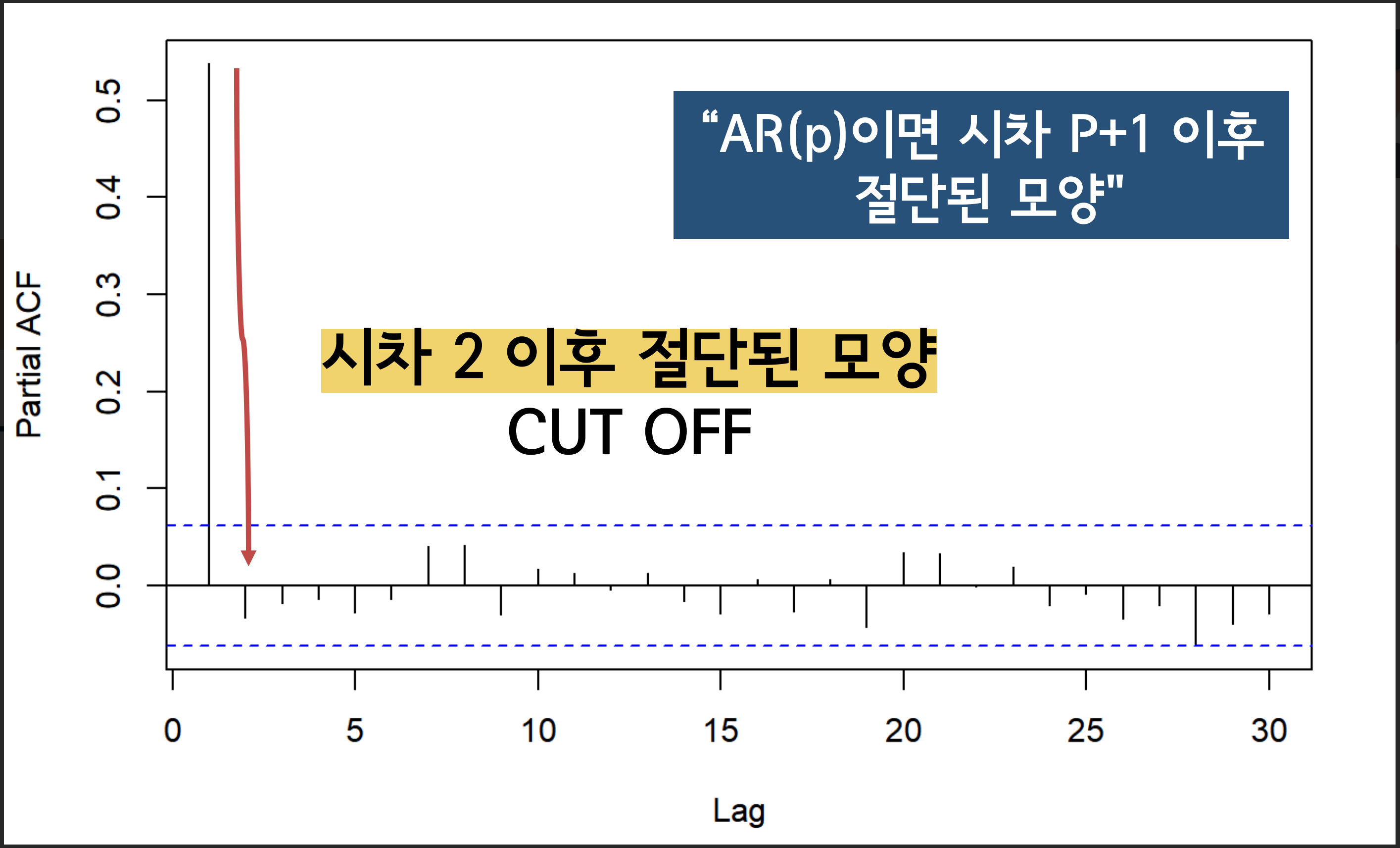
AR모형

AR(1) 모형의 ACF



AR모형

AR(1) 모형의 PACF



MA 모형

이동 평균 모형
= **M**oving **A**verage Model

현재의 관측값이 현재와 과거의 설명해 주지 못하는 부분(**오차**)의 **선형결합**으로 표시되는 모형

현재의 자료 ← 선형결합 ← 과거의 자료의 오차

MA 모형

백색잡음

$\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$

$$MA(q) : Z_t = \boxed{\varepsilon_t} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \ldots - \theta_q \varepsilon_{t-q}$$

$$Z_t = \underline{\theta(B)}\, \varepsilon_t$$

특성함수(characteristic function)

이때, $\theta(B) = 1 - B\theta_1 - B^2\theta_2 - \ldots - -B^q\theta_q$

MA 모형

백색잡음 $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$

$$MA(q) : Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \ldots - \theta_q \varepsilon_{t-q}$$

$$Z_t = \theta(B)\, \varepsilon_t$$

특성함수(characteristic function)

이때, $\theta(B) = 1 - B\theta_1 - B^2\theta_2 - \ldots - -B^q\theta_q$

“ q시점 전까지의 **오차**들의 **선형결합**으로 표현 ”

$$MA(1) : Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

MA 모형

1. 정상성

시계열의
통계적 특성이
시점에 의존하지
않아야 함.

오차항으로만 표현 되기 때문에 만족

2. 인과성

현재 시점의
관측값을 과거시점의
오차항으로 설명 가능

MA의 정의적 특성 때문에 만족

3. 가역성

오차항을
과거시점의
관측값으로 설명 가능

조건 필요

MA모형

MA모형이
가역성을 만족하려면?

3. 가역성

시계열의
통계적 특성이

현재 시점의
관측값을 과거시점의

오차항을
과거시점의

“θ(B) = 0 의 근의 절댓값” > 1

ε_t로만 표현되기 때문에 만족

AR의 정의적 특성 때문에 만족

조건 필요

AR모형

가역성 조건은 왜 필요한가요?

1) ACF와 모형 사이의 일대일 대응관계를 성립하도록 해줍니다.
2) t시점의 오차항을 과거시점의 관측값으로 표현할 수 있습니다.

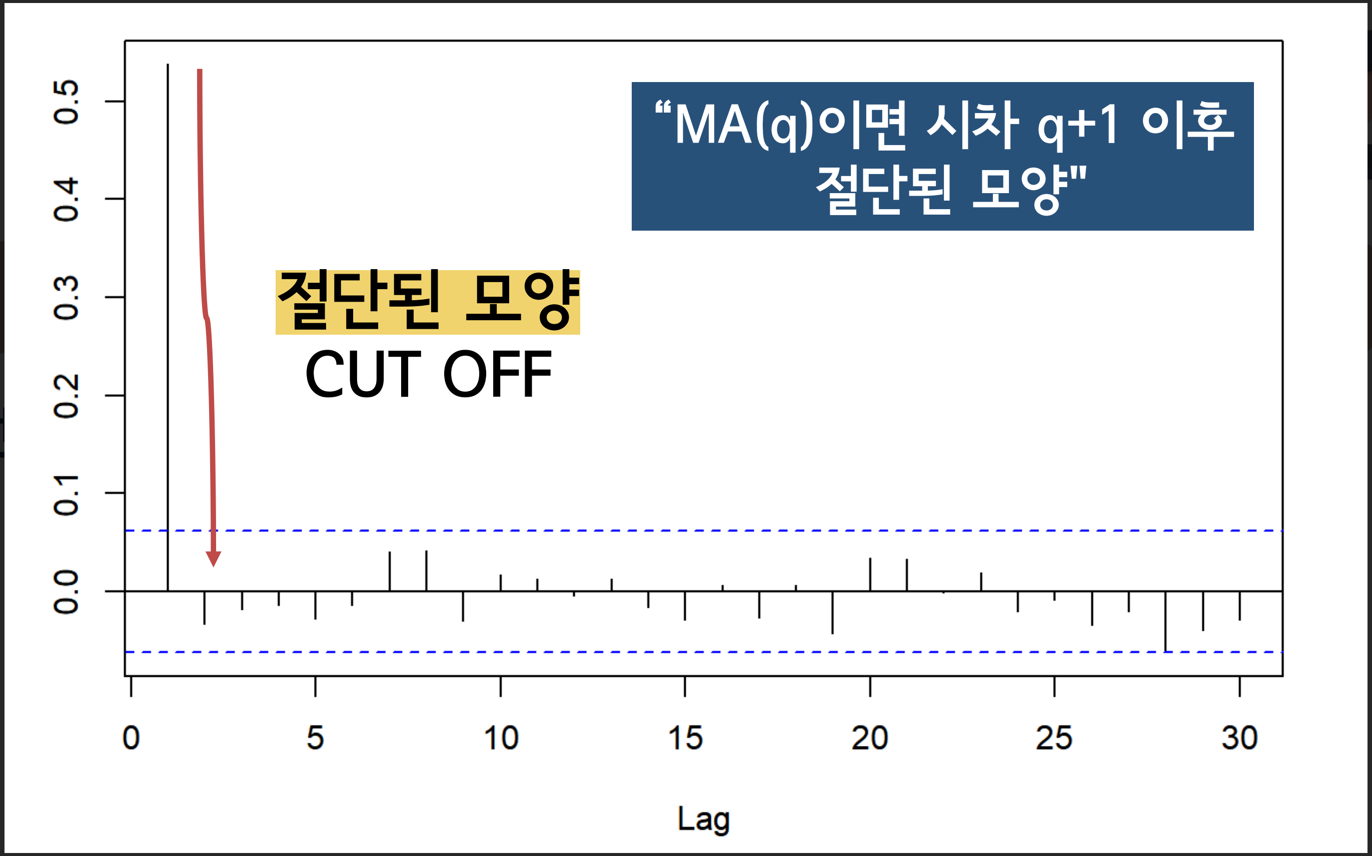
MA 모형

ACF와 PACF로 시각적으로 판단하기!

ACF	PACF
시차 q+1이후 절단된 모양	지수/사인함수 모양으로 감소
Cut off	Tail's off

MA모형

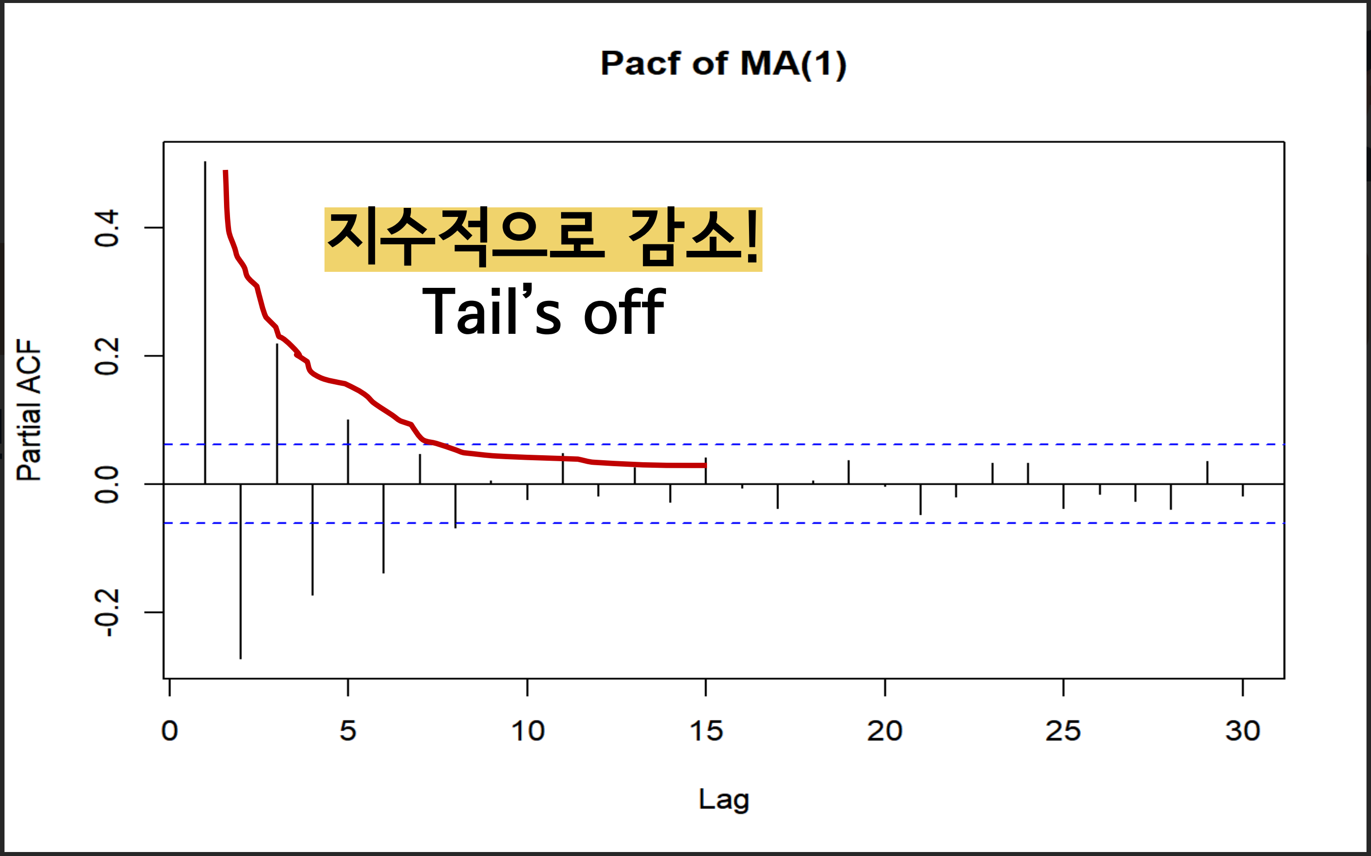
MA(1) 모형의 ACF



곡선

MA모형

MA(1) 모형의 PACF



곡선

쌍대성이란? AR과 MA의 쌍대성

AR과 MA의 쌍대성

변신!!

쌍대성

- 수학적 구조의 쌍대 : 구조를 '뒤집어서' 구성한 것

- 어떤 대상과 그 쌍대는 서로 일종의 한 '켤레'를 이룸
  → 쌍대의 쌍대는 자기 자신

- AR과 MA 사이에는 쌍대성이 존재!!

AR과 MA의 쌍대성

1.

AR(p) MA(∞)

- 유한차수의 정상 AR모형 → 무한차수의 MA모형

MA(q) AR(∞)

- 유한차수의 가역성을 갖는 MA모형 → 무한차수의 AR모형

AR과 MA의 쌍대성

2.

- 유한차수 AR 모형의 ACF & 유한차수 MA 모형의 PACF
  : 지수적으로 감소하는 모양(tail's-off)

- 유한차수 AR 모형의 PACF & 유한차수 MA 모형의 ACF
  : 절단하는 모양(cut-off)

AR과 MA의 쌍대성

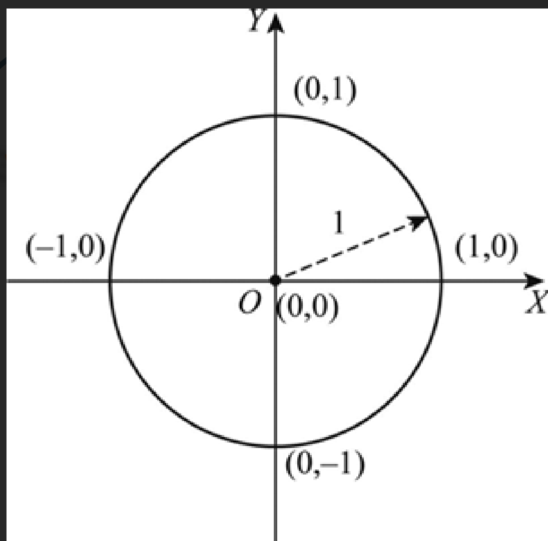
3.

- AR 모형은 정상성 조건이 필요 : $\phi(B) = 0$ 의 근들이 단위원 밖에 존재해야 한다

- MA 모형은 가역성 조건이 필요 : $\theta(B) = 0$의 근들이 단위원 밖에 존재해야 한다

AR과 MA의 쌍대성

[단위원이란?]



: 반지름의 길이가 1인 원

…건이 필요 : $\phi(B) = 0$ 의 근들이 단위원

…조건이 필요 : $\theta(B) = 0$의 근들이 단위원

밖에 존재해야 한다

즉, 해당 조건은 근의 크기가 1보다 커야함을 의미 !!

ARMA

ARMA모형의 필요성

AR, MA 단일 모형 → 단일 모형으로는 분석 모형의 차수 p, q ↑

추정모수가 많아지면서 **효율성 하락 , 해석의 어려움**

AR + MA = **ARMA**

- AR과 MA를 동시에 사용한다면 추정해야 할 모수의 개수를 줄일 수 있다!

ARMA

ARMA 모형 : 자기회귀 이동평균 모형

ARMA(p,q) : AR 모형의 차수 p, MA 모형의 차수 q인 모형

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

ARMA

ARMA 모형 : 자기회귀 이동평균 모형

ARMA(p,q) : AR 모형의 차수 p, MA 모형의 차수 q인 모형

$$\mathrm{Z_t} = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + \quad \textbf{AR}$$

$$\varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad \textbf{MA}$$

ARMA

ARMA 모형 : 자기회귀 이동평균 모형

ARMA(p,q) : AR 모형의 차수 p, MA 모형의 차수 q인 모형

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

AR

MA

<참고>

AR(p) + MA(q) = ARMA(p,q)

AR(1) = ARMA(1,0)

MA(1) = ARMA(0,1)

ARMA

후향연산자로 ARMA 모형 표현

AR MA

$$\left(1-\phi_1 B-\cdots-\phi_p B^p\right) Z_t=\left(1-\theta_1 B-\cdots-\theta_q B^q\right) \varepsilon_t$$

ARMA 모형 $\phi(B)Z_t = \theta(B)\varepsilon_t$ 로 표현 가능

ARMA

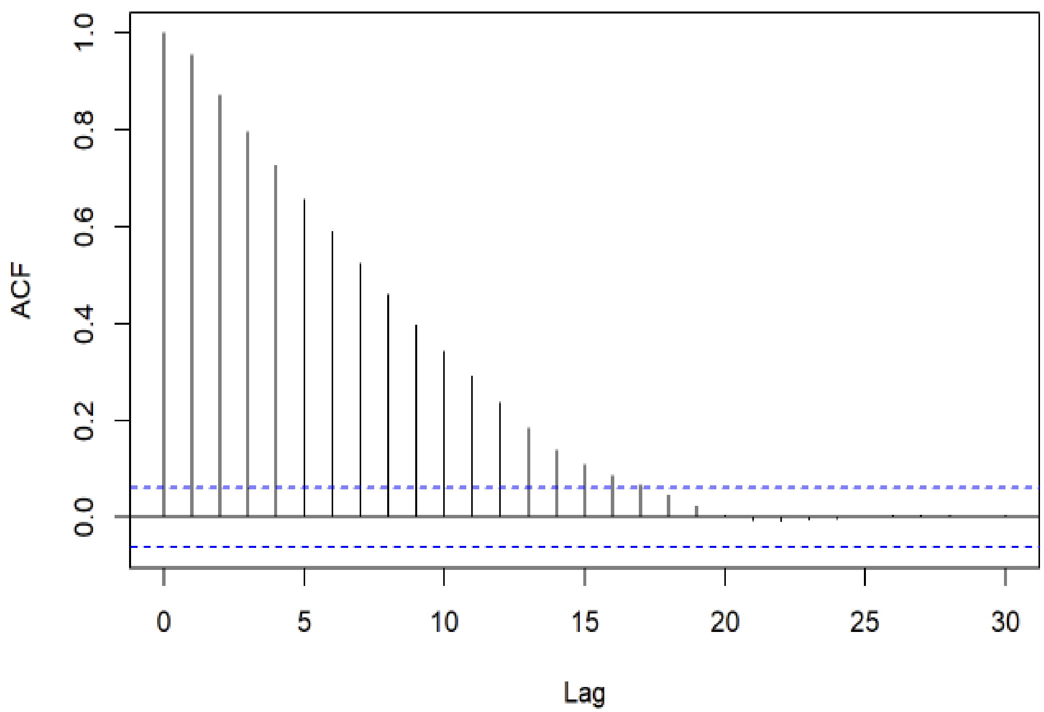
ARMA 모형의 조건

정상성과 **인과성** 만족
: $\phi(B) = 0$ 의 근들이 단위원 밖에 존재해야 한다

가역성 만족
: $\theta(B) = 0$ 의 근들이 단위원 밖에 존재해야 한다

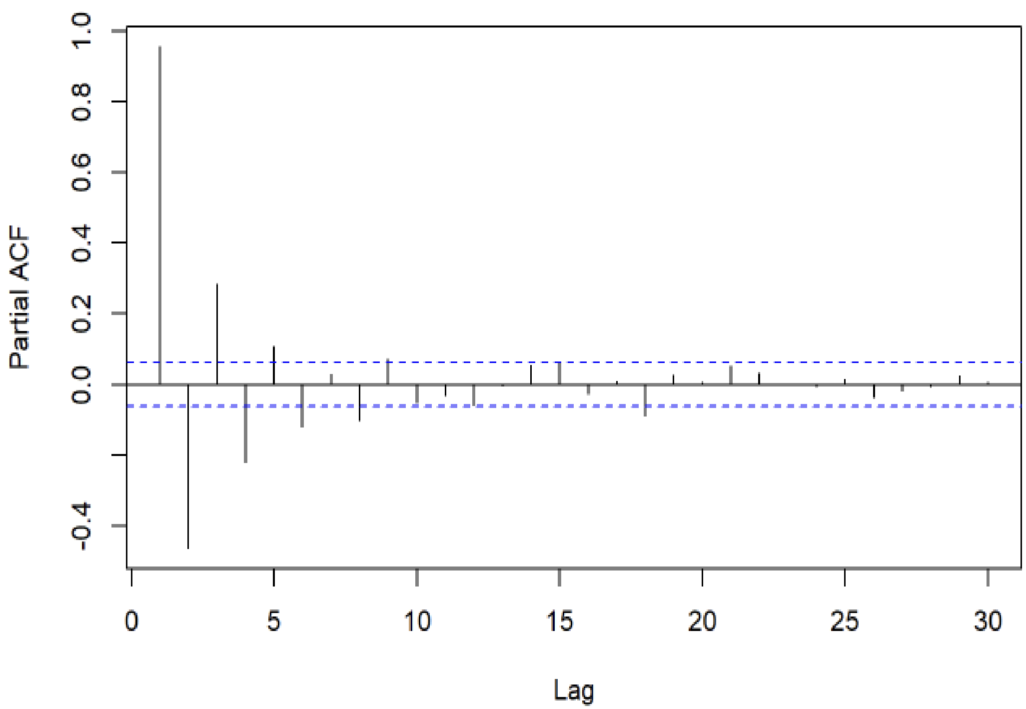
ARMA

ARMA 모형의 ACF : 지수적으로 감소 OR 싸인함수 형태로 소멸



ARMA

ARMA 모형의 PACF : 지수적으로 감소 OR 싸인함수 형태로 소멸



정리

(1) 모형의 정상성과 가역성

	AR(p)	MA(q)	ARMA(p,q)
정상성	조건필요	자체만족	조건필요
가역성	자체만족	조건필요	조건필요

정리

(2) 모형의 ACF와 PACF 패턴

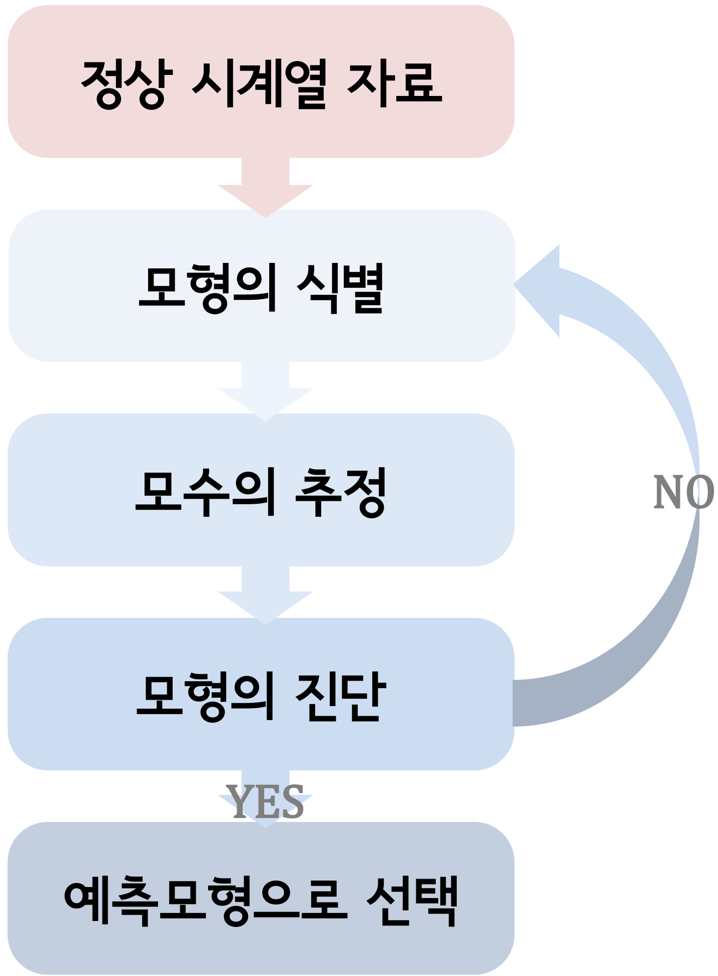
	AR(p)	MA(q)	ARMA(p,q)
ACF	지수적으로 감소	q+1차부터 절단	q+1시점부터 지수적으로 감소
PACF	p+1차부터 절단	지수적으로 감소	p+1시점부터 지수적으로 감소

4

모형 적합절차

4 모형적합 절차

[모형적합 Flow]



모형의 식별

1.

- ARMA(p,q)모형의 차수 p, q를 결정하는 단계
- ACF와 PACF를 보고 차수를 결정
- 모형 선택의 기준으로 사용되는 통계량 이용(AIC, BIC)
  : 가장 작은 AIC값을 갖는 모형 선택

모수의 추정

2.

- **최대가능도추정법** : $Z_1, Z_2, Z_3, \ldots$ 의 결합확률밀도함수인 **가능도함수(likelihood function)를 최대로** 하는 모수의 추정량을 구하는 방법

- **최소제곱추정법** : **오차제곱합을 최소로** 하는 추정법

- **적률추정법** : 모집단의 적률에 대응되는 **표본 적률의 방정식**을 풀어 추정량을 구하는 방법

모형의 진단

3.

- 잔차분석 : 잔차를 이용하여 모형에 대한 가정이 옳은지 확인

- 과대적합진단 : 잠정모형에 모수를 추가하여 더 많은 개수의 모수를 포함하는 모형을 적합

모형의

[과대적합진단]

- 추가된 모수가 유의하다고 판정
- 잠정모형의 모수추정값이 과대 적합 후의 모수추정값과 큰 차이가 존재
- 과대적합된 모형의 잔차들의 분산 < 잠정모형의 잔차들의 분산

➡ 하나 이상 충족 : 잠정모형을 새로운 모형으로 대체

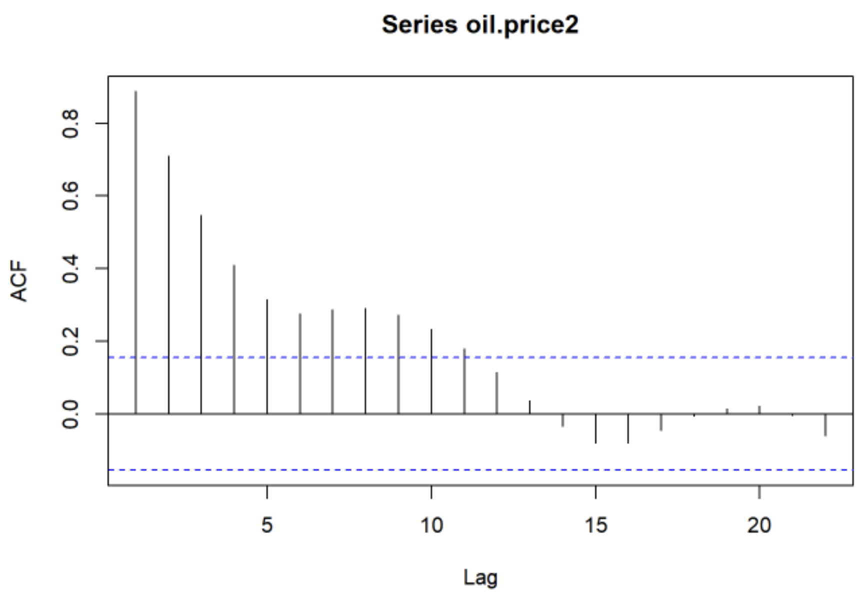
이 때, AR과 MA항을 동시에 추가하면 안됨

모형적합 절차

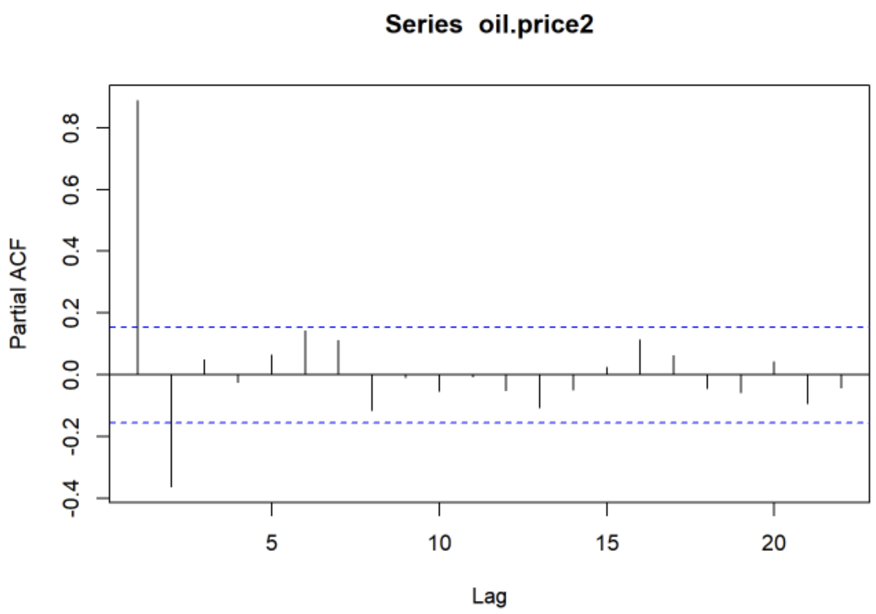
예시를 통해
알아보자!

모형의 식별

ACF



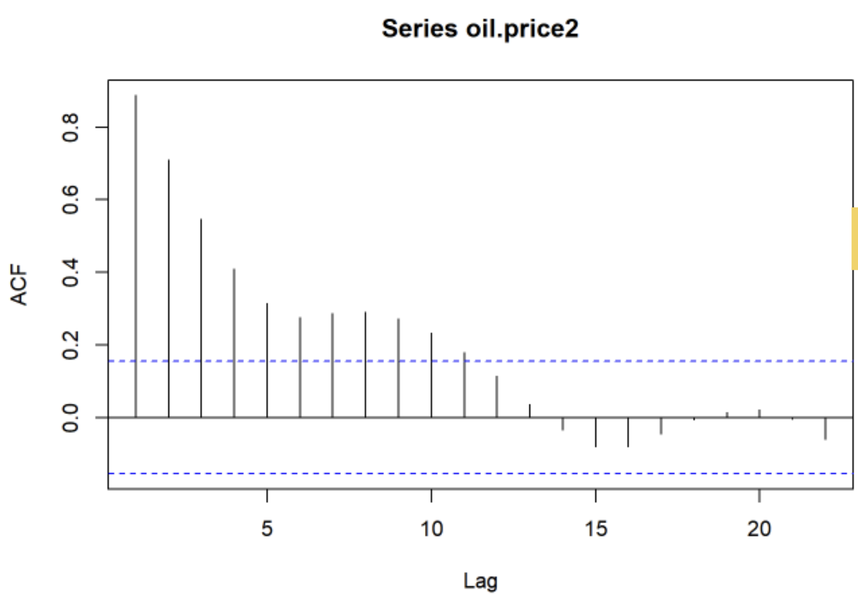
PACF



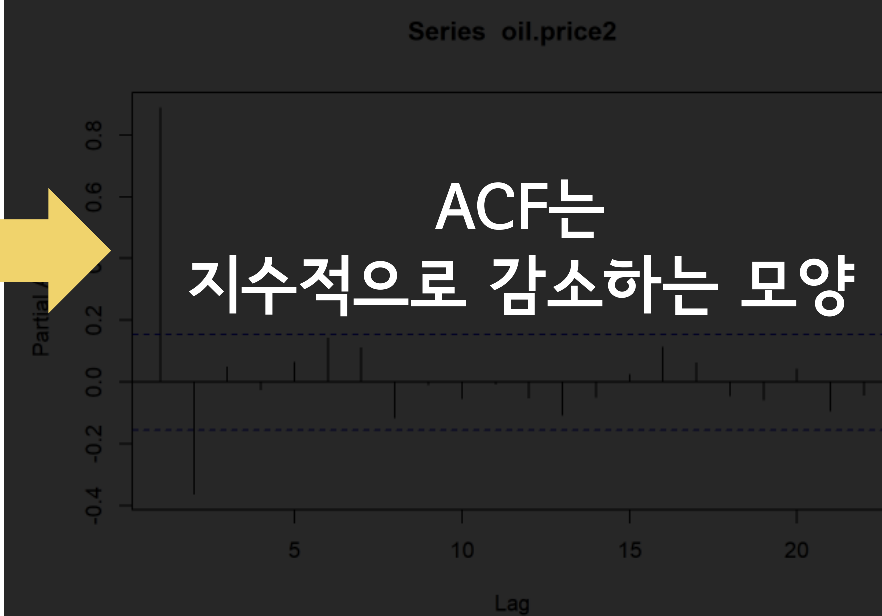
ACF와 PACF를 그려서 차수 p, q 선택

모형의 식별

ACF



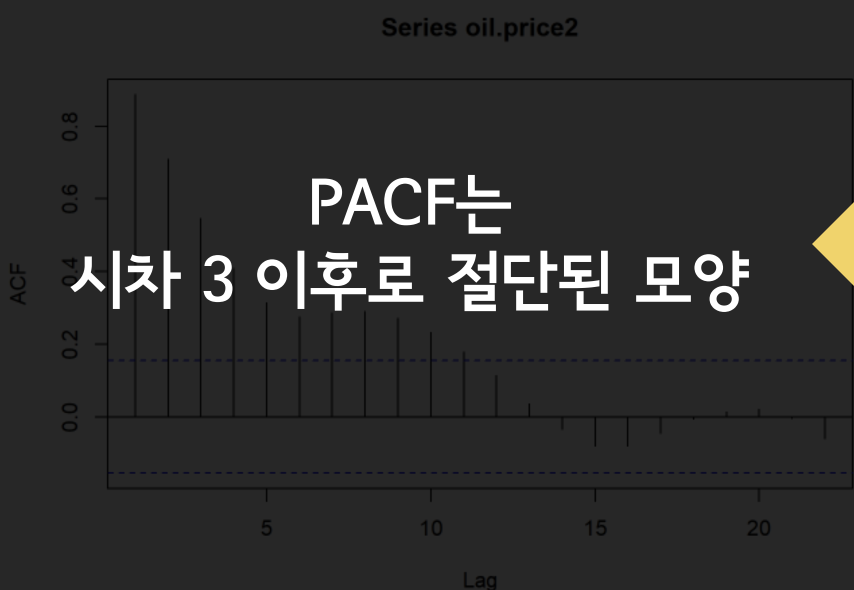
PACF



ACF는
지수적으로 감소하는 모양

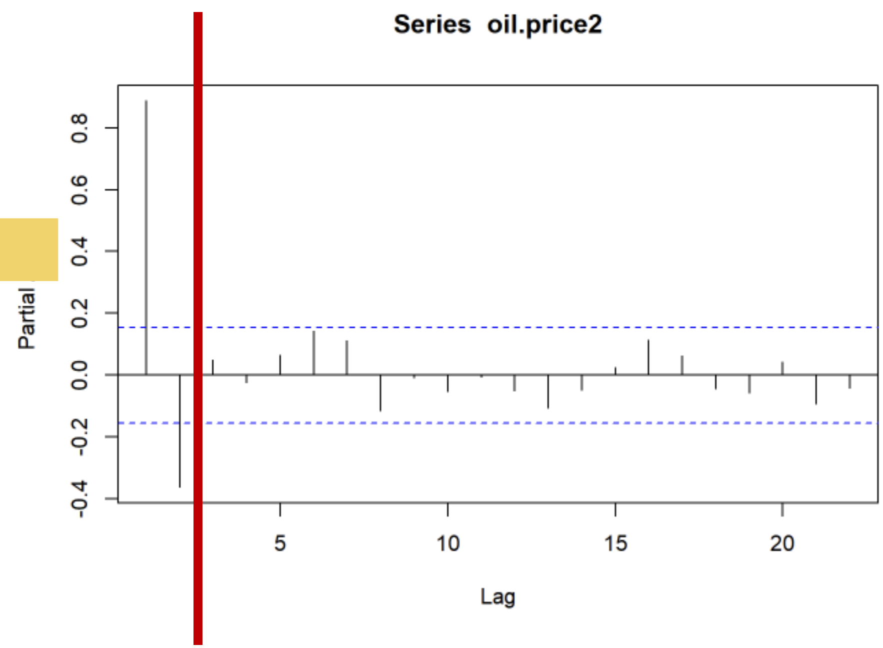
모형의 식별

ACF



PACF는
시차 3 이후로 절단된 모양

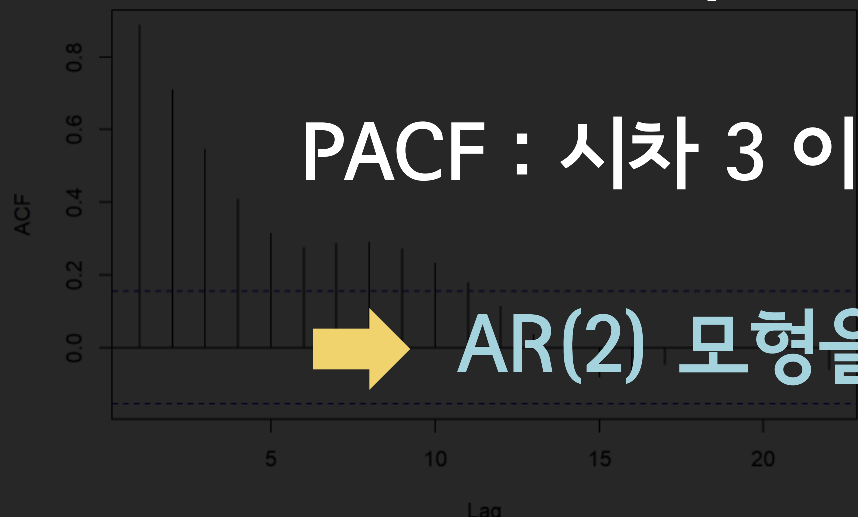
PACF



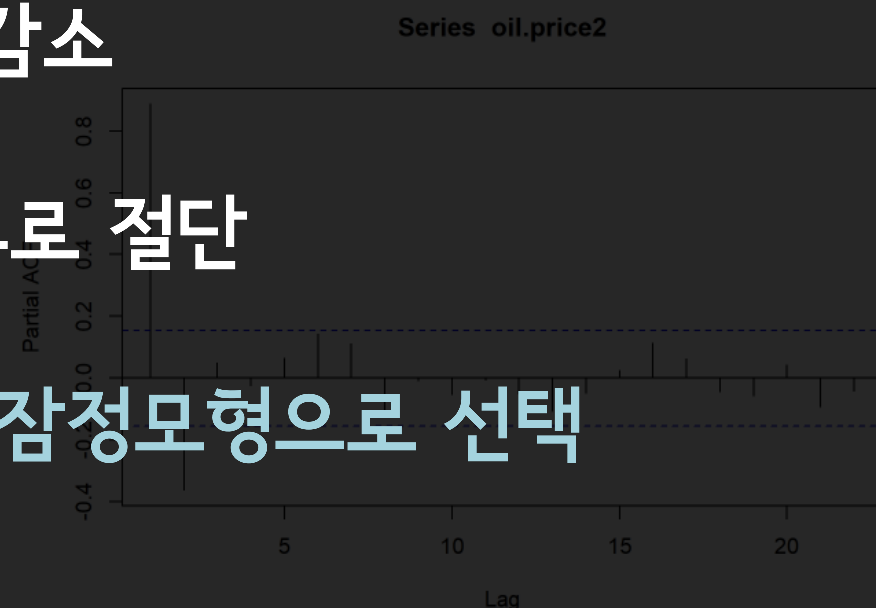
3 이후로 cut-off

모형의 식별

ACF



PACF



ACF : 지수적으로 감소

PACF : 시차 3 이후로 절단

➡ AR(2) 모형을 잠정모형으로 선택

➡ ACF와 PACF를 그려서 차수 p, q 선택

모수의 추정

```
fit<-arima(oil.price2, order=c(2,0,0))
summary(fit)
```

```
##
## Call:
## arima(x = oil.price2, order = c(2, 0, 0))
##
## Coefficients:
##          ar1      ar2  intercept
##       1.3267  -0.4835    19.1463
## s.e.  0.0746   0.0750     0.7397
##
## sigma^2 estimated as 2.205:  log likelihood = -291.35,  aic = 588.71
##
## Training set error measures:
```

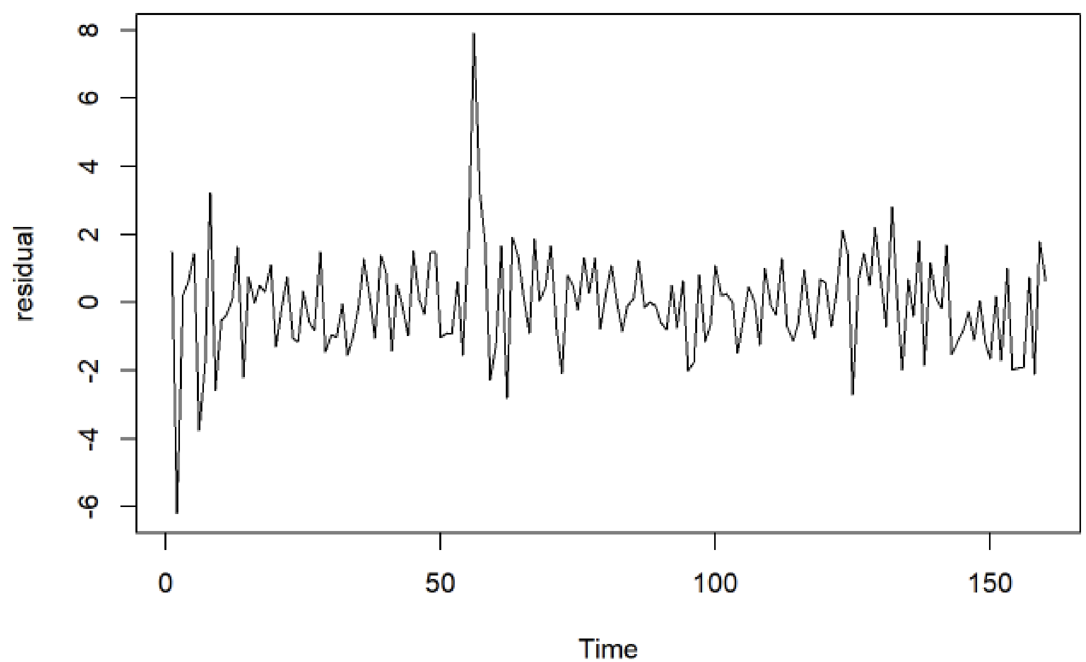
arima() 함수를 사용

order =(p,q,d)를 입력
→ p, q : ARMA 모형의 차수
d : 차분 차수(AR, MA모형은 0)

식 :
Z_t = 19.1463 + 1.3267Z_{t-1} - 0.4835Z_{t-2} + I_t

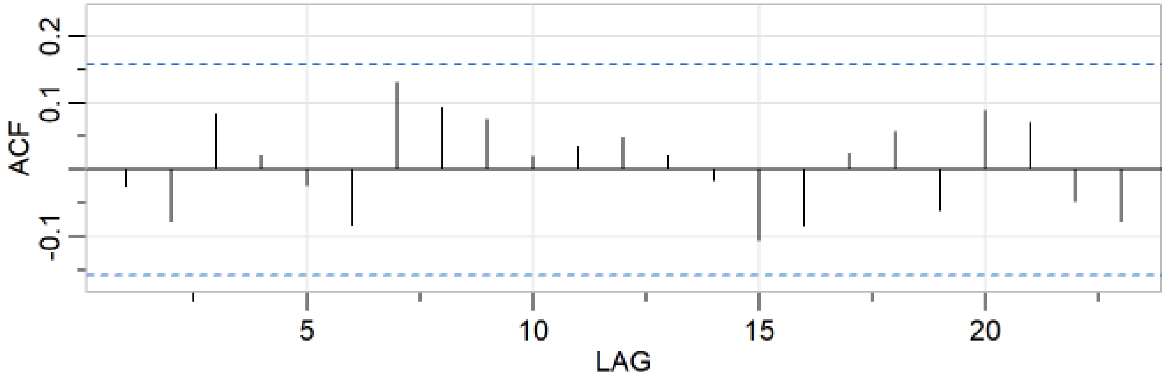
모형의 진단 - 잔차분석

잔차의 시계열 그림

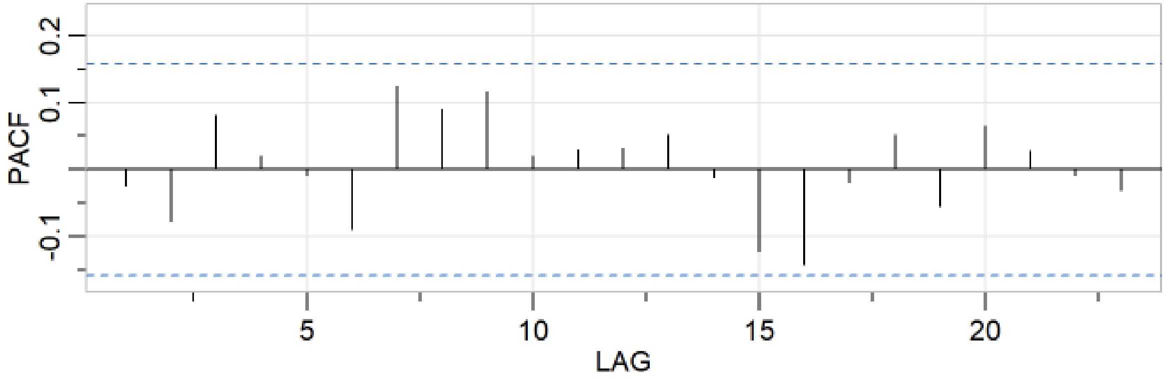


모형의 진단 - 잔차분석

잔차의 ACF



잔차의 PACF



모형의 진단 - 잔차분석

1주차에 배운
Ljung-Box 검정 사용!!

```
Box.test(resid(fit),lag=10,type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  resid(fit)
## X-squared = 9.0404, df = 10, p-value = 0.5283
```

귀무가설을 기각하지 못하므로 자기상관성이 없다고 할 수 있다.

→ 백색잡음이다

모형의 진단 - 과대적합진단

- 잠정모형으로 선택한 AR(2) (= ARMA(2,0))
- 차수 p를 +1 : AR(3)
- 차수 q를 +1 : ARMA(2,1)

→ 3개의 모형을 비교해보자

모형의 진단 - 과대적합진단

AR(2)

```
##
## z test of coefficients:
##
##           Estimate Std. Error z value  Pr(>|z|)
## ar1        1.326711   0.074592 17.7862 < 2.2e-16 ***
## ar2       -0.483522   0.074985 -6.4482 1.132e-10 ***
## intercept 19.146313   0.739656 25.8854 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

AR(3)

```
##
## z test of coefficients:
##
##           Estimate Std. Error z value  Pr(>|z|)
## ar1        1.366725   0.082171 16.6326 < 2.2e-16 ***
## ar2       -0.606097   0.130607 -4.6406 3.474e-06 ***
## ar3        0.096749   0.084588  1.1438    0.2527
## intercept 19.121889   0.805924 23.7267 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ARMA(2,1)

```
##
## z test of coefficients:
##
##           Estimate Std. Error z value  Pr(>|z|)
## ar1        1.11072    0.19741  5.6266 1.838e-08 ***
## ar2       -0.28991    0.18378 -1.5775    0.1147
## ma1        0.26864    0.20418  1.3157    0.1883
## intercept 19.11950    0.81221 23.5401 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

과대적합한 두 모형에
유의하지 않은 모수 존재

모형의 진단 - 과대적합진단

AR(2)

```
## Call:
## arima(x = oil.price2, order = c(2, 0, 0))
##
## Coefficients:
##          ar1      ar2  intercept
##       1.3267  -0.4835    19.1463
## s.e.  0.0746   0.0750     0.7397
##
## sigma^2 estimated as 2.205:  log likelihood = -291.35,  aic = 588.71
```

AR(3)

```
## Call:
## arima(x = oil.price2, order = c(3, 0, 0))
##
## Coefficients:
##          ar1      ar2     ar3  intercept
##       1.3667  -0.6061  0.0967    19.1219
## s.e.  0.0822   0.1306  0.0846     0.8059
##
## sigma^2 estimated as 2.187:  log likelihood = -290.7,  aic = 589.41
```

ARMA(2,1)

```
## Call:
## arima(x = oil.price2, order = c(2, 0, 1))
##
## Coefficients:
##          ar1      ar2     ma1  intercept
##       1.1107  -0.2899  0.2686    19.1195
## s.e.  0.1974   0.1838  0.2042     0.8122
##
## sigma^2 estimated as 2.181:  log likelihood = -290.49,  aic = 588.98
```

AR(2)의 AIC 값이 가장 작다

모형의 진단 - 과대적합진단

AR(2)

```
## Call:
## arima(x = oil.price2, order = c(2, 0, 0))
##
## Coefficients:
##          ar1      ar2  intercept
##       1.3267  -0.4835    19.1463
## s.e.  0.0746   0.0750     0.7397
##
## sigma^2 estimated as 2.205:  log likelihood = -291.35,  aic = 588.71
```

AR(3)

```
## Call:
## arima(x = oil.price2, order = c(3, 0, 0))
##
## Coefficients:
##          ar1      ar2     ar3  intercept
##       1.3667  -0.6061  0.0967    19.1219
## s.e.  0.0822   0.1306  0.0846     0.8059
##
## sigma^2 estimated as 2.187:  log likelihood = -290.7,  aic = 589.41
```

ARMA(2,1)

```
## Call:
## arima(x = oil.price2, order = c(2, 0, 1))
##
## Coefficients:
##          ar1      ar2     ma1  intercept
##       1.1107  -0.2899  0.2686    19.1195
## s.e.  0.1974   0.1838  0.2042     0.8122
##
## sigma^2 estimated as 2.181:  log likelihood = -290.49,  aic = 588.98
```

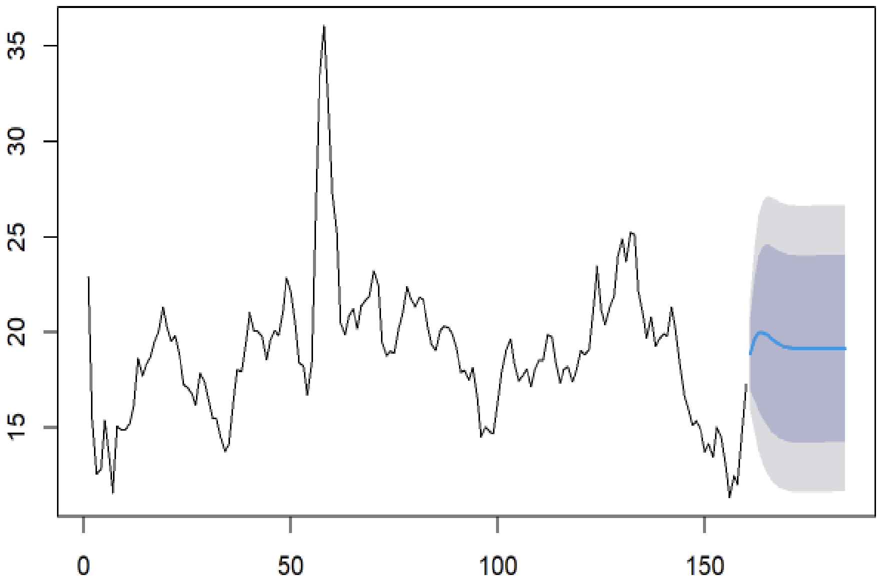
AR(2)의 AC 값이 가장 작다

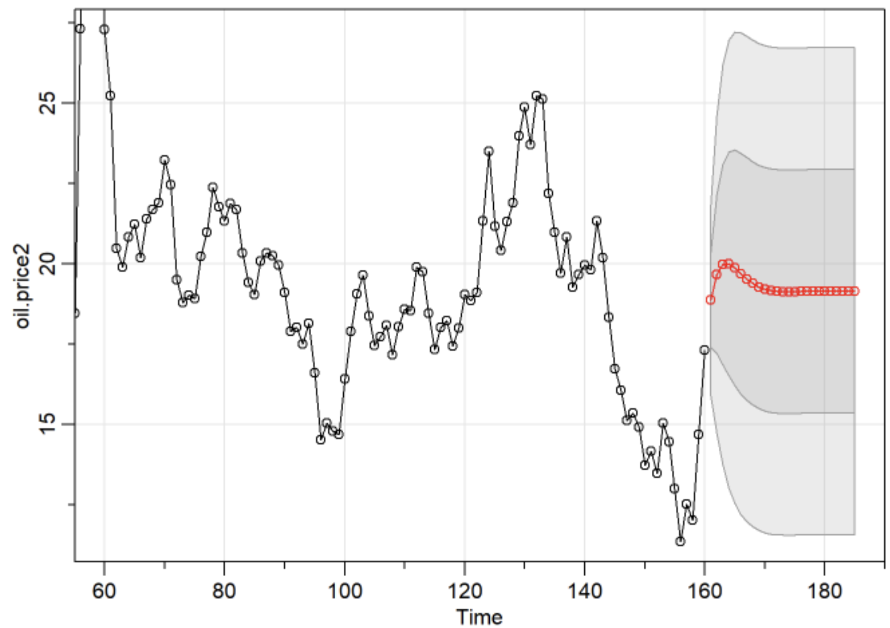
[결론]

최종 예측 모형으로 잠정모형으로 선택한 AR(2) 모형을 선택!!

미래 예측

forecast, sarima.for() 함수를 사용해서 **미래예측**





THANK YOU

3주차때 만나용 안녀어엉~~!~!~!~!