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) (정의) * 시계열 교재 def. 1.4.4.

ACF

$\gamma(0) = Var[Z_t]$

$\gamma(k) = E[Z_t Z_{t-k}]$

$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$

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

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

$k<n$

- 표본 ACF

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

- 표준오차 (바틀렛 근사식 사용)

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

$\hat{\rho} \approx N(\rho, n^{-1}W)$ * 교재 53페이지

W: covariance matrix

$w_{ij} = \sum_{k=-\infty}^{\infty} \{\rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(i)\rho(k)\rho(k+i)\}$

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

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

→ 복잡해서 소프트웨어 사용계산

표본 ACF 및 표본 PACF

$\hat{\gamma}(0) = \frac{8.6382}{15}$ $\hat{\gamma}(k) = \frac{1}{15}\sum_{t=1}^{n-k}(Z_t-\bar{Z})(Z_{t+k}-\bar{Z})$

(예 3.1) 다음의 시계열 데이터에 대한 표본 ACF를 구하라.

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

$\bar{Z} = -0.11$

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

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

$\hat{\rho}(1) = \frac{1}{8.6382}[(\underset{1}{\underline{-1.01}}+0.11)(\underset{2}{\underline{-0.81}}+0.11) + \cdots + (\underset{14}{\underline{-0.41}}+0.11)(\underset{15}{\underline{-0.22}}+0.11)]$

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

$\hat{\rho}(2)$

$= \frac{1}{8.6382}[(\underset{1}{\underline{-1.01}}+0.11)(\underset{3}{\underline{-0.33}}+0.11) + \cdots + (\underset{13}{\underline{0.10}}+0.11)(\underset{15}{\underline{-0.22}}+0.11)]$

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

	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)

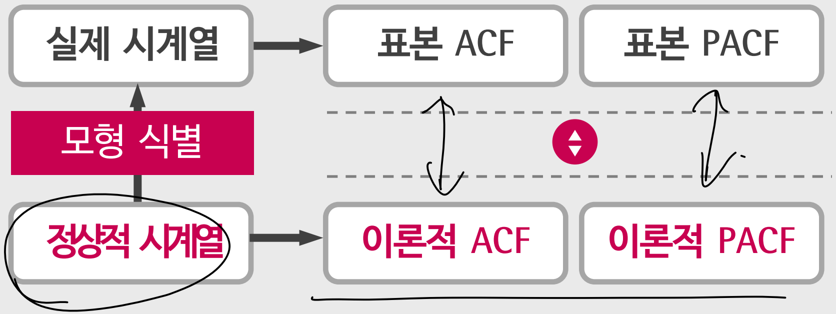
정상적인 시계열에서 이론적 ACF, PACF 알 수 있음.

실제 시계열에서 표본 ACF, PACF 계산

⇓

어떤 모형과 가까운지 판단

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

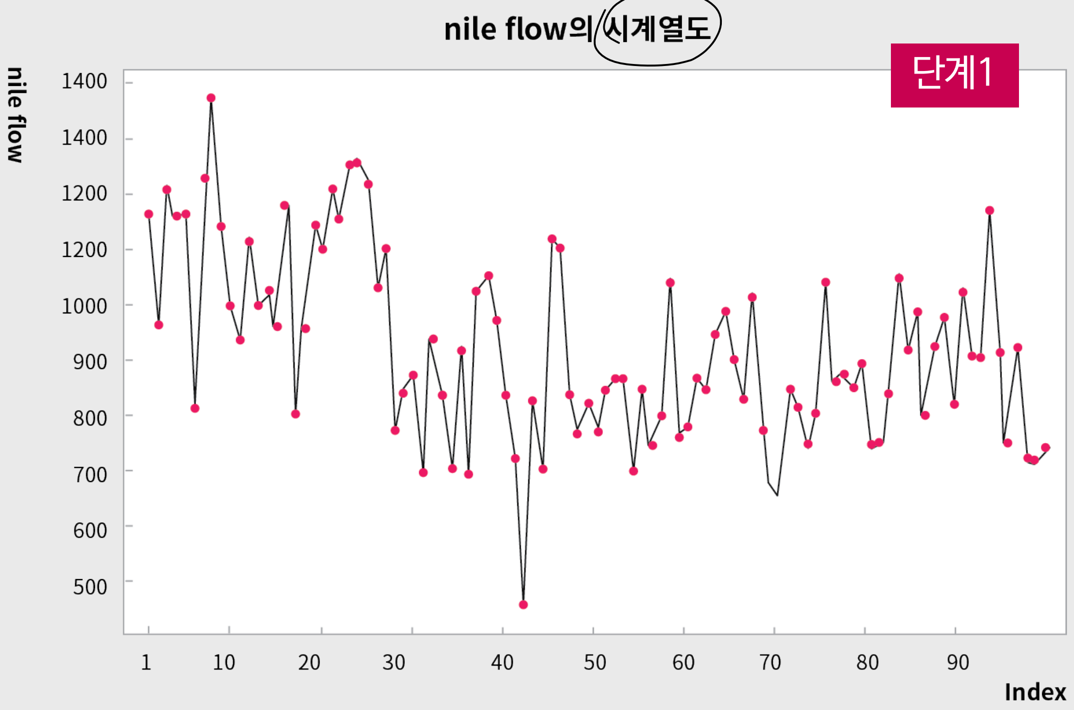


ARMA 모형의 식별

추세나 계절성이 있는지 판별.

예 3.2 (나일강 유량)

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



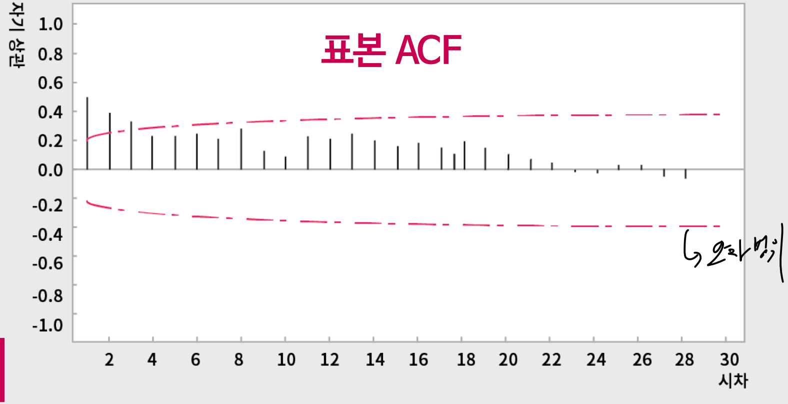
#1

ARMA 모형의 식별

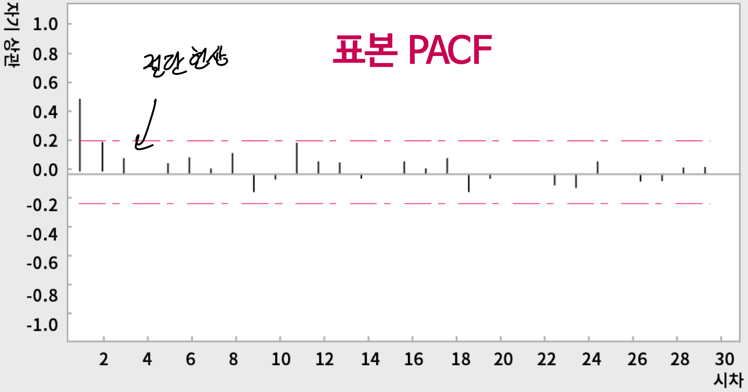
- 예 3.2 계속 (나일강 유량)

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

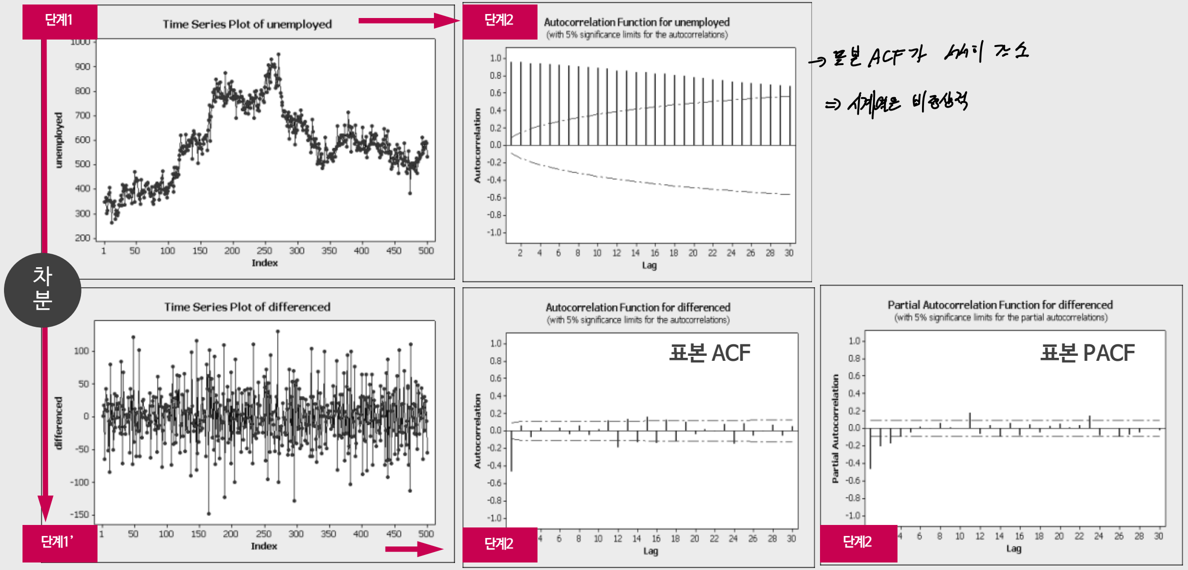
- 표본 ACF는 지수적으로 감소하고 표본 PACF는 시차 1또는 2 이후 절단되므로 AR(1) 또는 AR(2) 모형으로 보임
- 즉, p=1 또는 2; q=0

#1

ARMA 모형의 식별

- 예 3.3(미국 여성 실업자수)

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



⇒ 표본 ACF가 서서히 감소

⇒ 시계열은 비정상적

단계3'

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

※ 차분: t 시점과 t-1 시점의 데이터가 존재할 때 그 차이

원 시계열 대신 차이 계산하여 사용

시계열 모형 추정방법

- 최소자승법 (least squares method): AR모형의 경우 가능
- 비선형 최소자승법 (nonlinear least squares method): ARMA모형에 적용
- 최우 추정법 (maximum likelihood estimation)
  - 오차항이 서로 독립인 정규분포를 따르므로 우도함수 (likelihood function)을 유도하여 이를 최대로 하는 모형 계수들을 추정. ARMA 모형의 경우 정확한 우도함수 도출이 어렵고 초기치에 대한 가정이 필요.
  - 정확한 우도함수
  - 조건있는 우도함수: 임의로 초기치 가정하여 사용
  - 조건없는 우도함수: 과거의 초기치를 후방예측 (backcasting)하여 사용

최우추정법

- 정확한 우도함수
  - 평균 μ, 분산 σ^2의 정규분포를 따르는 확률변수 X에 대한 관측치 $(x_1, \ldots, x_n)$가 있다하자.
  - 이 때 관측치는 서로 독립이므로 n개 관측치의 결합확률분포는 정규확률밀도함수의 곱으로 표현됨

$$f(x_1, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

  - 관측 데이터를 대입하면 모수 μ와 σ^2의 함수로 간주할 수 있으며 이를 우도함수라 함.

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}}$$

  - 이 우도함수를 최대로 하는 모수 μ와 σ^2를 추정하는 방법을 최우추정법이라 함.
  - 로그함수를 취한 로그우도함수를 최대로 하는 것이 편함

$$logL(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x_i-\mu)^2}{-2\sigma^2}} = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \cdot e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}}$$

$$\ln L(\mu,\sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{\sum(x_i-\mu)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial\mu}\ln L = \frac{2\sum(x_i-\mu)}{2\sigma^2} = \frac{1}{\sigma^2}\sum(x_i-\mu) = \frac{n}{\sigma^2}(\bar{x}-\mu) = 0 \qquad \mu = \bar{x}$$

$$\frac{\partial}{\partial\sigma^2}\ln L = -\frac{n}{2\sigma^2} + \frac{\sum(x_i-\mu)^2}{2\sigma^4} = 0 \qquad \sigma^2 = \frac{1}{n}\sum(x_i-\mu)^2$$

※ $\frac{\partial^2}{\partial\mu\partial\mu}\ln L$, $\frac{\partial^2}{\partial\mu\partial\sigma^2}\ln L$, $\frac{\partial^2}{\partial\sigma^2\partial\sigma^2}\ln L$ 의 부호 확인

최우추정법

- ARMA모형의 경우 관측치 $(Z_1, Z_2, \dots, Z_n)$들은 서로 독립이 아니므로 이에 대한 우도함수의 구성이 어렵다.
- 대신 백색잡음 $(a_1, a_2, \dots, a_n)$이 서로 독립임을 활용하여 이에 대한 우도함수를 구성한다.

$$L(\theta; a_1, \dots, a_n) = (2\pi\sigma_a^2)^{-n/2} exp\left(-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n} a_t^2\right)$$

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} exp\left(-\frac{\Sigma(x_i-\mu)^2}{2\sigma^2}\right)$$

- (예 – AR(1) 모형) $Z_t = \phi_1 Z_{t-1} + a_t$
  - $a_t = Z_t - \phi_1 Z_{t-1}$
  - $a_1 = Z_1 - \phi_1 Z_0$ (즉, a_1을 구하기 위해 초기치 Z_0 필요)
- 초기치 처리 방법
  - 적절한 가정으로 초기치 결정 ⇒ 조건있는 우도함수 (conditional likelihood function)
  - 초기치를 예측하여 사용 ⇒ 조건없는 우도함수 (unconditional likelihood function)

조건있는 우도함수

- $(a_1, a_2, \ldots, a_n)$의 로그 우도함수

$$logL(\theta|a_1, \ldots, a_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}log\sigma_a^2 - \frac{\sum_{t=1}^{n} a_t^2}{2\sigma_a^2}$$

- ARMA (p,q) 모형에서 $(a_1, a_2, \ldots, a_n)$ 산출

$$a_t = \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + Z_t - \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p}$$
$$a_1 = \theta_1 a_0 + \cdots + \theta_q a_{1-q} + Z_1 - \phi_1 Z_0 - \cdots - \phi_p Z_{1-p}$$

다음의 초기치 필요

$$a_* = (a_{1-q}, \ldots, a_{-1}, a_0) = (0, \ldots, 0, 0)$$
$$z_* = (z_{1-p}, \ldots, z_{-1}, z_0) = (\bar{z}, \ldots, \bar{z}, \bar{z})$$

초기 조건하에서 로그 우도함수

$$logL_*(\theta|a_1, \ldots, a_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}log\sigma_a^2 - \frac{S_*(\theta)}{2\sigma_a^2}$$

조건제곱합

$$S_*(\theta) = \sum_{t=1}^{n} a^2{}_t\,(\theta|a_*, z_*, z_1, \ldots, z_n) \Rightarrow \text{ 최소화} \Rightarrow \boxed{\hat{\theta} \Rightarrow \hat{\sigma}_a^2 = \frac{S_*(\hat{\theta})}{df}}\ (df = n - 2p - q - 1)$$

조건있는 우도함수

(예- AR(1) 모형) 다음 10개 시계열 데이터가 AR(1)모형을 따른다.

−0.17, 0.24, −0.51, 1.10, 0.08, 0.32, −0.44, −1.16, −0.58, −0.28

$\phi_1 = 0.2$와 $\phi_1 = 0.4$ 일때 조건제곱합을 각각 구하라.

(풀이)

$\phi_1 = 0.2$: $a_t = Z_t - 0.2Z_{t-1}$

$\phi_1 = 0.4$: $a_t = Z_t - 0.4Z_{t-1}$

$z_0 = 0$ 가정

$$S_*(\phi_1 = 0.2) = 3.5234$$

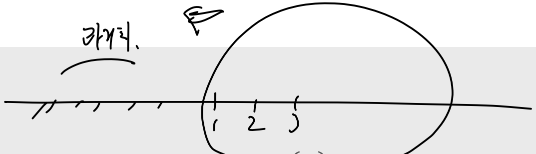
√

$$S_*(\phi_1 = 0.4) = 3.7106$$

최적화 알고리즘을 통해 제곱합이 최소화하거나 우도함수를 최대화하는 ϕ를 찾아나감.

Z_t	$\phi_1 = 0.2$		$\phi_1 = 0.4$	
	a_t	a_t^2	a_t	a_t^2
-0.17	-0.17	0.0289	-0.17	0.0289
0.24	0.274	0.0751	0.308	0.0949
-0.51	-0.558	0.3114	-0.606	0.3672
1.1	1.202	1.4448	1.304	1.7004
0.08	-0.14	0.0196	-0.36	0.1296
0.32	0.304	0.0924	0.288	0.0829
-0.44	-0.504	0.2540	-0.568	0.3226
-1.16	-1.072	1.1492	-0.984	0.9683
-0.58	-0.348	0.1211	-0.116	0.0135
-0.28	-0.164	0.0269	-0.048	0.0023

조건없는 우도함수



- 조건없는 우도함수 → 필요한 과거치는 예측

$$logL(\theta|z_1,\ldots,z_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}log\sigma_a^2 - \frac{S(\theta)}{2\sigma_a^2}$$

- 조건없는 제곱합

  $S(\theta) = \sum_{t=-M}^{n} E^2[a_t|\theta; z_1,\ldots,z_n]$ (M: 과거치 예측수)

  백색잡음의 기대치로 오차항 예측 - 이 때 과거 관측치 예측 필요

- 과거값 예측에 후방예측 (backcasting) 사용 - 시간축을 반대방향으로 생각 ⇒ 과거치를 미래치로 생각.

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

$$\Rightarrow Z_t = \phi_1 Z_{t+1} - \cdots - \phi_p Z_{t+p} + b_t - \theta_1 b_{t+1} - \cdots - \theta_q b_{t+q}$$

⇒ 미래값 (관측치)으로 과거값 예측

- 예 –AR(1)모형 $Z_t = \phi_1 Z_{t-1} + a_t \Leftrightarrow Z_t = \phi_1 Z_{t+1} + b_t$

$$E[a_0] = E[Z_0] - \phi_1 E[Z_{-1}]$$

$$E[Z_0] = \phi_1 E[Z_1] + E[b_0] = \phi_1 z_1$$

$$E[Z_{-1}] = \phi_1 E[Z_0] + E[b_{-1}] = \phi_1 \phi_1 z_1$$

조건없는 우도함수

(예- AR(1))다음 10개 시계열 데이터가 AR(1)모형을 따른다.

-0.17, 0.24, -0.51, 1.10, 0.08, 0.32, -0.44, -1.16, -0.58, -0.28

$\phi_1 = 0.2$와 $\phi_1 = 0.4$ 일때 조건없는 제곱합을 각각 구하라.

$S(\theta) = \sum_{t=-M}^{n} E^2[a_t|\theta; z_1, \dots, z_n]$

- $Z_t = \phi_1 Z_{t-1} + a_t \Leftrightarrow Z_t = \phi_1 Z_{t+1} + b_t$
- $E[a_t] = Z_t - \phi_1 Z_{t-1},\ t = 2,3,\dots$
- $E[a_1] = E[Z_1] - \phi_1 E[Z_0] = (1-\phi_1^2)z_1$
- $E[a_{-j}] = \phi_1^{j+1}(1-\phi_1^2)z_1$, j= 0,1,2, …
- $S(\phi_1 = 0.2) = 3.5236$
- $S(\phi_1 = 0.4) = 3.7060$

Z_t		$\phi_1 = 0.2$		$\phi_1 = 0.4$	
		a_t	a_t^2	a_t	a_t^2
-3	-	-0.0040	0.0000	-0.00366	0.0000
-2	-	-0.0100	0.0001	-0.00914	0.0001
-1	-	-0.0250	0.0006	-0.02285	0.0005
0	-	-0.0626	0.0039	-0.05712	0.0033
1	-0.17	-0.1564	0.0245	-0.1428	0.0204
2	0.24	0.2740	0.0751	0.308	0.0949
3	-0.51	-0.5580	0.3114	-0.606	0.3672
4	1.1	1.2020	1.4448	1.304	1.7004
5	0.08	-0.1400	0.0196	-0.36	0.1296
6	0.32	0.3040	0.0924	0.288	0.0829
7	-0.44	-0.5040	0.2540	-0.568	0.3226
8	-1.16	-1.0720	1.1492	-0.984	0.9683
9	-0.58	-0.3480	0.1211	-0.116	0.0135
10	-0.28	-0.1640	0.0269	-0.048	0.0023

$\sum$ $S(\phi=0.2)$

$\sum$ $S(\phi=0.4)$

ARMA 모형의 추정

- 예 (나일강 유량) 본 시계열이 AR(2)모형을 따른다고 식별되므로 다음의 모형 추정 $Z_t = c + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$ 추정 결과는 다음과 같다.

	추정치	표준오차	t-값	p-값
상수 (c)	357.5	14.55	24.56	0.000
AR1 (ϕ_1)	0.4039	0.1000	4.04	0.000
AR2 (ϕ_2)	0.2064	0.1006	2.05	0.043

#1
유의한지 확인

- 두 계수 ϕ_1, ϕ_2의 추정치는 유의수준 5%에서 유의
- 추정식은 다음과 같다. $\hat{Z}_t = 357.5 + 0.4039 Z_{t-1} + 0.2064 Z_{t-2}$

$\boxed{\widehat{\sigma_a^2} = MSE = 20943}$

ARMA 모형의 검증

- ## 모형의 검증
  - 모형의 오차항이 평균 0, 분산 σ_a^2의 정규분포를 따르는 $\boxed{\text{백색잡음}}$이라 가정하고 있기 때문에 이에 대한 검증이 필요
  - 잔차에 대하여 다음을 확인
    - 정규성: 정규확률도표 활용
    - 등분산성 및 패턴 유무: 잔차 산점도 활용
    - 랜덤성:

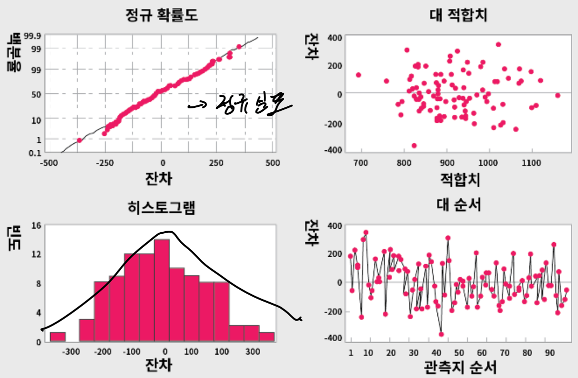
      잔차가 랜덤성을 가지면 모든 시차에 대해 표본 ACF, PACF가 0
      - ✓ ACF/PACF로 시차별 상관계수가 모두 0인지 확인
      - ✓ 포트만토 검정 (portmanteau test): 시차별 자기상관계수가 모두 0인지 검정 (Ljung-Box 검정 등 사용)

        ※ 교재 page 30.

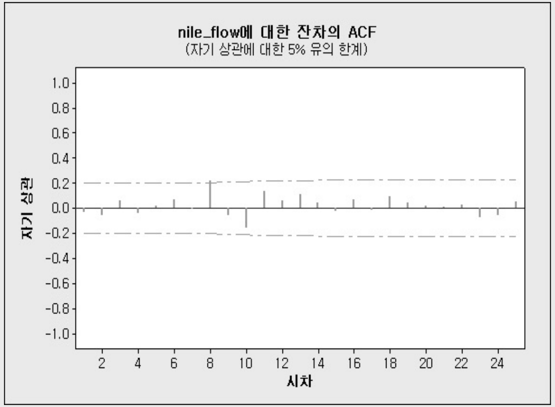
ARMA 모형의 식별과 추정

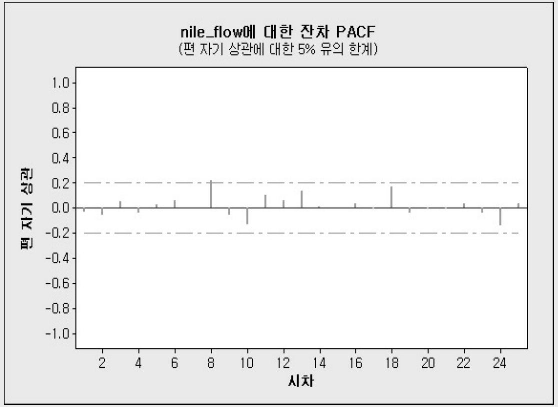
- 예 (나일강 유량)

산점도 → 패턴 X



nile_flow의 잔차 그림





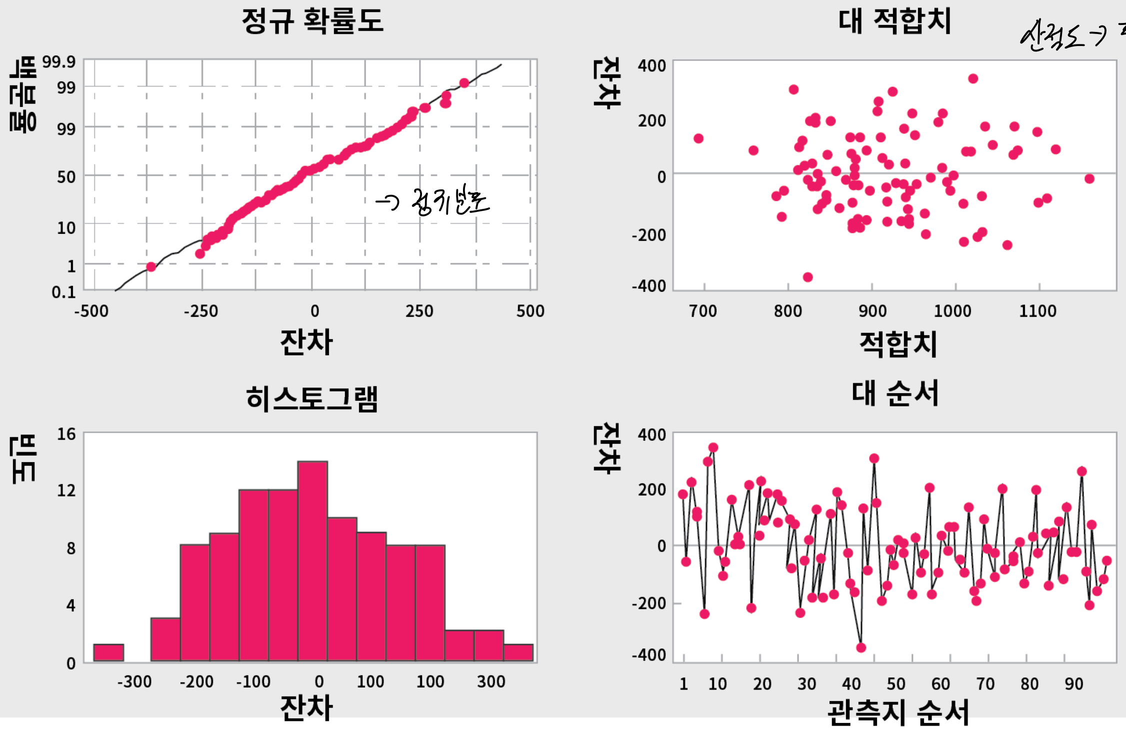
#1

ARMA 모형의 식별과 추정

- 예 (나일강 유량)
  - 정규확률도를 볼 때 잔차가 정규분포를 따름
  - 잔차 산점도로 부터 등분산성, 패턴없음이 관찰

nile_flow의 잔차 그림

산점도 → 패턴 X

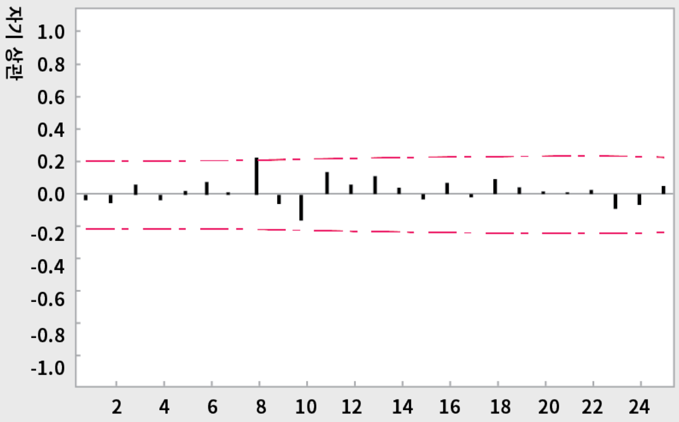


ARMA 모형의 식별과 추정

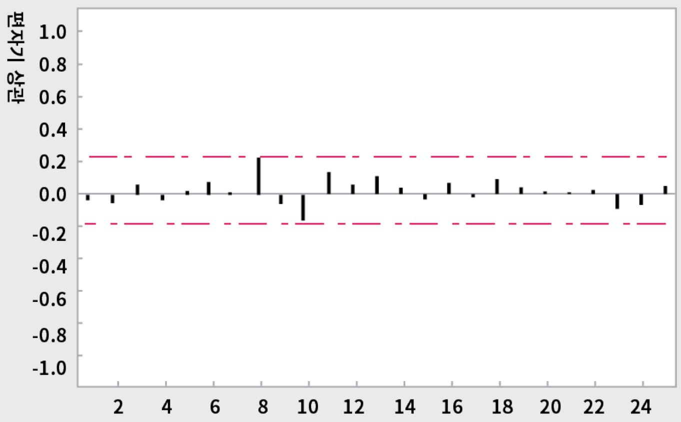
- 예 (나일강 유량)
  - 잔차에 대한 ACF/PACF로 부터 백색잡음 확인

오차범위 내

nile_flow에 대한 잔차의 ACF
(자기 상관에 대한 5% 유의 한계)



nile_flow에 대한 잔차의 PACF
(편 자기 상관에 대한 5% 유의 한계)



#1

ARMA 모형의 식별과 추정

- 포트만토 (Ljung-Box) 검정

시차	12	24	36	48
카이제곱	11.9	16.6	25.0	37.3
자유도	9	21	33	45
p-값	0.222	0.735	0.841	0.787

p값이 크다 → 유의하지 않다.

- 포트만토 검정에서 시차별 자기상관이 없음의 귀무가설을 채택
- 즉, 잔차에 시차별 자기상관이 없다고 판정

자기상관이 없다는 귀무가설 기각 못함.

시계열 예측

모형을 식별하고 모수 추정 후 미래값 예측

최소 평균제곱오차 예측치

- 관측치: $Z_1, \dots, Z_n$
  - $f_{n,k}$ (k =1,2,⋯): 시점 n에서 k시점 이후 (즉, $n+k$ 시점) 시계열 예측치
  - 과거 시계열 관측치의 선형결합으로 예측

$$f_{n,k} = b_k Z_n + b_{k+1} Z_{n-1} +$$

$$\Downarrow \; Z_t = a_t - \psi_1 a_{t-1} - \psi_2 a_{t-2} - \cdots$$ (MA 표현 방법)

$$f_{n,k} = c_k a_n + c_{k+1} a_{n-1} + \cdots$$

- 평균제곱오차 (mean squared error)

$$Q = E\left[\left(Z_{n+k} - f_{n,k}\right)^2\right] = E\left[\left(\sum_{j=0}^{\infty} \psi_j a_{n+k-j} - \sum_{j=k}^{\infty} c_j a_{n+k-j}\right)^2\right]$$

$$= \sigma_a^2 \sum_{j=0}^{k-1} \psi_j^2 + \sigma_a^2 \sum_{j=0}^{\infty} \left(\psi_{j+k} - c_{j+k}\right)^2$$

↳최소

시계열 예측

최소 평균제곱오차 예측치

- Q를 최소로 하는 c_{j+k}의 추정치

  ☆ 시점 n 까지의 관측치가 조건으로 주어졌을 때 해당 예측값은 시점 $n+k$ 에서 시계열에 대한 기댓값이 됨.

  $\hat{c}_{j+k} = \psi_{j+k},\ k = 1,2,\ldots$

  $f_{n,k} = \psi_k a_n + \psi_{k+1} a_{n-1} + \cdots$

  $f_{n,k} = E[Z_{n+k} | Z_n, Z_{n-1}, \ldots]$: k시점 이후 예측치는 조건부 기대치와 동일

- 예측오차

  $e_{n,k} = Z_{n+k} - f_{n,k} = a_{n+k} + \psi_1 a_{n+k-1} + \cdots + \psi_{k-1} a_{n+1},\ k = 1,2,\ldots$

  예측오차 분산

  $v_{n,k} = \boxed{Var[e_{n,k}] = \sigma_a^2 \sum_{j=0}^{k-1} \psi_j^2}$

  $v_{n,k} = Var[Z_{n+k} | Z_n, Z_{n-1}, \ldots]$: k시점 이후 예측오차분산은 조건부 분산과 동일

ARMA모형 예측

- 예측식: 시점 n까지의 과거 데이터로 부터 미래 시점 예측은 조건부 기대치 사용
  - 한단계 예측식: $f_{(n,1)} = E[Z_{(n+1)} \mid Z_n, Z_{(n-1)}, \cdots]$
  - k-단계 예측식: $f_{(n,k)} = E[Z_{(n+k)} \mid Z_n, Z_{(n-1)}, \cdots]$, $k = 1, 2, \cdots$
  - 계수는 추정값 이용, 과거 오차항은 잔차를 이용하여 추정
- 예측오차 분산: 예측구간이 필요한 경우 사용
  - 한단계 예측오차: $e_{(n,1)} = Z_{(n+1)} - f_{(n,1)} = a_{(n+1)}$
  - k-단계 예측오차: $e_{(n,k)} = Z_{(n+k)} - f_{(n,k)}$
  - 한단계 예측오차 분산: $Var[e_{(n,1)}] = Var[a_{(n+1)}] = \sigma_a^2$ σ_a^2
  - k-단계 예측오차 분산: $Var[e_{(n,k)}] = Var[Z_{(n+k)} \mid Z_n, Z_{(n-1)}, \cdots]$, $k = 1, 2, \cdots$

ARMA모형 예측

AR모형의 예측

- AR(2) 모형: $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$
  - 한단계 예측식: $f_{n,1} = E[Z_{n+1} | Z_n, Z_{n-1}, \dots] = \phi_1 Z_n + \phi_2 Z_{n-1}$ (a_{n+1}의 기대치는 0)
  - 2-단계 예측식: $f_{n,2} = E[Z_{n+2} | Z_n, Z_{n-1}, \dots] = \phi_1 f_{n,1} + \phi_2 Z_n$
  - 한단계 예측오차 분산: $\mathrm{Var}[e_{n,1}] = \sigma_a^2$
  - 2-단계 예측오차 분산: $\mathrm{Var}[e_{n,2}] = \mathrm{Var}[Z_{n+2} | Z_n, Z_{n-1}, \dots] = (1 + \phi_1^2)\sigma_a^2$

- AR(4) 모형: $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_4 Z_{t-4} + a_t$
  - 한단계 예측식: $f_{n,1} = E[Z_{n+1} | Z_n, Z_{n-1}, \dots] = \phi_1 Z_n + \phi_2 Z_{n-1} + \cdots + \phi_4 Z_{n-3}$
  - 2-단계 예측식: $f_{n,2} = E[Z_{n+2} | Z_n, Z_{n-1}, \dots] = \phi_1 f_{n,1} + \phi_2 Z_n + \phi_3 Z_{n-1} + \phi_4 Z_{n-2}$
  - 한단계 예측오차 분산: $\mathrm{Var}[e_{n,1}] = \sigma_a^2$
  - 2-단계 예측오차 분산: $\mathrm{Var}[e_{n,2}] = \mathrm{Var}[Z_{n+2} | Z_n, Z_{n-1}, \dots] = (1 + \phi_1^2)\sigma_a^2$

ARMA모형 예측

(예- AR(2) 모형) 시점 1에서 20까지의 시계열데이터가 다음과 같다.

시점 1-10: -1.356, -1.567, -0.994, -0.417, 0.840, -0.991, 0.166, 0.889, 0.514, -0.491

시점 11-20: -0.766, -1.936, -2.223, -1.395, -1.512, -0.582, 1.204, 1.706, -0.768, -0.313

- 과거데이터를 바탕으로 시계열모형은 다음의 AR(2)를 따르는 것으로 추정되었다.

$$Z_t = 0.7Z_{t-1} - 0.2Z_{t-2} + a_t \ (\sigma_a^2 = 1.0)$$

- 시점 10에서부터 한단계이후 예측치를 구하고 다음 시점의 실제값과 비교하여 예측오차를 산출하라.

(풀이) 한단계이후 예측치는 다음과 같다.

$$f_{n,1} = 0.7Z_n - 0.2Z_{n-1}$$
$$f_{11,1} = 0.7Z_{11} - 0.2Z_{10}$$

n	Z_n	예측치	오차
10	-0.491	-	-
11	-0.766	-0.4465	-0.3195
12	-1.936	-0.4380	-1.4980
13	-2.223	-1.2020	-1.0210
14	-1.395	-1.1689	-0.2261
15	-1.512	-0.5319	-0.9801
16	-0.582	-0.7794	0.1974
17	1.204	-0.1050	1.3090
18	1.706	0.9592	0.7468
19	-0.768	0.9534	-1.7214
20	-0.313	-0.8788	0.5658

$f_{11,1} = 0.7 \times (-0.766) - 0.2 \times (-0.491) = -0.438$

시계열 예측

MA모형의 예측

- MA(2) 모형: $Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

- 한단계 예측식: $f_{n,1} = E[Z_{n+1}|Z_n, Z_{n-1}, \dots] = -\theta_1 a_n - \theta_2 a_{n-1}$
- 2-단계 예측식: $f_{n,2} = E[Z_{n+2}|Z_n, Z_{n-1}, \dots] = -\theta_2 a_n$
- 한단계 예측오차 분산: $\mathrm{Var}[e_{n,1}] = \sigma_a^2$
- 2-단계 예측오차 분산: $\mathrm{Var}[e_{n,2}] = \mathrm{Var}[Z_{n+2}|Z_n, Z_{n-1}, \dots] = (1 + \theta_1^2)\sigma_a^2$

- MA(q) 모형: $Z_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$

- 한단계 예측식: $f_{n,1} = E[Z_{n+1}|Z_n, Z_{n-1}, \dots] = -\theta_1 a_n - \theta_2 a_{n-1} - \cdots - \theta_q a_{n+1-q}$
- 2-단계 예측식: $f_{n,2} = E[Z_{n+2}|Z_n, Z_{n-1}, \dots] = -\theta_2 a_n - \theta_3 a_{n-1} - \cdots - \theta_q a_{n+2-q}$
- 한단계 예측오차 분산: $\mathrm{Var}[e_{n,1}] = \sigma_a^2$
- 2-단계 예측오차 분산: $\mathrm{Var}[e_{n,2}] = \mathrm{Var}[Z_{n+2}|Z_n, Z_{n-1}, \dots] = (1 + \theta_1^2)\sigma_a^2$
- 실제 예측치 산출을 위해서는 추정치 사용, 이 때 a_n의 추정치로는 해당 시점의 잔차를 사용

ARMA모형 예측

(예- MA(2) 모형) 시점 1에서 20까지의 시계열데이터가 다음과 같다.
시점 1-10: 1.377, 1.856, -0.655, -0.587, -0.188, 1.414, 0.731, -1.628, -0.511, -0.294
시점 11-20: 0.499, -0.442, 1.019, -1.705, -0.139, 0.219, 1.131, -0.508, 0.541, -0.809

- 과거데이터를 바탕으로 시계열모형은 다음의 MA(2)를 따르는 것으로 추정되었다.

$$Z_t = a_t - 0.6a_{t-1} - 0.2a_{t-2} \ (\sigma_a^2 = 1.0)$$

- 시점 11에서부터 한단계이후 예측치를 구하고 다음 시점의 실제값과 비교하여 예측오차를 산출하라.

(풀이) 한단계이후 예측치는 다음과 같다.

$$f_{n,1} = -0.6a_n - 0.2a_{n-1}$$

여기서 $a_t = (1 - 0.6B - 0.2B^2)^{-1} \approx (1 + 0.6B + 0.56B^2 + 0.456B^3 + 0.3856B^4)Z_t$

$f_{11,1} = -0.6 a_{11} - 0.2 a_{10} = 0.3814$

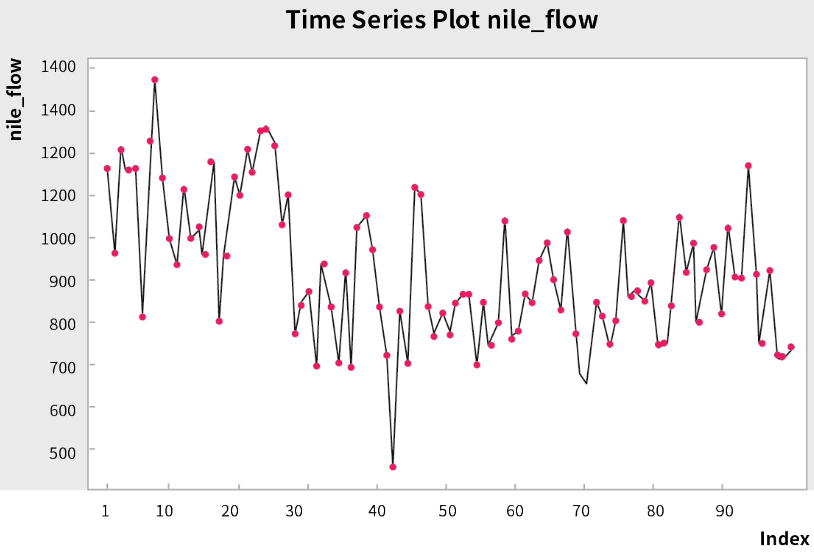
$f_{12,1} = -0.6 a_{12} - 0.2 a_{11} = 0.7856$

n	Z_n	a_n	예측치	오차
10	-0.294	-0.6337	-	-
11	0.499	-0.4240	0.4814	0.0176
12	-0.442	-1.1680	0.3812	-0.8232
13	1.019	0.7021	0.7856	0.2334
14	-1.705	-1.2269	-0.1877	-1.5173
15	-0.139	-0.6005	0.5957	-0.7347
16	0.219	-0.5250	0.6057	-0.3867
17	1.131	0.8000	0.4351	0.6959
18	-0.508	-0.4276	-0.3750	-0.1330
19	0.541	0.9158	0.0965	0.4444
20	-0.809	-0.1687	-0.4640	-0.3450

시계열 예측

ex) $357.5 + 0.4039 \times 718 + 0.2064 \times 919 = 837.1818 \fallingdotseq 837.2$

- 예 (나일강 유량) 1872-1970년 사이 나일강 유량 시계열의 추정 모형은 AR(2)이며 추정식은 다음과 같다.
  - $\hat{Z}_t = 357.5 + 0.4039Z_{t-1} + 0.2064Z_{t-2}$. 1967년 부터 5년간 한단계 예측값은 표와 같다.



년도	실제값	예측값	잔차
1967	919	847.0	72.0
1968	718	882.7	-164.7
1969	714	837.2	-123.2
1970	740	794.1	-54.1
1971	-	803.8	

#1