시계열 분석 기법과 응용

Week 3. ARMA 모형의 식별과 예측

3-1. ARMA모형의 식별

전치혁 교수
(포항공과대학교 산업경영공학과)

ARMA 모형의 식별

- 모형의 식별, 추정 및 검증 과정
  - 단계 1: 시계열 데이터를 그래프로 나타내어 시각적으로 관찰하고 정상성 여부를 검토한다. 비정상적인 경우에는 추세제거, 계절성제거, 분산안정화 등을 통하여 정상적 시계열로 변환한다. (이 부분은 비정상적 시계열에서 설명).
  - 단계 2. 시계열 데이터에 대한 표본 ACF 및 표본 PACF를 산출하고 정상성 여부를 확인한다. 비정상적인 경우 단계 1을 반복한다.
  - 단계 3. (모형의 식별) 표본 ACF 및 표본 PACF를 다양한 ARMA모형들의 이론적 ACF 및 PACF와 비교하여 ARMA모형의 차수인 p와 q를 구한다.
  - 단계 4. (모형의 추정) 단계 3에서 얻은 모형에 대한 계수들을 추정하고 잔차(residual)를 구한다.
  - 단계 5. (모형의 검증) 잔차가 백색잡음을 따르는지 검정한다. 잔차가 백색잡음을 따르면 단계 3의 모형이 제대로 식별되었으며, 아니면 다른 차수 p와 q를 구한 후 과정을 반복한다.

표본 ACF 및 표본 PACF

- 표본 ACF 산출 (단계 2 관련)

시계열 데이터: $Z_1, \dots, Z_n$

- 표본 분산

$$\hat{\gamma}(0) = \frac{1}{n}\sum_{t=1}^{n}(Z_i - \bar{Z})^2$$

- 시차 k 표본 자기공분산

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

- 표본 ACF

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-k}(Z_t - Z)(Z_{t+k} - Z)}{\sum_{t=1}^{n}(Z_i - Z)^2}, \quad k = 1,2, \dots$$

- 표준오차 (바틀렛 근사식 사용)

$$se(\hat{\rho}(k)) = \sqrt{\frac{1}{n}[1 + 2\hat{\rho}^2(1) + \cdots + 2\hat{\rho}^2(m)]}$$

표본 ACF 및 표본 PACF

- 표본 PACF 산출 (단계 2 관련)
  - 시차별 PACF는 ACF로 표현되므로 ACF를 추정하여 표본 PACF를 구할 수 있다.
  - Durbin (1960)이 제안한 반복적 공식 사용

$$\hat{P}(1) = \hat{\phi}_{11} = \hat{\rho}(1)$$

$$\hat{P}(k) = \hat{\phi}_{kk} = \frac{\hat{\rho}(k) - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j}\hat{\rho}(k-j)}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j}\hat{\rho}(j)}, k = 2,3,\ldots$$

여기서

$$\hat{\phi}_{k,j} = \hat{\phi}_{k-1,j} - \hat{\phi}_{k,k}\hat{\phi}_{k-1,k-j}, \mathrm{j} = 1, 2, \ldots, k-1$$

표본 ACF 및 표본 PACF

(예 3.1) 다음의 시계열 데이터에 대한 표본 ACF를 구하라.

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-k}(Z_t - Z)(Z_{t+k} - Z)}{\sum_{t=1}^{n}(Z_i - Z)^2}$$

$$\bar{Z} = -0.11$$

$$\hat{\gamma}(0) = \frac{1}{15}[(-1.01 + 0.11)^2 + (-0.81 + 0.11)^2 + \cdots + (-0.22 + 0.11)^2]$$

$$= \frac{8.6382}{15} = 0.5759$$

$$\hat{\rho}(1) = \frac{1}{8.6382}[(-1.01 + 0.11)(-0.81 + 0.11) + \cdots + (-0.41 + 0.11)(-0.22 + 0.11)]$$

$$= \frac{4.1431}{8.6382} = 0.4796$$

$$\hat{\rho}(2) = \frac{1}{8.6382}[(-1.01 + 0.11)(-0.33 + 0.11) + \cdots + (0.10 + 0.11)(-0.22 + 0.11)]$$

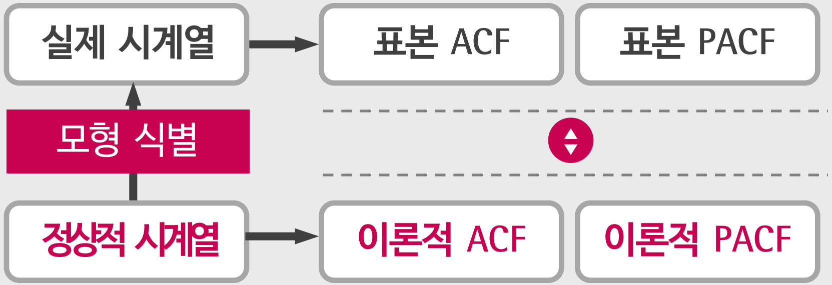
$$= \frac{1.3848}{8.6382} = 0.1603$$

	Z_t		Z_t
1	-1.01	9	1.30
2	-0.81	10	1.08
3	-0.33	11	-0.33
4	-0.40	12	0.31
5	-0.95	13	0.10
6	-1.33	14	-0.41
7	0.72	15	-0.22
8	0.63		

ARMA 모형의 식별

- ARMA모형의 이론적 ACF와 PACF의 패턴 (단계 3 관련)

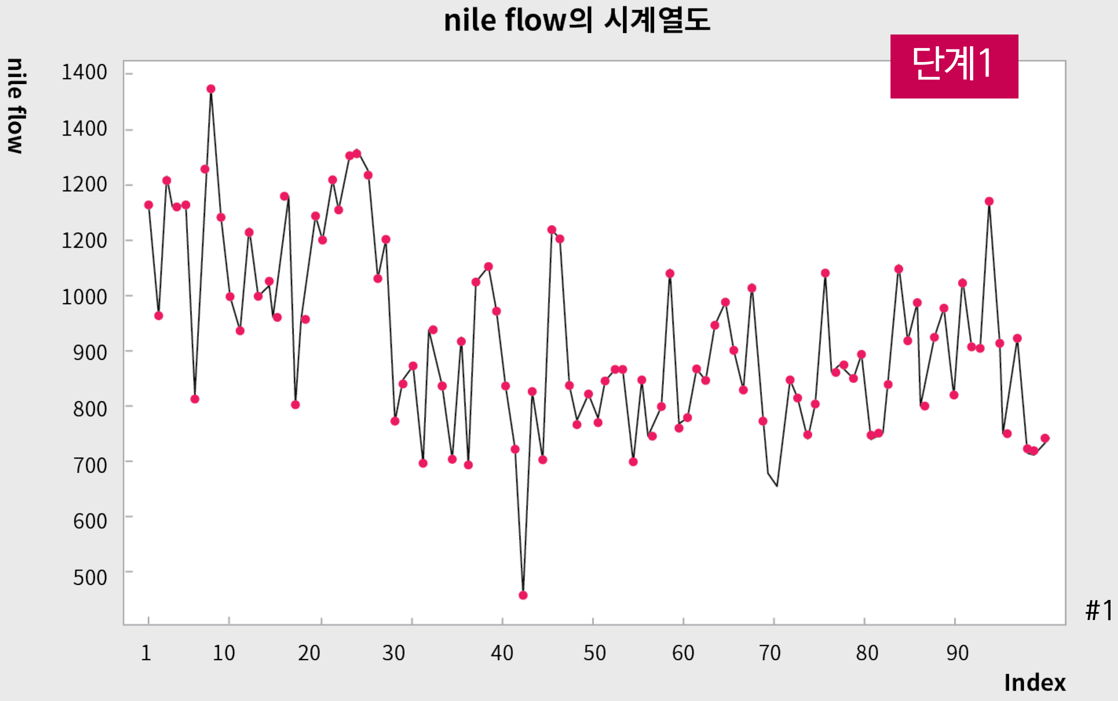
모형	ACF	PACF
AR(p)	지수적으로 감소	시차 p 이후 절단
MA(q)	시차 q 이후 절단	지수적으로 감소
ARMA(p, q)	시차 (q-p) 이후 지수적 감소	시차 (p-q) 이후 지수적 감소



ARMA 모형의 식별

예 3.2 (나일강 유량)

다음 그림은 1872년부터 1970년까지의 나일강 (Nile river) 연간평균 유량 데이터를 나타낸다. 본 데이터에 적합한 모형을 식별해 보자.

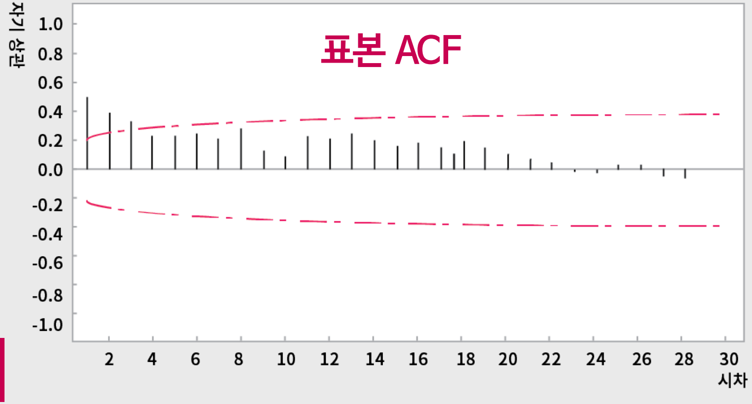


ARMA 모형의 식별

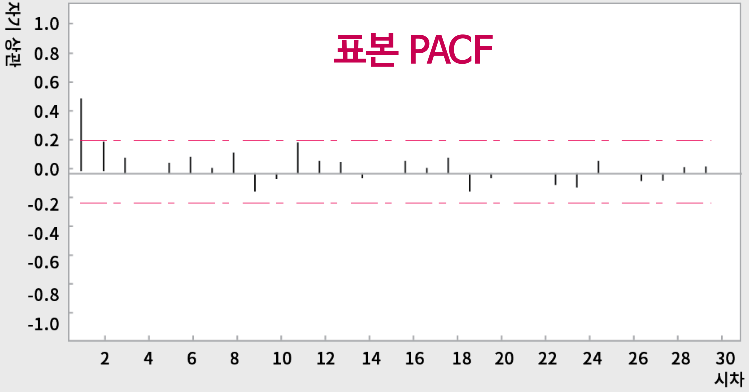
- 예 3.2 계속 (나일강 유량)

단계2

nile flow에 대한 자기 상관 함수
(자기 상관에 대한 5% 유의 한계)



nile flow에 대한 편 자기 상관 함수
(편 자기 상관에 대한 5% 유의 한계)



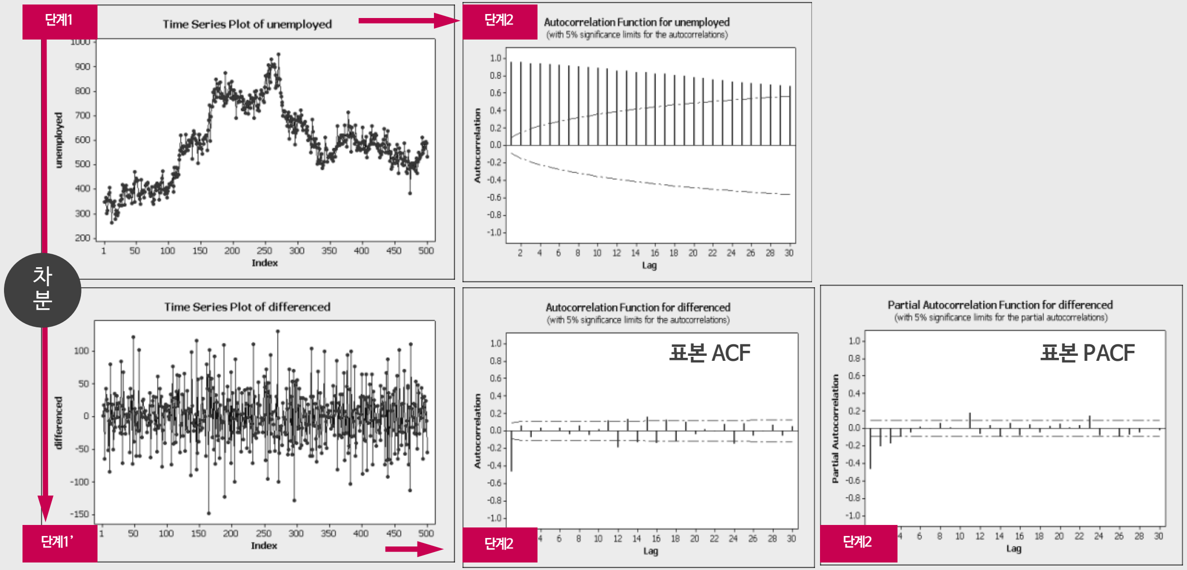
단계3

- 표본 ACF는 지수적으로 감소하고 표본 PACF는 시차 1또는 2 이후 절단되므로 AR(1) 또는 AR(2) 모형으로 보임
- 즉, p=1 또는 2; q=0

ARMA 모형의 식별

- 예 3.3(미국 여성 실업자수)

  \- 다음 그림은 1961-2002년 사이 월별 미국의 16-19세 젊은 여성 실업자수를 나타낸다.



단계3'

\- ACF는 시차 1이후 절단되고 PACF는 지수적으로 감소하므로 MA(1)모형

Reference

#1. Qlik DataMarket https://datamarket.com/data/set/22w8/mean-annual-nile-flow-1871-1970#!ds=22w8&display=line 2019.12

#2. W.W.S. Wei (2006), Time Series Analysis: Univariate and Multivariate Methods 2nd ed.