시계열자료분석팀

5팀

문병철
손재민
주혜인
한지원
홍현경

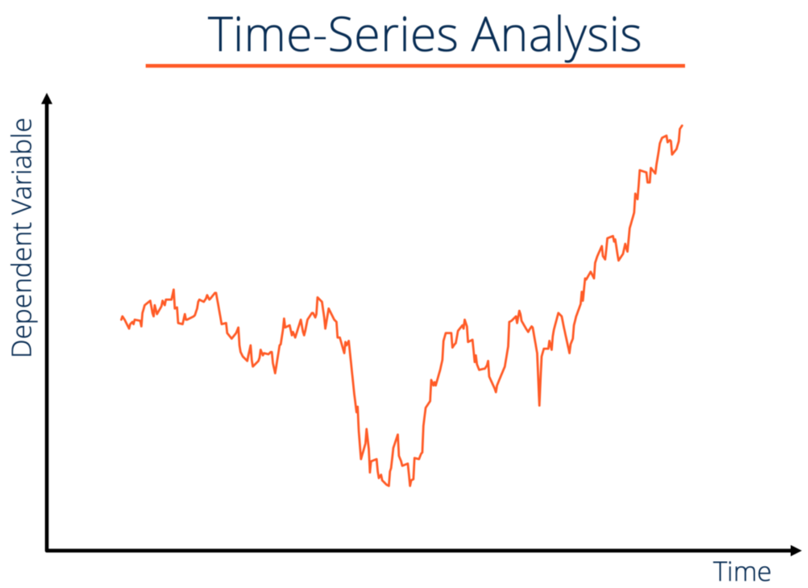
INDEX

1

시계열 알아가기

정의

시계열 자료란?



시간에 따라 관측된 자료 → 대개 서로 독립이 아니라는 특성을 가짐

시간 t에 따라 이산형 / 연속형으로 구분 가능

정의

시계열 자료란?

예측을 위한 분석

(추세분석, 평활법, 분해법, ARIMA 모형 등)

시계열 자료 분석 목적

시스템을 **이해**하고 **제어**하기 위한 분석

(스펙트럼 분석, 개입분석, 전이함수 모형 등)

시계열 분석

: 시간 순서대로 정렬된 데이터에서 의미 있는 요약과 통계 정보를 추출하는 것

1 시계열 알아가기

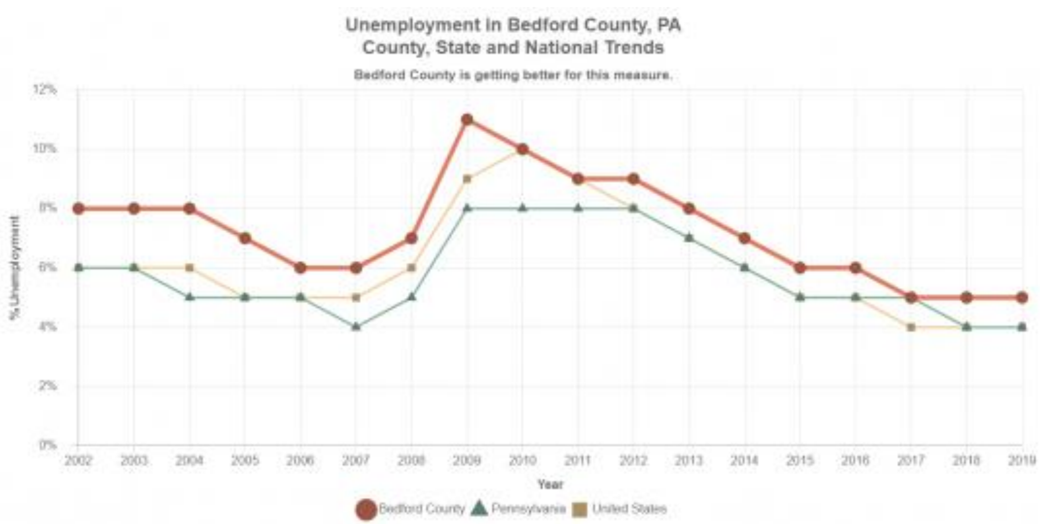
시계열자료의 구성요소

시계열 자료의 구성 요소

시간의 영향을 받는	시간의 영향을 받지 않는
체계적 성분	**불규칙 성분**
추세변동(Trend)	우연변동 / 불규칙 성분
순환변동(Cycle)	(Random Fluctuation)
계절변동(Seasonal Variation)	

시계열자료의 구성요소

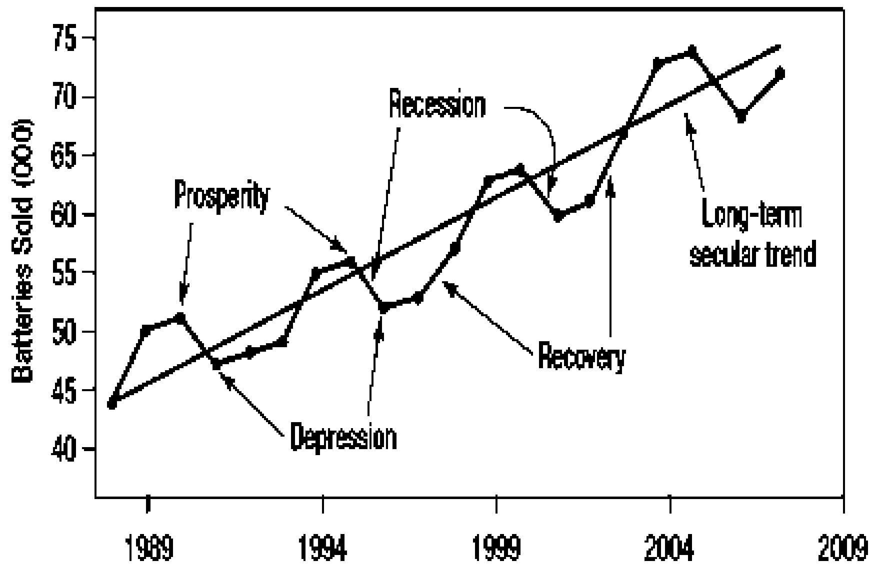
1. 추세변동(Trend)



시간의 경과에 따라 관측 값이 증가, 감소 하는 추세를 갖는 변동

시계열자료의 구성요소

2. 순환변동(Cycle)

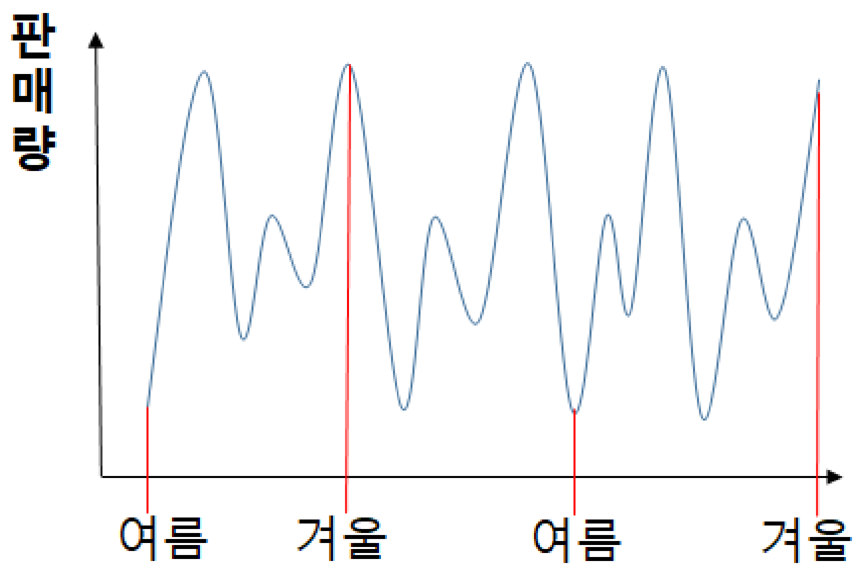


주기적인 변화를 가지지만 해당 변화가 계절에 의한 것이 아닌, 주기가 매우 긴 변동

ex) 태양 흑점 수의 변화

시계열자료의 구성요소

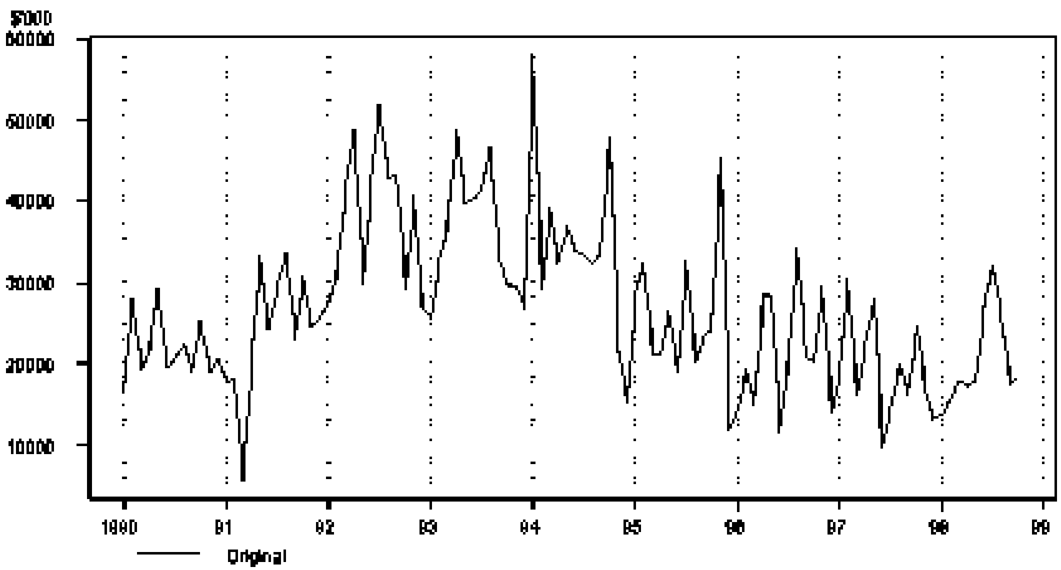
3. 계절변동(Seasonal Variation)



주별, 월별, 계절별과 같은 주기적인 성분에 의한 변동

시계열자료의 구성요소

4. 불규칙성분(Random Fluctuation)



시간에 따른 규칙적인 움직임과 무관하게 랜덤한 원인에 의해 나타나는 변동

2

정상성

정상성의 필요성과 정의

정상성의 필요성

시계열 모델은 **결합 분포**의 구체화

→ 미래의 값 예측을 위해서는 무한한 시점들의 **결합 분포** 고려 필요

현실적인 어려움 존재 !

정상성을 통해 몇 가지 가정으로 이를 대체,

대부분의 시계열 이론들이 정상성을 가정

정상성의 필요성과 정의

정상성의 필요성

→ 미래의 값 예측을 위해서는 무한한 시점들의 **결합 분포** 고려 필요

현실적인 어려움 존재!

정상성을 통해 몇 가지 가정으로 이를 대체,
대부분의 시계열 이론들이 정상성을 가정

“

정상성(Stationarity)

”

시계열의 확률적 성질 (평균, 분산 등)이
시간의 흐름에 영향을 받지 않는 것

강정상성
(Strict Stationarity)

약정상성
(Weak Stationarity)

정상성의 필요성과 정의

강정상성

$$(X_{t_1}, \dots, X_{t_n}) \overset{d}{=} (X_{t_1+h}, \dots, X_{t_n+h})$$

모든 h와 $n > 0$에 대하여 **결합확률밀도가 시간대(t)를 바꾸어도 동일**

하지만 !

강정상성을 만족하는 것은 현실적으로 매우 어려움

따라서 !

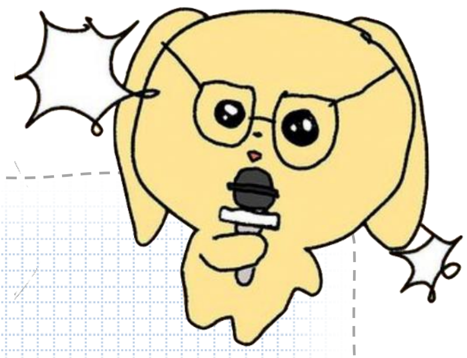
강정상성보다 완화된 조건의 약정상성을 사용

2 정상성(Stationarity)

정상성의 필요성과 정의

약정상성

중요하다 중요해! 모두 집중 집중 집중!!!



약정상성의 세가지 조건

$$i)\ \ E[|X_t|]^2 < \infty \ \ , \forall t \in \mathbb{Z}$$

$$ii)\ E[X_t] = m \ \ , \forall t \in \mathbb{Z}$$

$$iii)\ \gamma_X(h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$$

2 정상성(Stationarity)

정상성의 필요성과 정의

약정상성

약정상성의 세가지 조건

$i)\ \ E[|X_t|]^2 < \infty \ \ , \forall t \in \mathbb{Z}$

$ii)\ E[X_t] = m \ \ , \forall t \in \mathbb{Z}$

분산과 관련된 **2차 적률**이 존재하고, 이는 **시점 t에 관계없이 일정**하다!

2 정상성(Stationarity)

정상성의 필요성과 정의

약정상성

중요하다 중요해! 모두 집중 집중 집중!!!

약정상성의 세가지 조건

$i)\ E[|X_t|]^2 < \infty \quad , \forall t \in \mathbb{Z}$

$ii)\ E[X_t] = m \quad , \forall t \in \mathbb{Z}$

$iii)\ \gamma_X(h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$

평균이 상수이고, 이는 **시점 t 에 관계없이 일정**하다!

2 정상성(Stationarity)

정상성의 필요성과 정의

약정상성

중요하다 중요해! 모두 집중 집중 집중!!!

약정상성의 세가지 조건

$i)\ E[|X_t|]^2 < \infty \ , \forall t \in \mathbb{Z}$

자기공분산은 **시차** h **에만 의존**하고 시점 t 와는 무관하다!

$ii)\ E[X_t] = m \ , \forall t \in \mathbb{Z}$

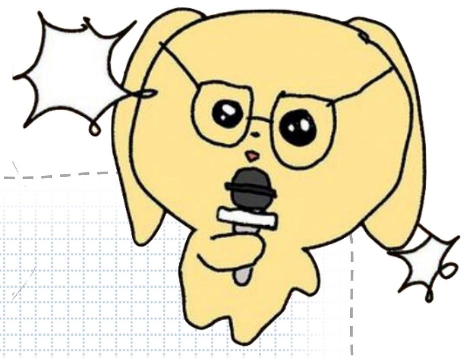
$$iii)\ \gamma_X(h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$$

2 정상성(Stationarity)

정상성의 필요성과 정의

약정상성

중요하다 중요해! 모두 집중 집중 집중!!!



약정상성의 세가지 조건

$$i)\ \ E[|X_t|]^2 < \infty \ \ , \forall t \in \mathbb{Z}$$

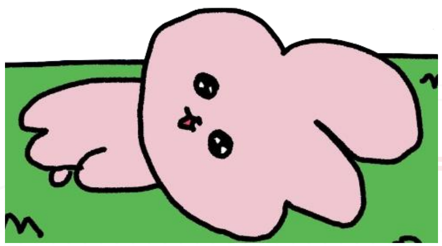
$$ii)\ E[X_t] = m \ \ , \forall t \in \mathbb{Z}$$

$$iii)\ \gamma_X(h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)]$$

약정상성을 만족하는지 확인하기 위해서는

1차 적률인 $\boldsymbol{E[X_t]}$와 자기공분산 $\boldsymbol{\gamma_X(h)}$만 고려하면 된다

정상성의 필요성과 정의



조건은 세가지인데

굳이 두가지만 고려하는 이유는 도대체 뭔데

자기공분산 $\gamma_X(h)$에서 $h = 0$ 인 경우

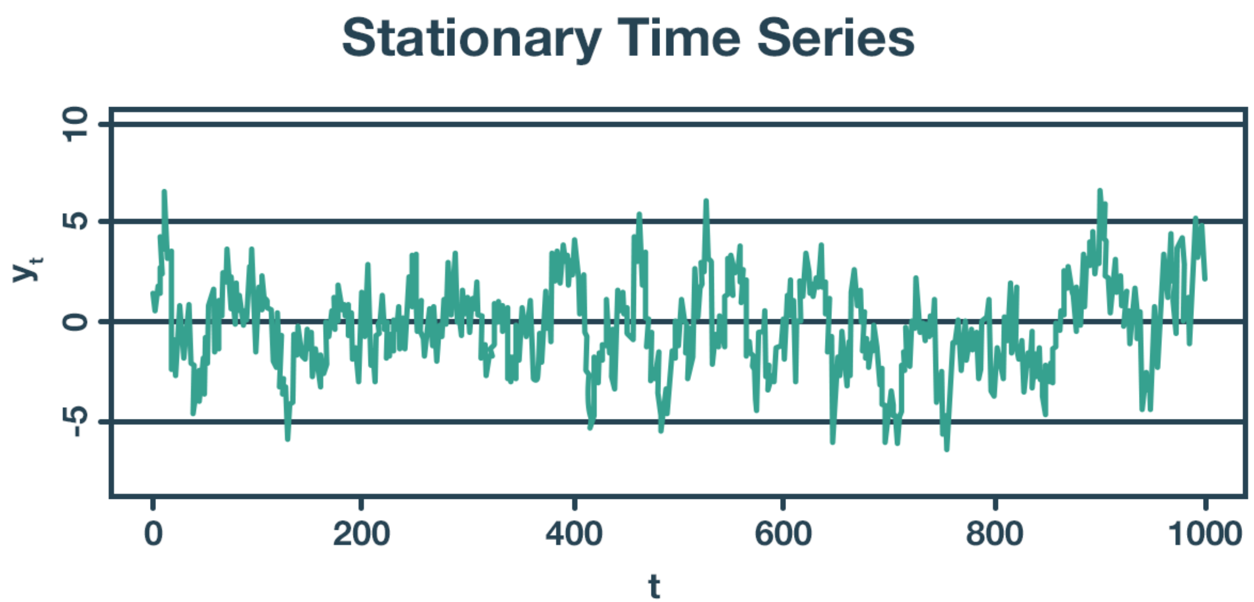
$Cov(X_t, X_{t+h}) = Cov(X_t, X_t) = Var(X_t)$가 되기 때문!

약정상성을 만족하는지 확인하기 위해서는

1차 적률인 $E[X_t]$와 자기공분산 $\gamma_X(h)$만 고려하면 된다

- 정상성을 만족하는 시계열

그래프로 알아보는 정상성



시간에 상관없이 **평균**과 **분산** 일정

3

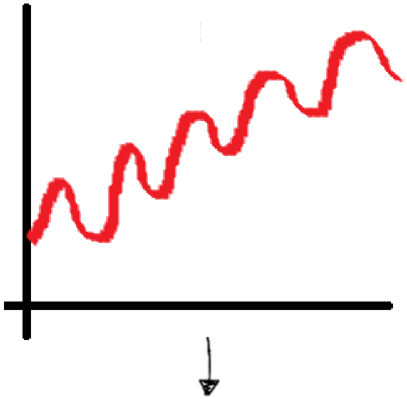
정상화

3 정상화

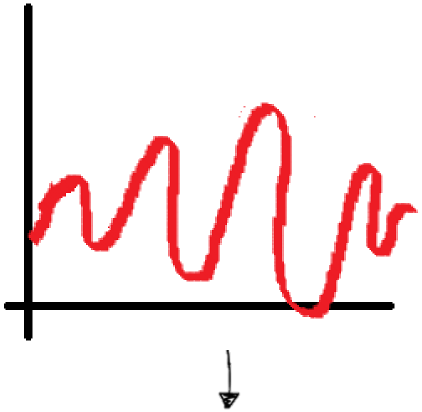
비정상 시계열

고심이와 함께 정리해보는 ~

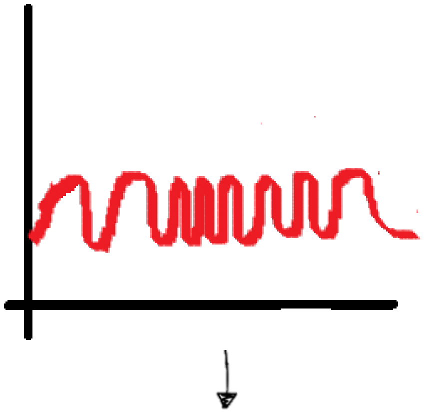
비정상 시계열이 될 수 있는 경우



평균이 일정하지 않은 경우



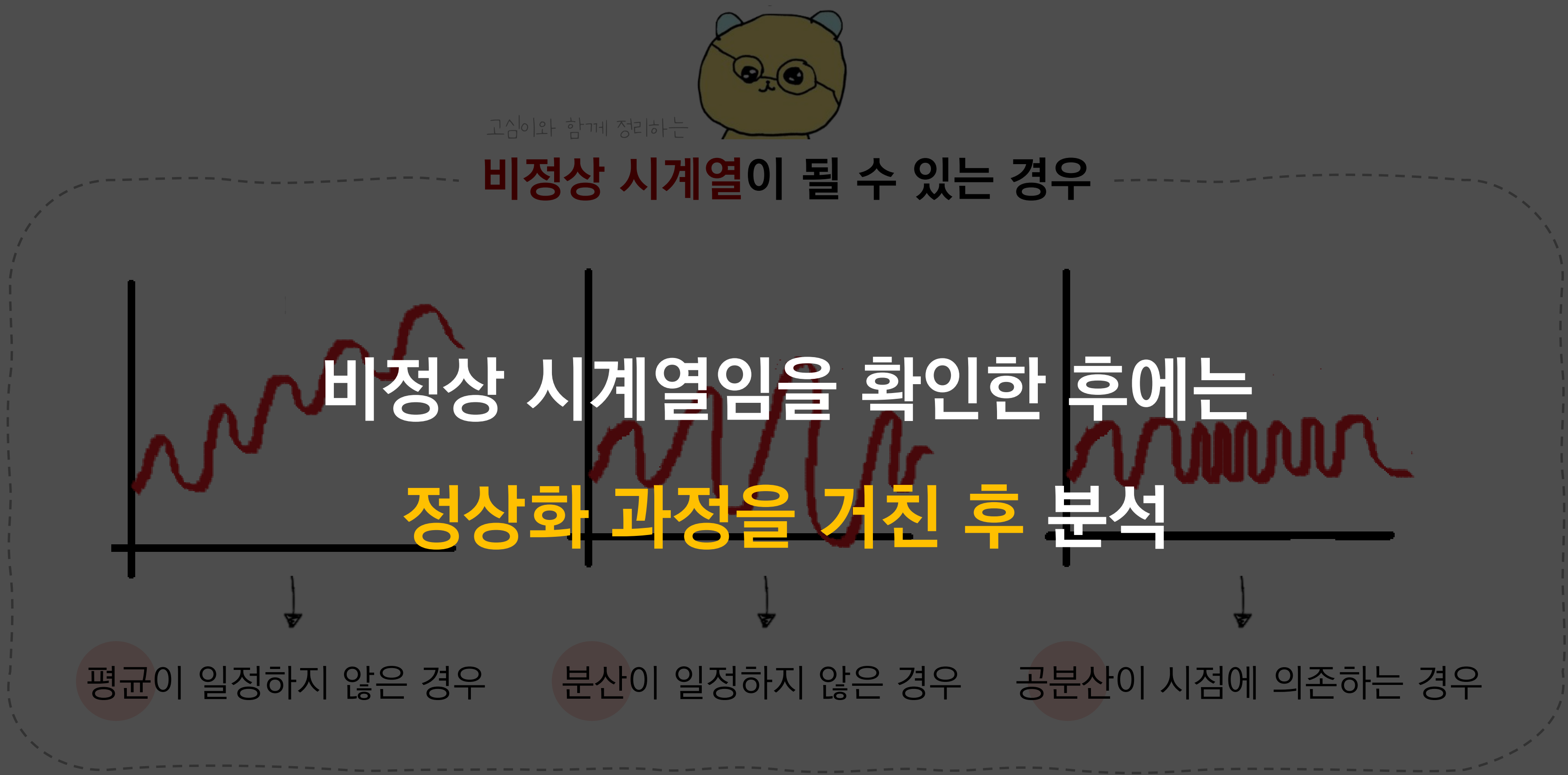
분산이 일정하지 않은 경우



공분산이 시점에 의존하는 경우

3 정상화

비정상 시계열

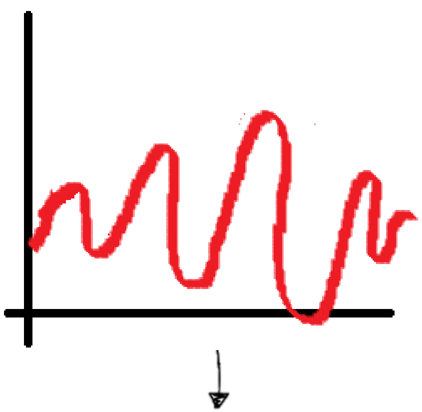


비정상 시계열임을 확인한 후에는

정상화 과정을 거친 후 분석

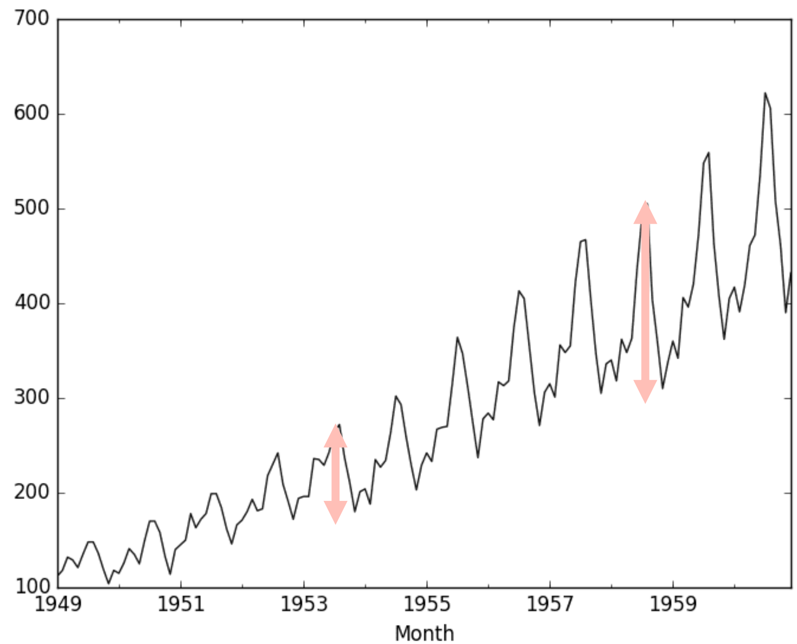
3 정상화

비정상 시계열



분산이 일정하지 않은 경우

EX)



시간에 따라 변동 폭 증가

분산이 일정하지 않은 경우

분산 안정화(VST)

시간에 따라 증가/감소하는 분산을 일정하게 변환

로그 변환

$$f(x) = \ln(x)$$

제곱근 변환

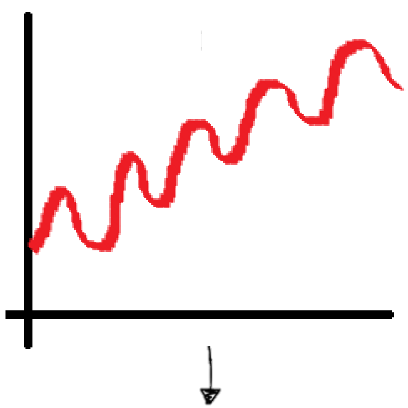
$$f(x) = \sqrt{x}$$

Box-cox 변환

$$f(x;\ \lambda) = \frac{x^{\lambda}-1}{\lambda}$$

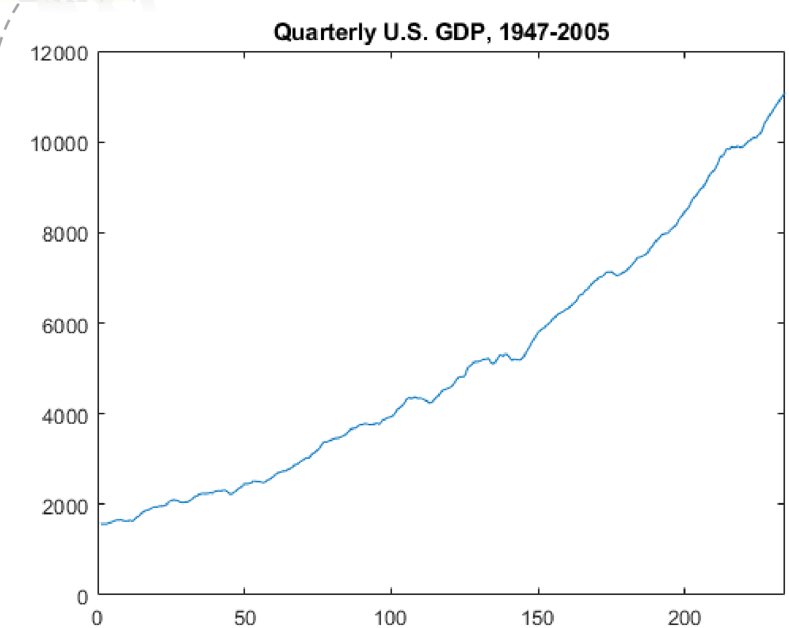
3 정상화

비정상 시계열



평균이 일정하지 않은 경우

EX)



시간에 따라 관측값이 증가하는 추세가 존재

평균이 일정하지 않은 경우 : 추세만 존재

회귀(Regression)

STEP 1

$$X_t = m_t + Y_t \, , E(Y_t) = 0$$

추세만 존재하는 시계열을 가정

평균이 일정하지 않은 경우 : 추세만 존재

회귀(Regression)

STEP 1

$$X_t = m_t + Y_t , E(Y_t) = 0$$

추세만 존재하는 시계열을 가정

STEP 2

$$m_t = c_0 + c_1 t + c_2 t^2 + \cdots + c_p t^p$$

추세성분을 시간 t에 대한 선형회귀식으로 표현

평균이 일정하지 않은 경우 : 추세만 존재

회귀(Regression)

STEP 3

$$(\widehat{c_o}, \ldots, \widehat{c_p}) = \operatorname*{argmin}_{c} \sum_{t=1}^{n} (X_t - m_t)^2$$

STEP 2의 선형식의 계수를 OLS로 추정

오차 제곱합을 최소화하는
값을 추정하는 방법

평균이 일정하지 않은 경우 : 추세만 존재

회귀(Regression)

STEP 3

$$(\widehat{c_o}, \dots, \widehat{c_p}) = \underset{c}{\operatorname{argmin}} \sum_{t=1}^{n} (X_t - m_t)^2$$

STEP 2의 선형식의 계수를 OLS로 추정

STEP 4

$$X_t - \widehat{m}_t = Y_t$$

추정한 추세를 시계열에서 제거

평균이 일정하지 않은 경우 : 계절성만 존재

조화회귀(Harmonic Regression)

STEP 1

$$X_t = s_t + Y_t \ , E(Y_t) = 0 \ , s_{t+d} = s_t = s_{t-d}$$

$$s_t = a_0 + \sum_{j=1}^{k} (a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t))$$

주기가 d인 계절성만을 가지는 시계열을 가정

계절 성분 s_t를 시간 t에 대한 회귀식으로 표현

평균이 일정하지 않은 경우 : 계절성만 존재

조화회귀(Harmonic Regression)

k는 주로 1~4 사이의 값 사용

STEP 2

$$s_t = a_0 + \sum_{j=1}^{k} \left(a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)\right)$$

λ_j는 2π인 함수의 주기와 데이터 주기를 맞추기 위한 값

적절한 λ_j와 k를 선택하여 OLS로 a_j와 b_j를 추정

평균이 일정하지 않은 경우 : 계절성만 존재

조화회귀(Harmonic Regression)

k는 주로 1~4 사이의 값 사용

STEP 2

$$s_t = a_0 + \sum_{j=1}^{k} \left(a_j \cos\left(\lambda_j t\right) + b_j \sin(\lambda_j t)\right)$$

λ_j는 2π인 함수의 주기와 데이터 주기를 맞추기 위한 값

적절한 λ_j와 k를 선택하여 OLS로 a_j와 b_j를 추정

STEP 3

$$X_t - \hat{s_t} = Y_t$$

추정한 계절성분인 $\hat{s_t}$ 를 X_t에서 제거

평균이 일정하지 않은 경우 : 추세 및 계절성 동시에 존재

회귀(Regression)

STEP 1 : 추세만 있는 시계열을 가정한 후 추세 제거

$$X_t = m_t + Y_t \, , E(Y_t) = 0$$

$$X_t - \hat{m}_t = Y_t$$

평균이 일정하지 않은 경우 : 추세 및 계절성 동시에 존재

회귀(Regression)

STEP 1 : 추세만 있는 시계열을 가정한 후 추세 제거

$$X_t = m_t + Y_t , E(Y_t) = 0 \longrightarrow X_t - \widehat{m}_t = Y_t$$

STEP 2 : 계절성만 있는 시계열을 가정한 후 계절성 제거

$$X_t = s_t + Y_t , E(Y_t) = 0 \longrightarrow X_t - \hat{s}_t = Y_t$$

평균이 일정하지 않은 경우 : 추세 및 계절성 동시에 존재

회귀(Regression)

STEP 1 : 추세만 있는 시계열을 가정한 후 추세 제거

$$X_t = m_t + Y_t \, , E(Y_t) = 0 \quad \rightarrow \quad X_t - \hat{m}_t = Y_t$$

STEP 2 : 계절성만 있는 시계열을 가정한 후 계절성 제거

$$X_t = s_t + Y_t \, , E(Y_t) = 0 \quad \rightarrow \quad X_t - \hat{s}_t = Y_t$$

STEP 3 : 여전히 추세가 존재 한다면, *STEP 1* 과정 반복

평균이 일정하지 않은 경우 : 추세 및 계절성 동시에 존재

회귀(Regression)

STEP 1 : 추세만 있는 시계열을 가정한 후 추세 제거

$$X_t = m_t + Y_t \,, E(Y_t) = 0 \longrightarrow X_t - \hat{m}_t = Y_t$$

BUT,

STEP 2 : 계절성만 있는 시계열을 가정한 후 계절성 제거

OLS에서 오차항의 독립성을 가정하는 회귀와 달리,

시계열의 오차항은 독립성을 가정하지 않아 추정이 정확하지 않을 가능성 존재

$$X_t = s_t + Y_t \,, E(Y_t) = 0 \longrightarrow X_t - \hat{s}_t = Y_t$$

STEP 3 : 여전히 추세가 존재 한다면, *STEP 1* 과정 반복

평균이 일정하지 않은 경우 : 추세만 존재

평활(Smoothing)

시점을 여러 구간으로 쪼갠 후, 구간의 평균을 구해 추세를 구하는 과정

1

이동평균 평활법

2

지수평활법

평균이 일정하지 않은 경우 : 추세만 존재

평활(Smoothing)

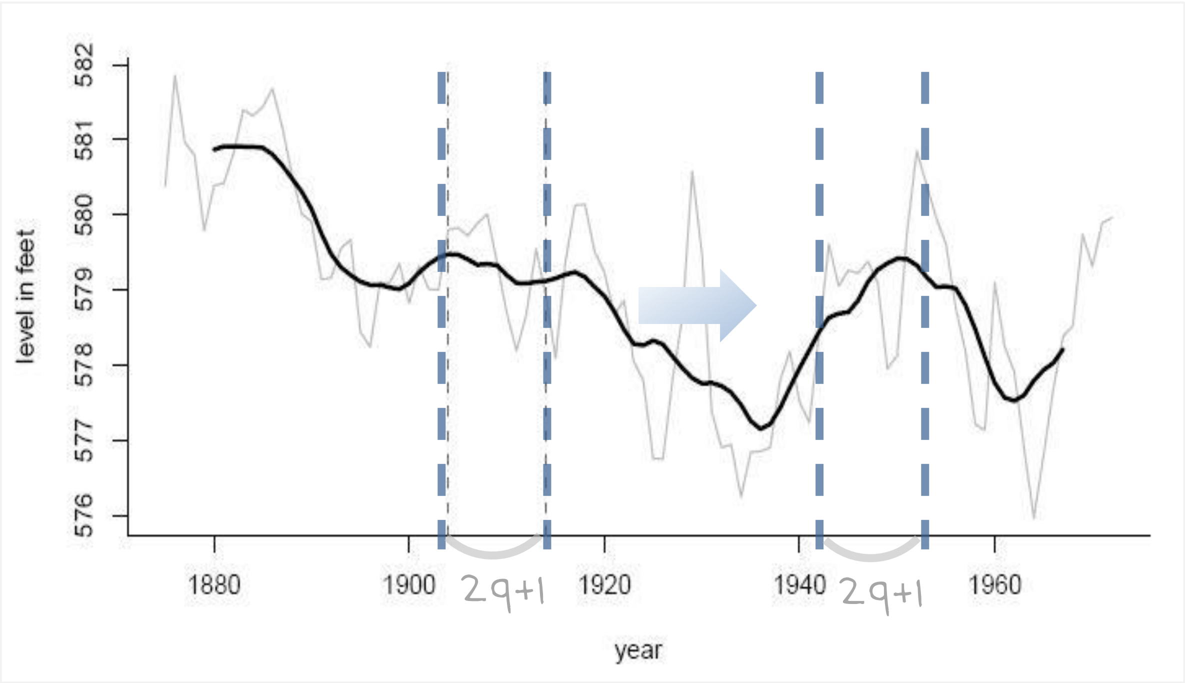
시점을 여러 구간으로 쪼갠 후, 구간의 평균을 구해 추세를 구하는 과정

1 이동평균 평활법

2 지수평활법

평균이 일정하지 않은 경우 : 추세만 존재

이동평균 평활법



일정 기간마다 평균을 계산하여 추세를 추정하는 방식

평균이 일정하지 않은 경우 : 추세만 존재

이동평균 평활법

STEP 1

$$W_t = \frac{1}{2q+1}\sum_{j=-q}^{j=q}\left(m_{t+j} + Y_{t+j}\right)$$

$$= \frac{1}{2q+1}\sum_{j=-q}^{q} m_{t+j} + \frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j}$$

길이가 $2q+1$인 구간의 평균을 구한다

+ 시점의 추세 추정을 위해 앞, 뒤 q개의 시계열 활용!

평균이 일정하지 않은 경우 : 추세만 존재

이동평균 평활법

STEP 2

$$m_t = c_0 + c_1 t$$

$$\frac{1}{2q+1}\sum_{j=-q}^{q} m_{t+j} = c_0 + c_1 t = m_t \quad , t \in [q+1, n-q]$$

$$\frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j} \approx E(Y_t) = 0 \; (by\ WLLN)$$

추세 성분 m_t를 대입 (추세가 linear하다고 가정할 때)

평균이 일정하지 않은 경우 : 추세만 존재

이동평균 평활법

STEP 2

$$m_t = c_0 + c_1 t$$

$$\frac{1}{2q+1}\sum_{j=-q}^{q} m_{t+j} = c_0 + c_1 t = m_t \;\;, t \in [q+1, n-q]$$

$$\frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j} \;\approx E(Y_t) = 0 \;(by\; WLLN)$$

추세 성분 m_t를 대입 (추세가 linear하다고 가정할 때)

STEP 3

$$X_t - \hat{w}_t = Y_t$$

추정한 추세인 $\widehat{w_t}$를 X_t에서 제거

평균이 일정하지 않은 경우 : 추세만 존재

이동평균 평활법

문제점

현실에서 t 시점을 예측할 때, t시점 이후의 데이터를 활용 → 거의 불가능

최근 자료와 과거 자료의 가중치가 $\frac{1}{2q+1}$로 동일함

과거 데이터에만 의존하는 지수평활법을 알아보자!

평균이 일정하지 않은 경우 : 추세만 존재

지수 평활법

시점 1일 때의 추세 추정값 : $\hat{m}_1 = X_1$

시점 2일 때의 추세 추정값 : $\hat{m}_2 = aX_2 + (1-A)\hat{m}_1 = aX_2 + (1-a)X_1$

⋮

시점 t일 때의 추세 추정값 : $\hat{m}_t = aX_t + (1-a)\hat{m}_{t-1}$

$$= \sum_{j=0}^{t-2} a(1-a)^j X_{t-j} + (1-a)^{t-1} X_1$$

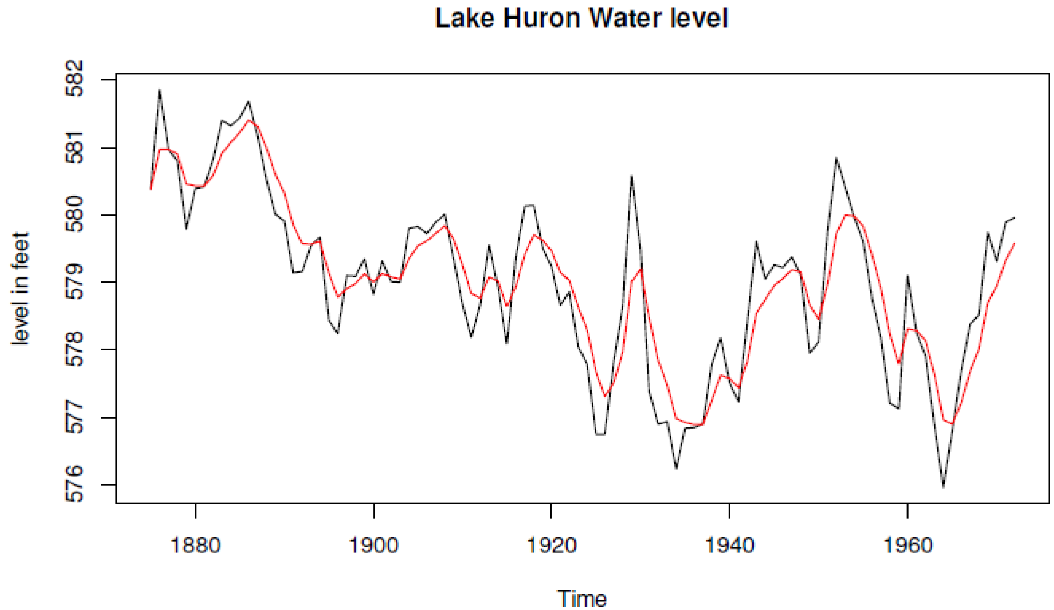
$0 < \alpha < 1$ 이므로 과거 시점일수록 더 작은 가중치 부여

시점 t까지의 관찰값, 즉 과거 시점의 데이터를 이용하여 추세를 추정

→ **시점별 가중평균**으로 추세를 파악하는 방법

평균이 일정하지 않은 경우 : 추세만 존재

지수 평활법



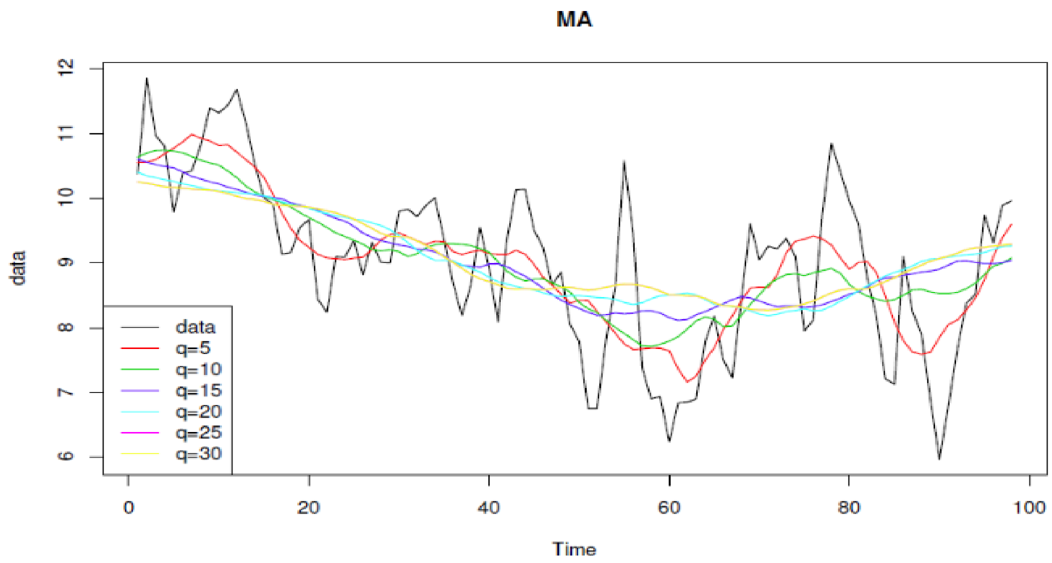
$\alpha = 0.4$ 일 경우, 지수평활법으로 추정된 추세

시점 t까지의 관찰값, 즉 과거 시점의 데이터를 이용하여 추세를 추정

→ **시점별 가중평균**으로 추세를 파악하는 방법

평균이 일정하지 않은 경우 : 추세만 존재

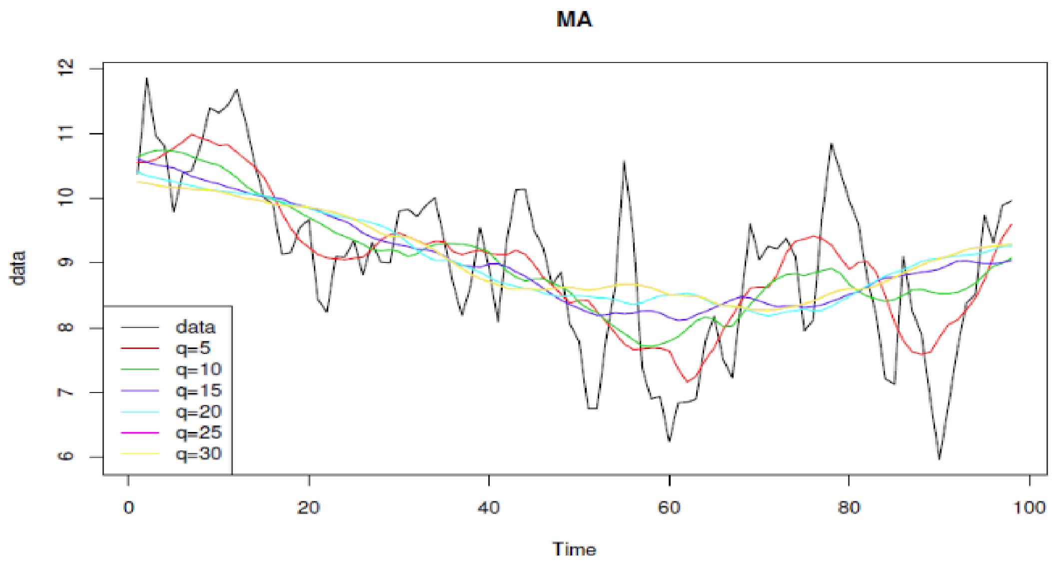
평활법의 tuning parameter



평활법은 tuning parameter를 선택해야 한다는 단점 존재

● 평균이 일정하지 않은 경우 : 추세만 존재

평활법의 tuning parameter



평활법은 tuning parameter를 선택해야 한다는 단점 존재

결과에 중요한 영향을 미치므로,

Cross-validation(CV)를 통해 최적의 파라미터를 선택

평균이 일정하지 않은 경우 : 추세만 존재

평활법의 tuning parameter

$$W_t = \underbrace{\frac{1}{2q+1}\sum_{j=-q}^{q} m_{t+j}}_{\approx m_t \ if \ q \ is \ small} + \underbrace{\frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j}}_{\approx 0 \ if \ q \ is \ large}$$

bias

variance

CV와 bias-variance trade off가 궁금하다면? 고심이는 데마팀 클린업을 참고하는 것을 강력 추천할게 ~

평활법에서의 bias – variance trade off

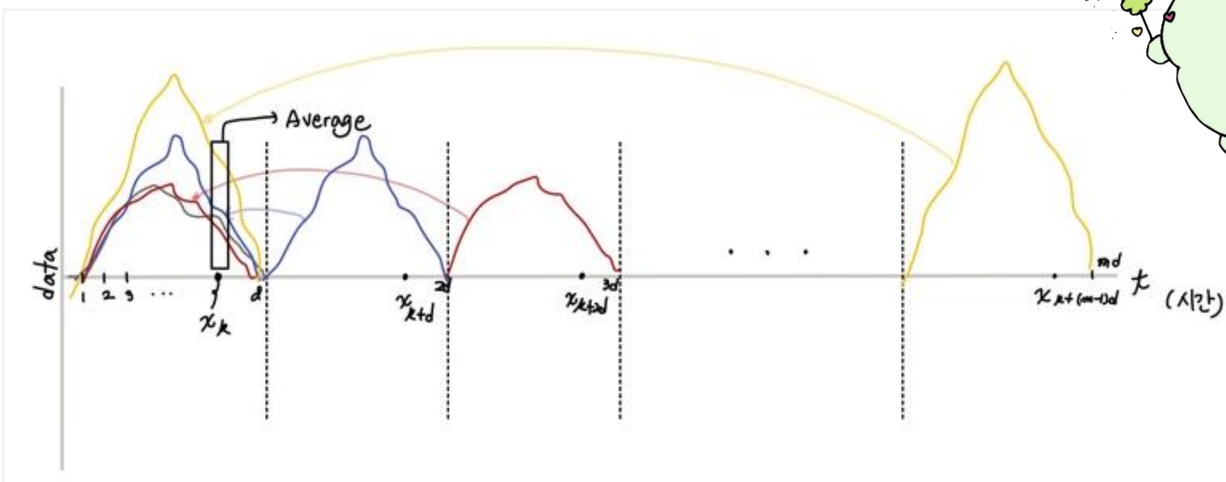
q ↓	$\longrightarrow$	bias ↓	variance ↑
q ↑	$\longrightarrow$	bias ↑	variance ↓

3 정상화

평균이 일정하지 않은 경우 : 계절성만 존재

Seasonal smoothing

내가 직접 그렸어! 내가! 내가! 내가!



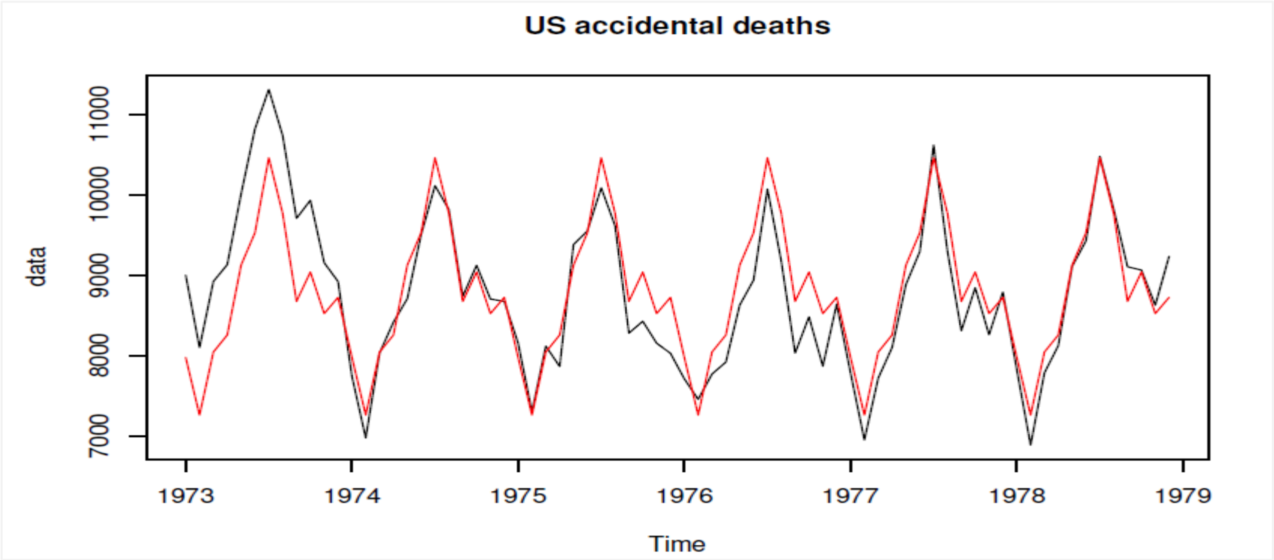
팀장님께서 한땀한땀 셀프 디자인한 그림 (made by 말랑콩떡...)

주기가 d인 관측치를 한 주기 안에 **겹친 후 평균**을 내자!

평균이 일정하지 않은 경우 : 계절성만 존재

Seasonal smoothing

$k = 1, \cdots, d$ 에 대하여 계절성분 추정

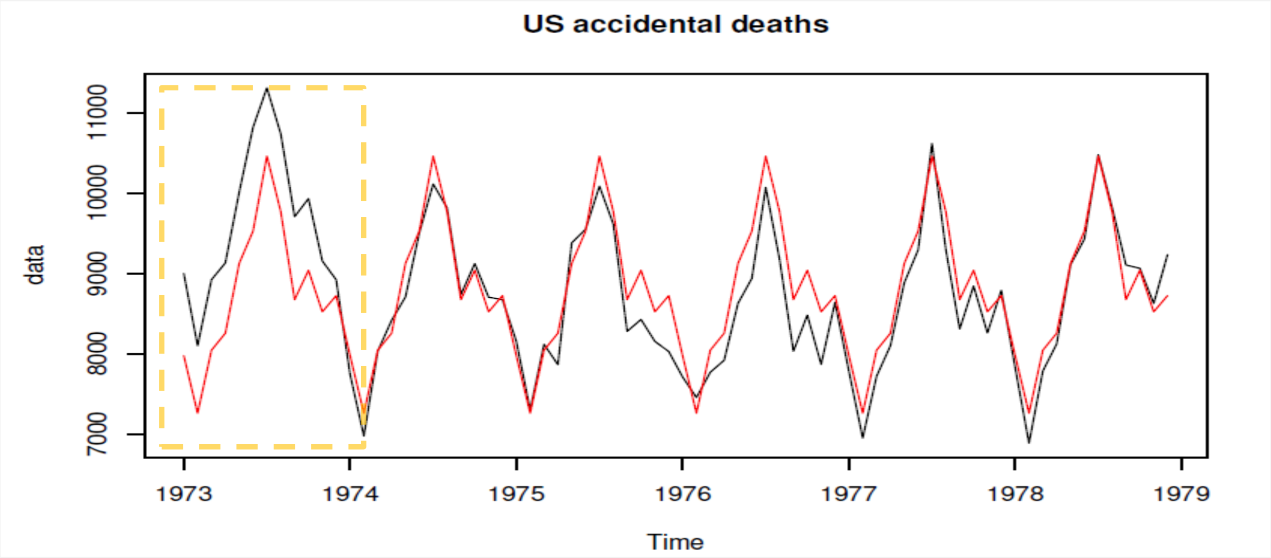


$$\hat{s}_k = \frac{1}{m}\left(x_k + x_{k+d} + \cdots + x_{k+(m-1)d}\right) = \frac{1}{m}\, \Sigma_{j=0}^{m-1} x_{k+jd}$$

평균이 일정하지 않은 경우 : 계절성만 존재

Seasonal smoothing

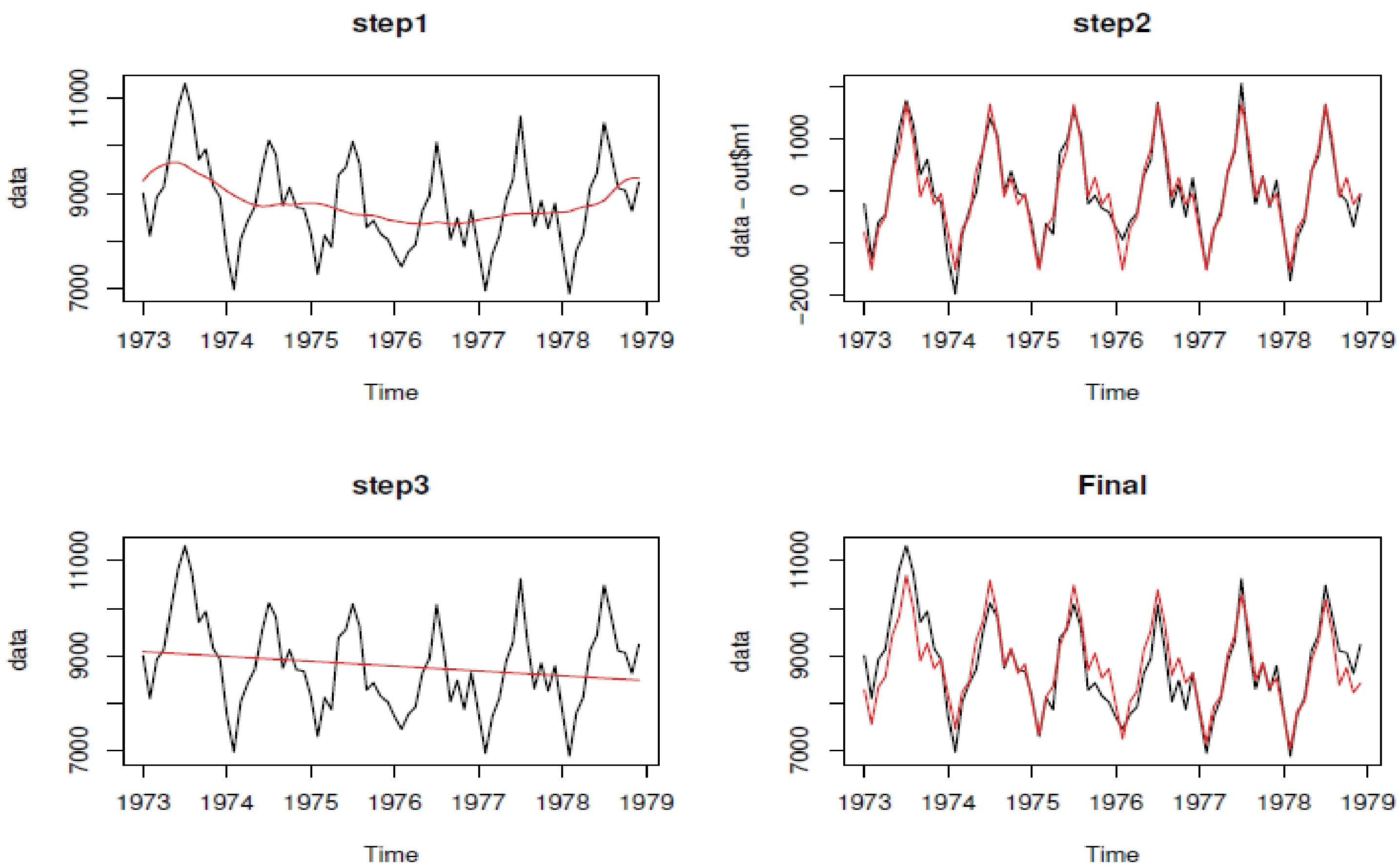
$k = 1, \cdots, d$ 에 대하여 계절성분 추정



첫번째 주기의 패턴이 반복됨을 확인 가능

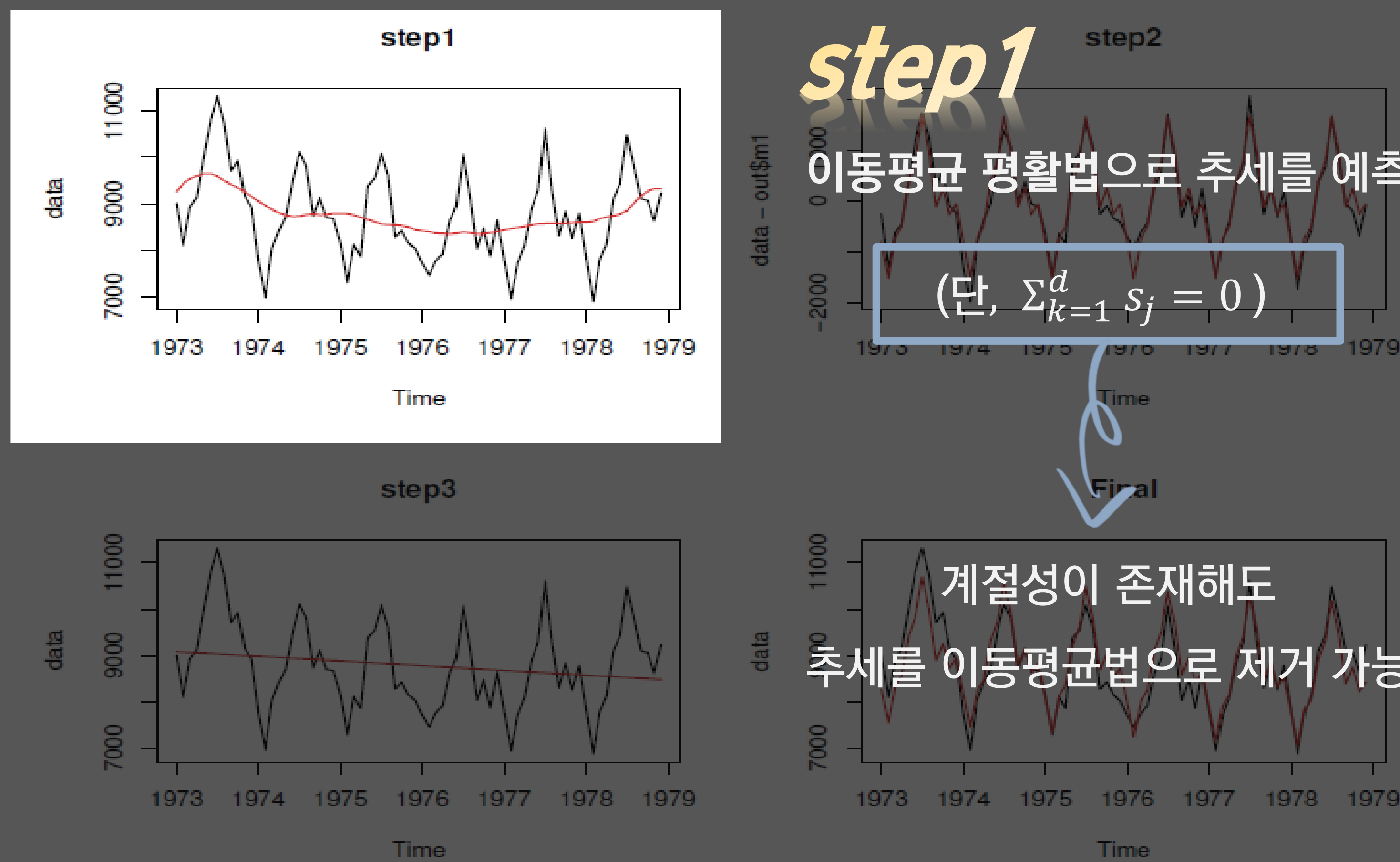
평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재

Classical decomposition algorithm



- 평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재

Classical decomposition algorithm



step1

이동평균 평활법으로 추세를 예측

(단, $\sum_{k=1}^{d} s_j = 0$)

계절성이 존재해도

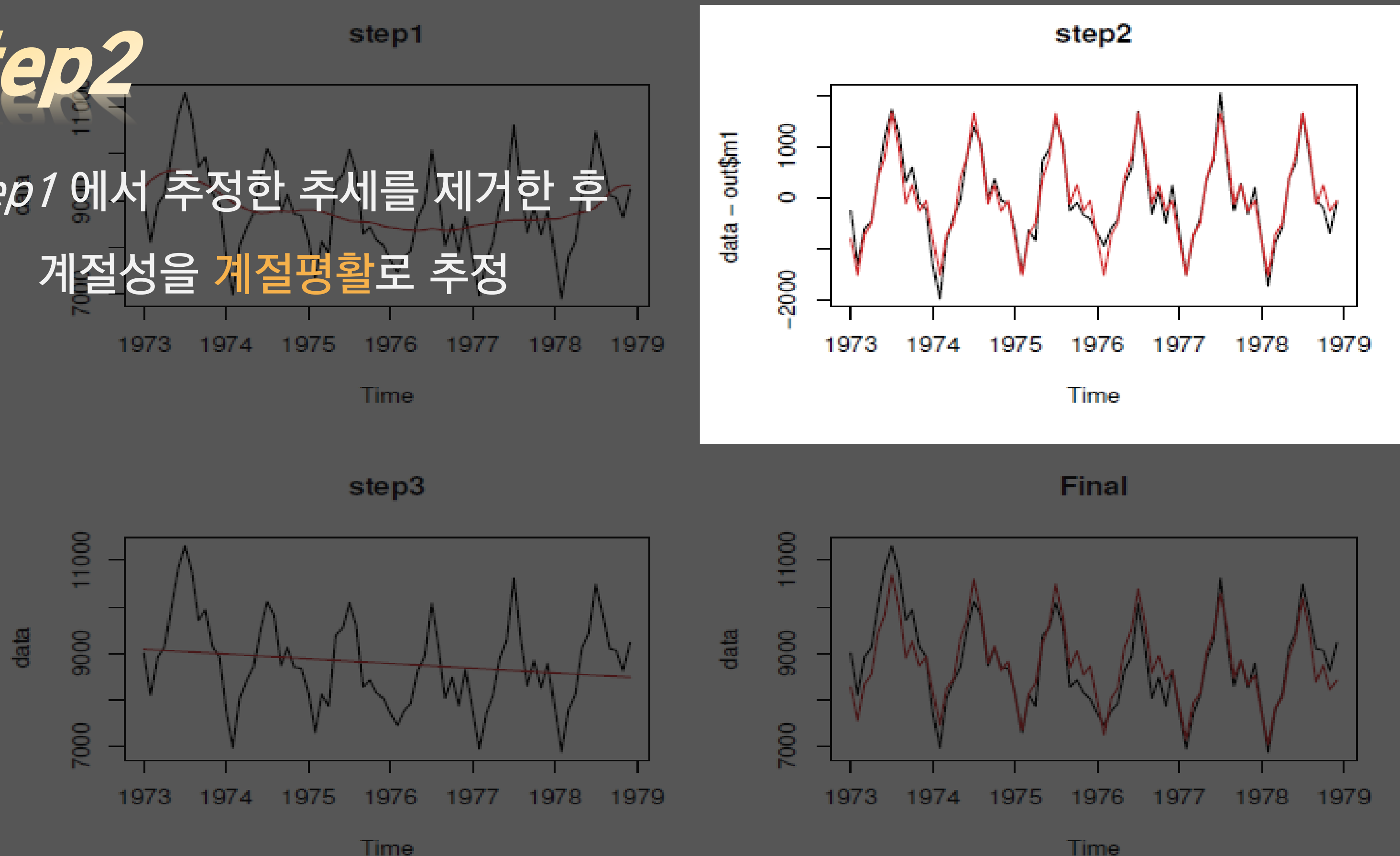
추세를 이동평균법으로 제거 가능

- **평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재**

Classical decomposition algorithm

step2

step1 에서 추정한 추세를 제거한 후 계절성을 계절평활로 추정



평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재

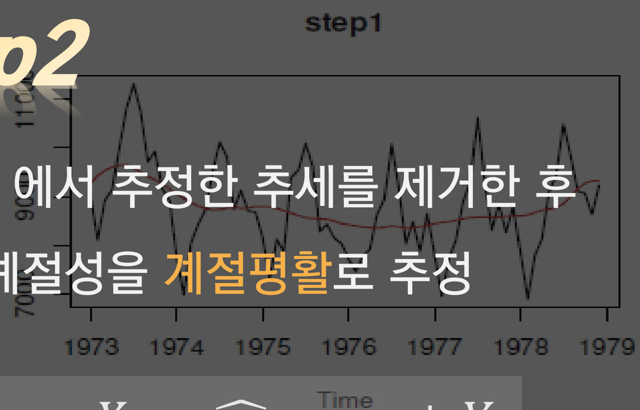
Classical decomposition algorithm

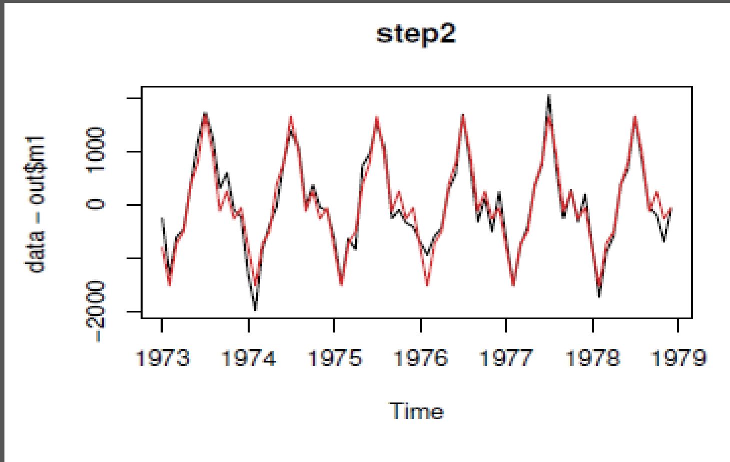
step2

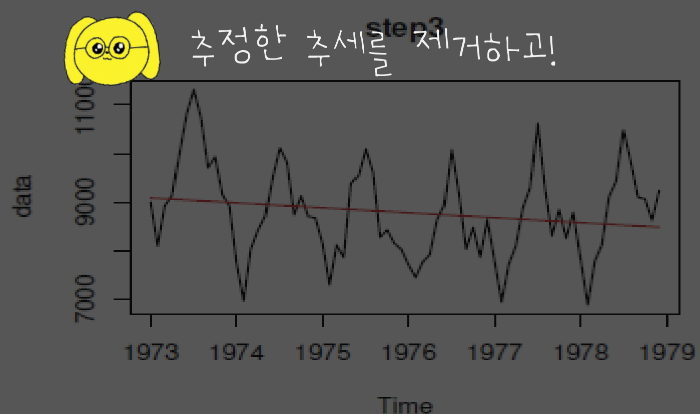
step1 에서 추정한 추세를 제거한 후 계절성을 **계절평활**로 추정

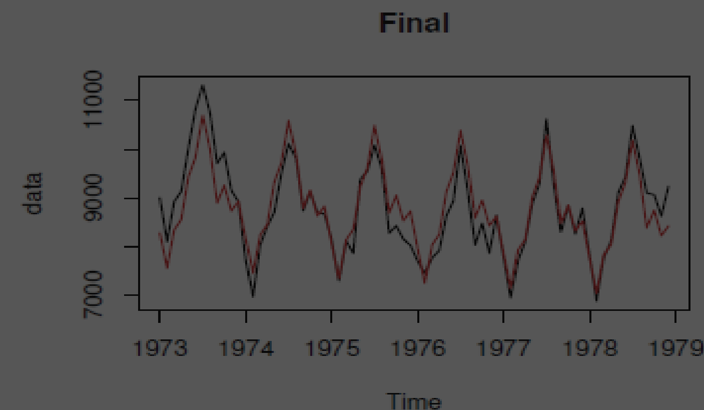
$$z_t = X_t - \widehat{m}_t \approx s_t + Y_t$$

추정한 추세를 제거하고!









- **평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재**

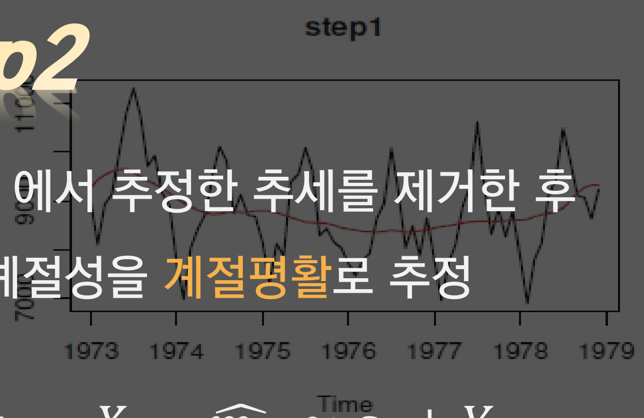
Classical decomposition algorithm

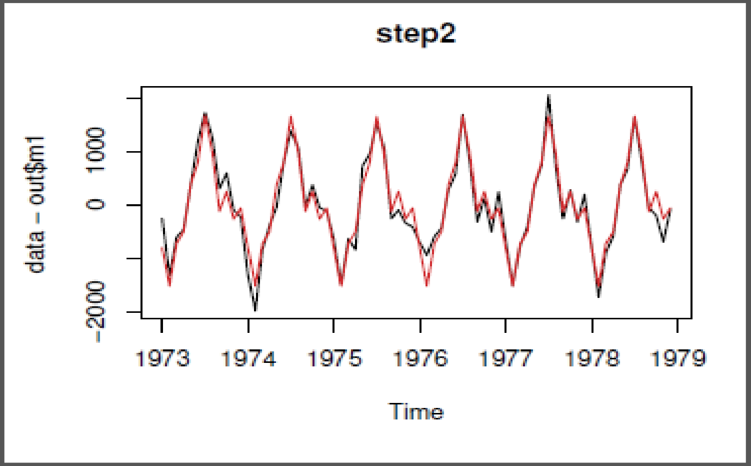
step2

step1 에서 추정한 추세를 제거한 후

계절성을 **계절평활**로 추정

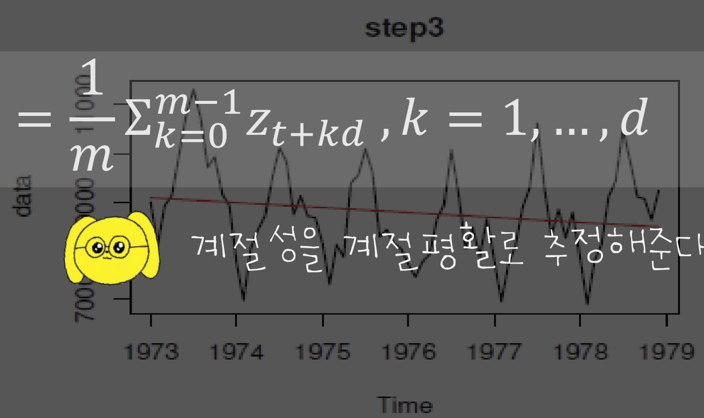
$$z_t = X_t - \widehat{m_t} \approx s_t + Y_t$$

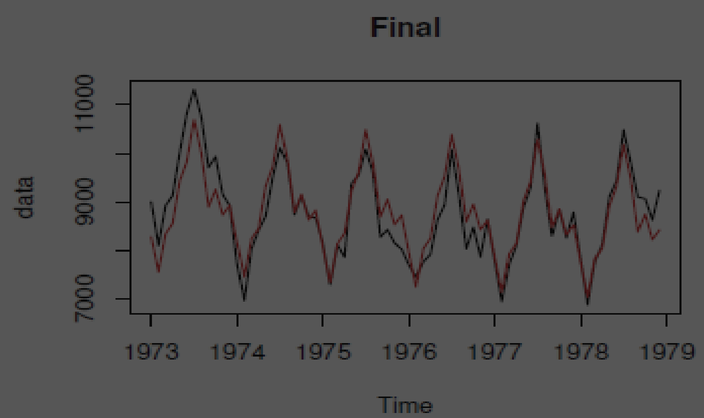




$$\hat{s}_t^* = \frac{1}{m}\Sigma_{k=0}^{m-1} z_{t+kd}\ , k = 1, \dots, d$$

계절성을 계절평활로 추정해준다!



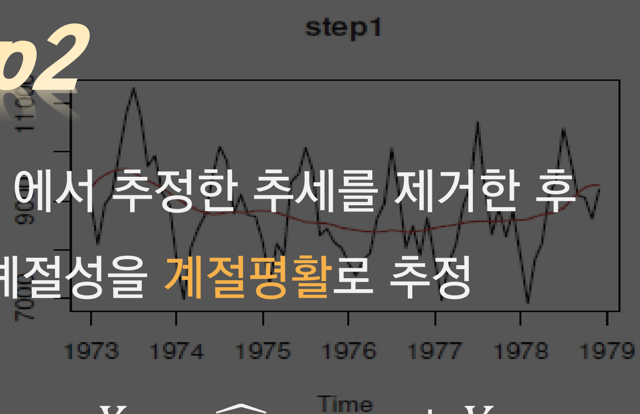


평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재

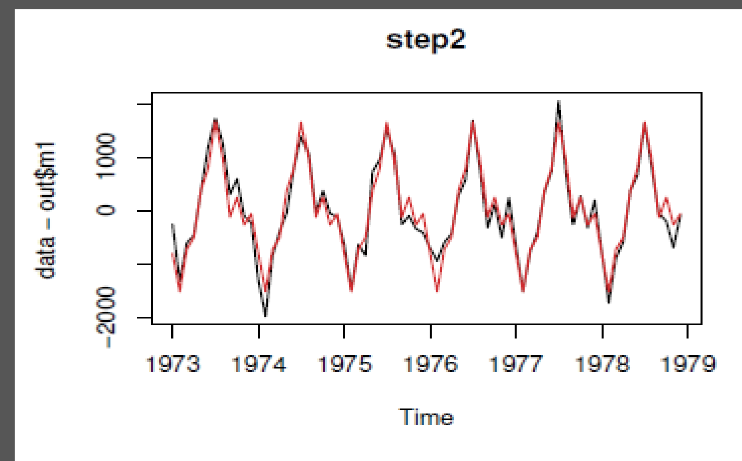
Classical decomposition algorithm

step2

step1 에서 추정한 추세를 제거한 후 계절성을 계절평활로 추정

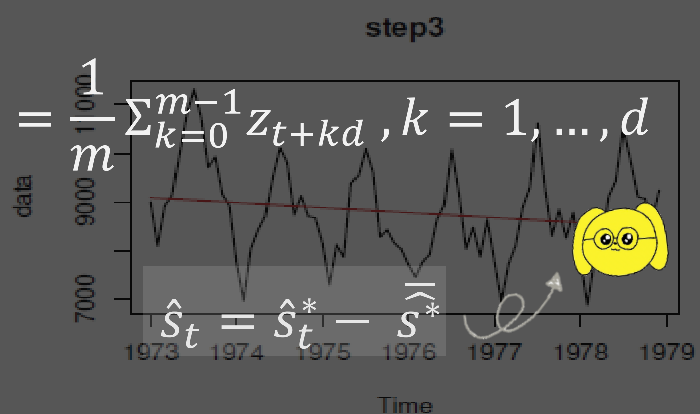


$$z_t = X_t - \widehat{m_t} \approx s_t + Y_t$$



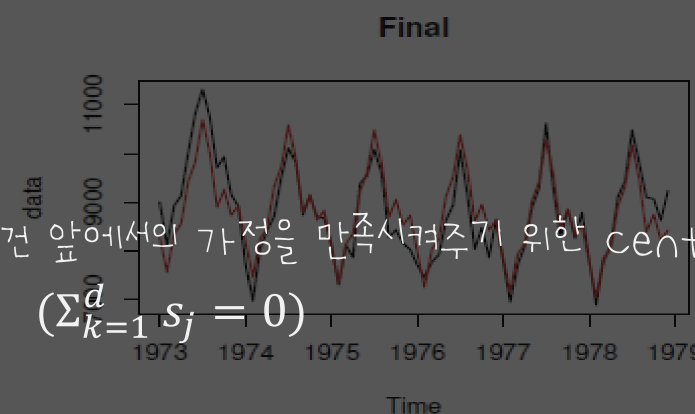
$$\hat{s}_t^* = \frac{1}{m} \Sigma_{k=0}^{m-1} z_{t+kd} \,, k = 1, \ldots, d$$

$$\hat{s}_t = \hat{s}_t^* - \overline{\hat{s}^*}$$



이건 앞에서의 가정을 만족시켜주기 위한 centering

$$(\Sigma_{k=1}^{d} s_j = 0)$$



평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재

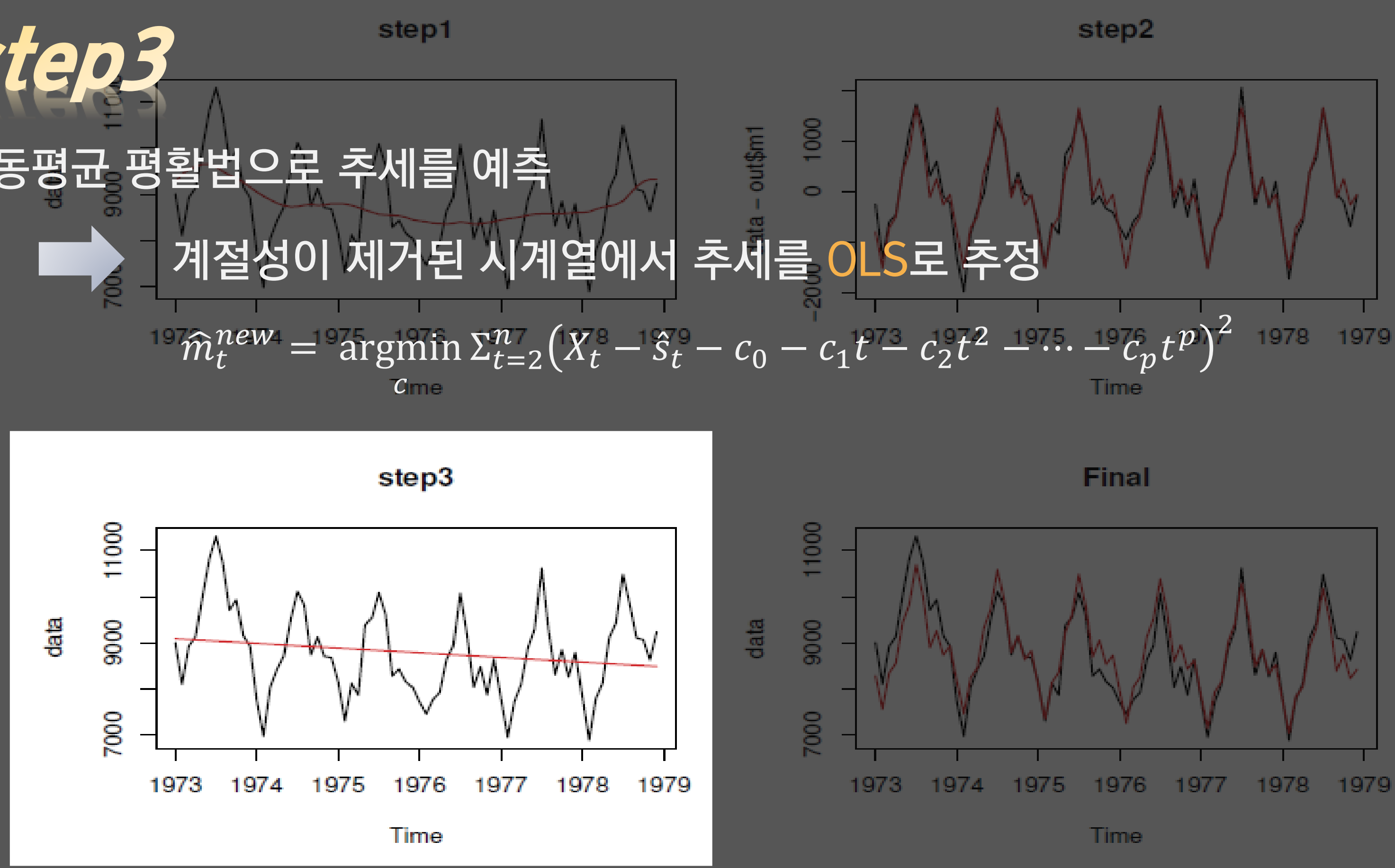
Classical decomposition algorithm

step3

이동평균 평활법으로 추세를 예측

계절성이 제거된 시계열에서 추세를 OLS로 추정

$$\hat{m}_t^{new} = \underset{c}{\operatorname{argmin}}\ \Sigma_{t=2}^{n}\left(X_t - \hat{s}_t - c_0 - c_1 t - c_2 t^2 - \cdots - c_p t^p\right)^2$$



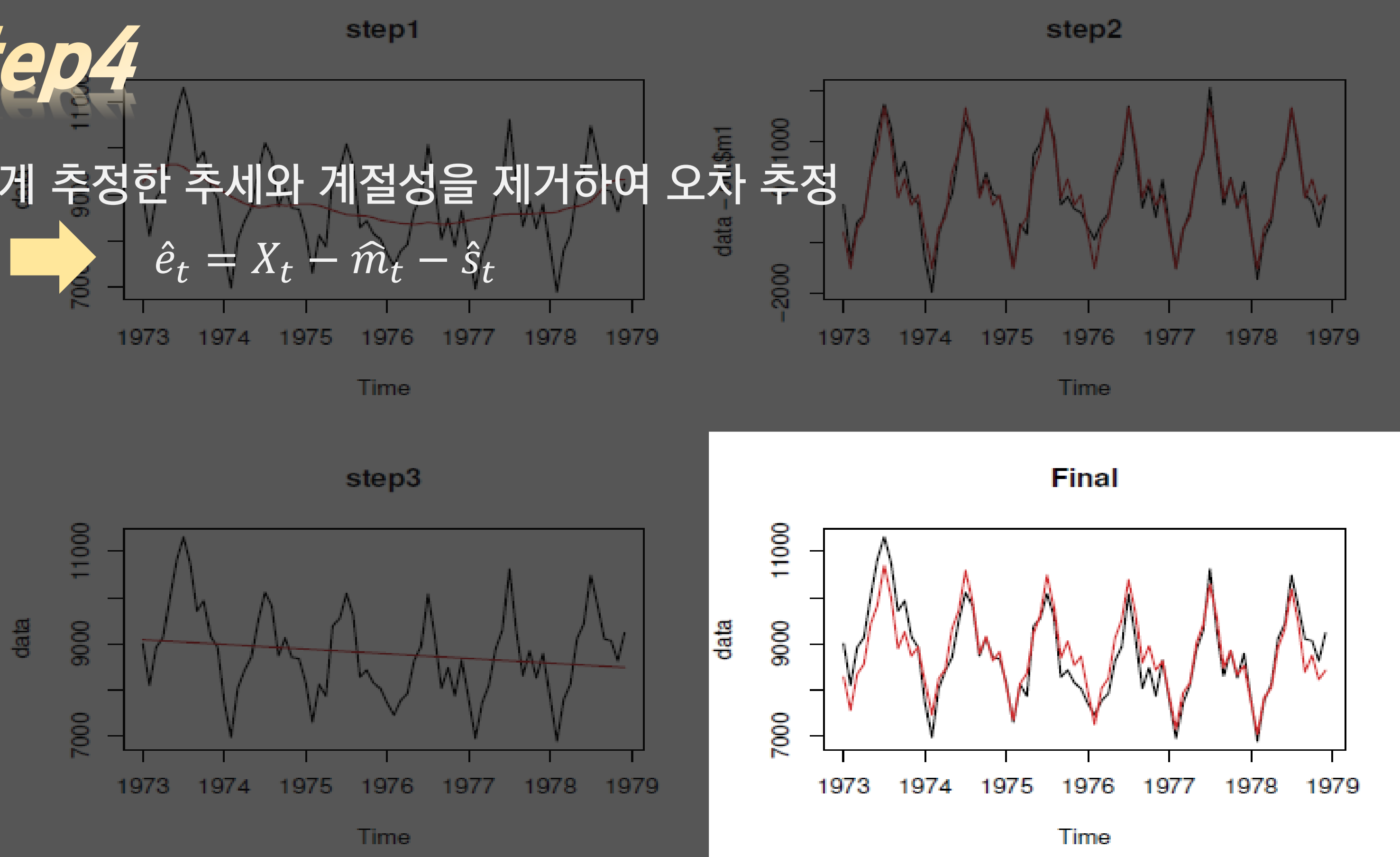
● **평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재**

Classical decomposition algorithm

step4

새롭게 추정한 추세와 계절성을 제거하여 오차 추정

$$\hat{e}_t = X_t - \hat{m}_t - \hat{s}_t$$



평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재

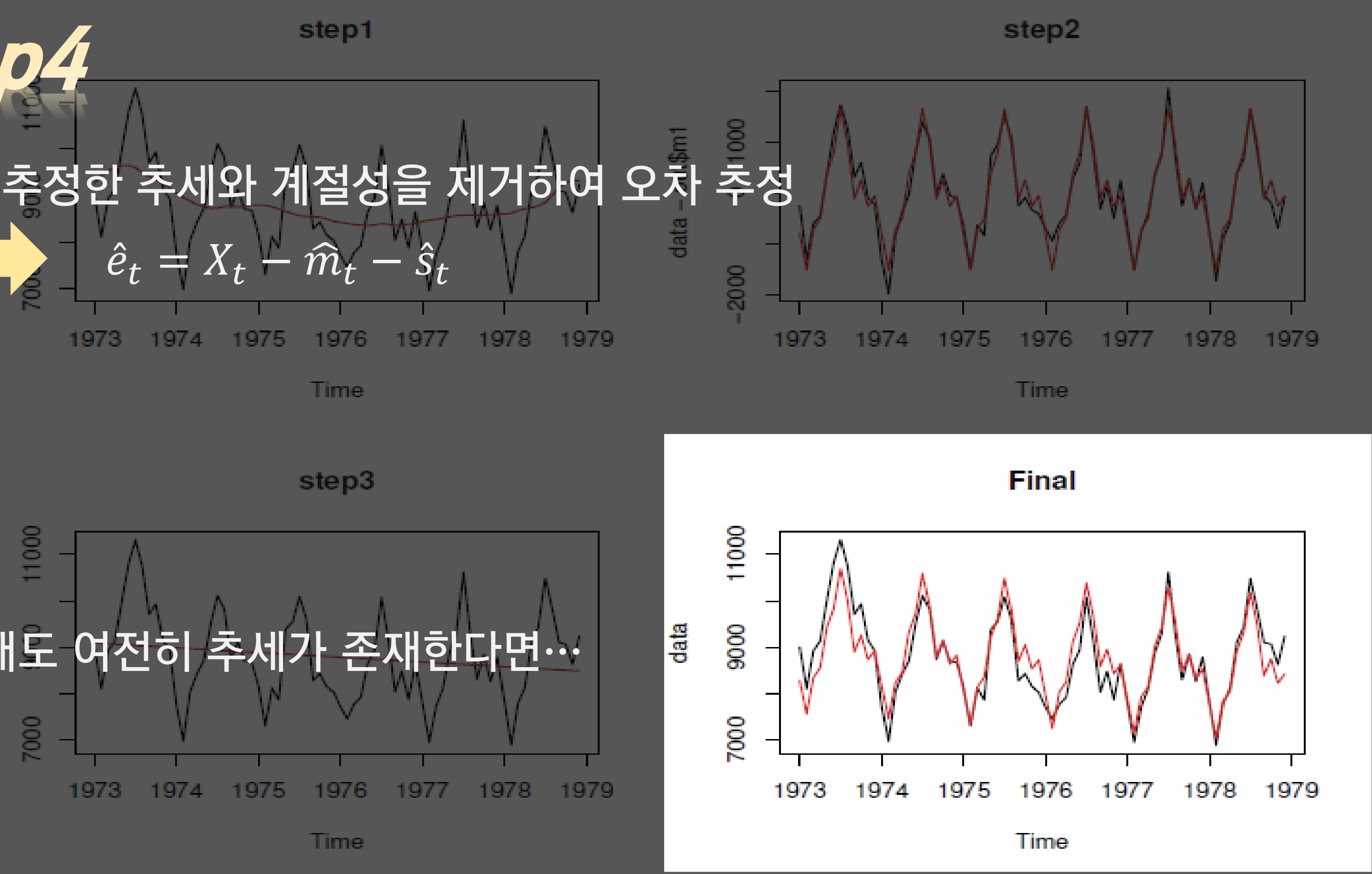
Classical decomposition algorithm

step4

새롭게 추정한 추세와 계절성을 제거하여 오차 추정

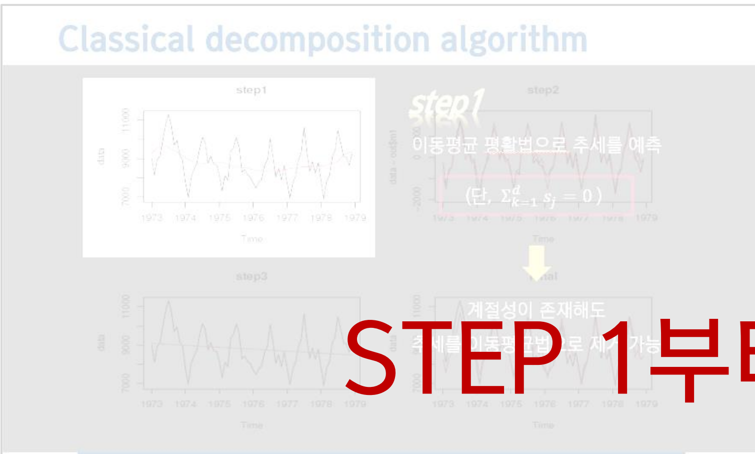
$$\hat{e}_t = X_t - \hat{m}_t - \hat{s}_t$$

그래도 여전히 추세가 존재한다면…

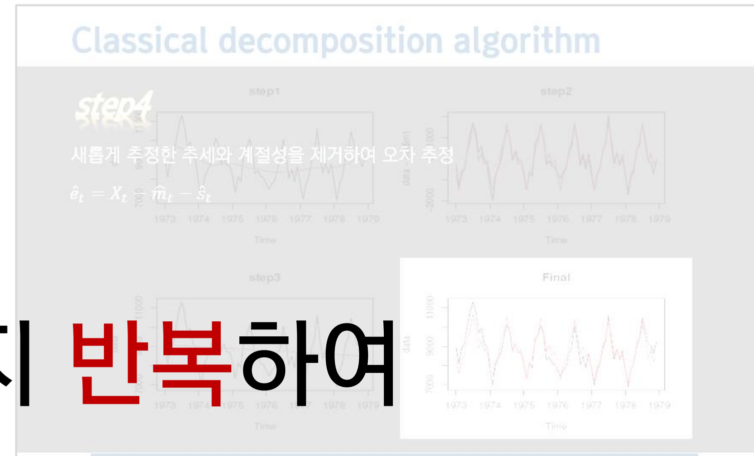


● **평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재**

Classical decomposition algorithm



• • •



STEP 1부터 4까지 반복하여

추세를 다시 제거하기

이거 문병철이 넣었어요 . . .

반복하기로 준이와 약속 ~

● 평균이 일정하지 않은 경우

차분(Differencing)

1차 차분 $\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$

2차 차분 $\nabla^2 X_t = \nabla(\nabla X_t) = \nabla(X_t - X_{t-1})$

$$= X_t - 2X_{t-1} + X_{t-2}$$

$$= (1 - B)^2 X_t$$

특정 시점 자료에서 그 앞에 있는 자료와의 차이

평균이 일정하지 않은 경우

이것도 문병철이 넣었어요. . .

$(1 - B)X_t$

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

한 시점 전으로 돌려주는 작용

후향 연산자(Backshift Operator)

평균이 일정하지 않은 경우 : 추세만 존재

차분(Differencing)

STEP 1

다음과 같은 추세를 가정

$$m_t = (c_0 + c_1 t)$$

평균이 일정하지 않은 경우 : 추세만 존재

차분(Differencing)

STEP 1

다음과 같은 추세를 가정

$$m_t = (c_0 + c_1 t)$$

STEP 2

추세에 1차 차분을 적용

$$\nabla m_t = (c_0 + c_1 t) - \big(c_0 + c_1 (t-1)\big) = c_1$$

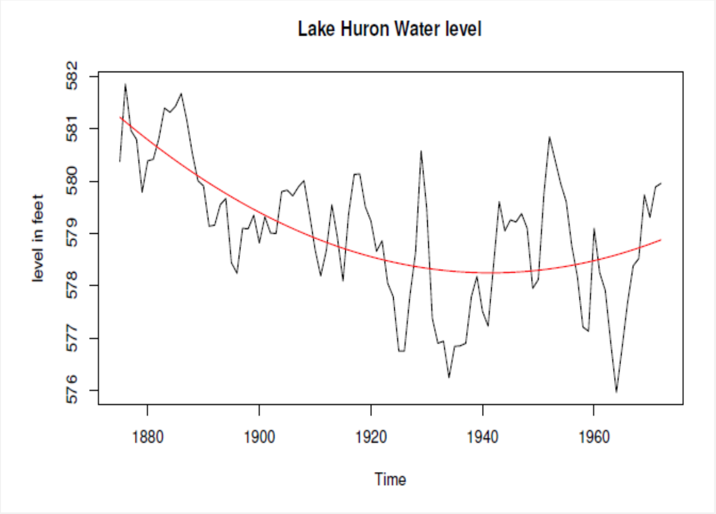
시간의 영향을 받지 않는 상수

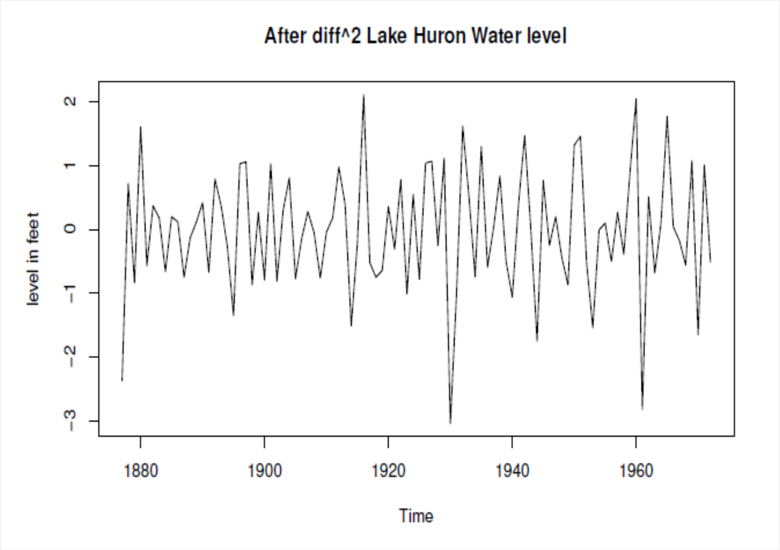
평균이 일정하지 않은 경우 : 추세만 존재

차분(Differencing)

k차 차분을 통해 **k**차 추세 제거 가능

$$\nabla^k X_t = k!\, c_k + \nabla^k Y_t = const + error$$





- 평균이 일정하지 않은 경우 : 계절성만 존재

차분(Differencing)

lag-d differencing을 통해 계절성 제거 가능

$$\nabla_d X_t = (1 - B^d)X_t \,, where\ t = 1, \dots, n$$

이 때, $s_t = s_{t+d}$ 를 만족하면

$$\nabla_d X_t = s_t - s_{t-d} + Y_t - Y_{t-d} = 0 + error$$

오차항만 남았다!

※ 주의 ※ 비슷하지만 너무나도 다르다! 조심! 조심!

$$\nabla_d = (1 - B^d) \quad \neq \quad \nabla^d = (1 - B)^d \ (d - th\ order\ differencing)$$

평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재

차분(Differencing)

STEP 1

추세와 계절성이 있는 모형에서 계절차분 적용

$$X_t = m_t + s_t + Y_t$$

$$\nabla_d X_t = \underbrace{m_t - m_{t-d}}_{\text{추세}} + \underbrace{Y_t - Y_{t-d}}_{\text{오차}}$$

평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재

차분(Differencing)

STEP 1

추세와 계절성이 있는 모형에서 계절차분 적용

$$X_t = m_t + s_t + Y_t$$

$$\nabla_d X_t = \underbrace{m_t - m_{t-d}}_{\text{추세}} + \underbrace{Y_t - Y_{t-d}}_{\text{오차}}$$

STEP 2

남아있는 추세를 제거하기 위해 p차 차분 적용

$$\nabla_d = \left(1 - B^d\right) = \underbrace{(1 - B)}_{\text{1차 차분을 포함}}(1 + B + \cdots + B^{d-1})$$

$\because \ m_t = c_0 + c_1 t + \cdots + c_p t^p$ 인 경우 $\nabla^{p-1} \nabla_d X_t$ 으로 차분을 함

평균이 일정하지 않은 경우 : 추세 및 계절성이 동시에 존재

차분(Differencing)

STEP 1

추세와 계절성이 있는 모형에서 계절차분 적용

$X_t = m_t + s_t + Y_t$

$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$

추세 오차

STEP 2

남아있는 추세를 제거하기 위해 d차 차분 적용

$\nabla_d = (1 - B^d) = (1 - B)(1 + B + \cdots + B^{d-1})$

1차 차분을 포함

$\because \; m_t = c_0 + c_1 t + \cdots + c_p t^p$ 인 경우 $\nabla^{p-1} \nabla_d X_t$ 으로 차분을 함

과대차분(overdifferencing)

이미 정상화가 되어있음에도 차분을 시도

정상성에는 문제가 없으나

ACF를 복잡하게 만들거나 분산이 커지는 문제점

4

정상성 검정

정상시계열

백색잡음(White Noise)

: 자기상관이 존재하지 않는 대표적 정상시계열

$\{X_t\}$가 백색잡음이기 위해서는?

↓

상관관계가 존재하지 않고(uncorrelated) 평균 = 0, 분산 $\sigma^2 < \infty$을 만족

$$\{X_t\} \sim WN(0, \sigma^2)$$

$IID(0, \sigma^2)$는 백색잡음이지만, **그 역은 성립하지 않는다**

정상시계열

백색잡음(White Noise)

내가 해냄 고심이와 함께 백색잡음 개념 정리 ~

기죽지마라!

백색잡음의 세가지 조건 !

- $E(X_t) = 0$ 평균이 0 이고,
- $Var(X_t) = \sigma^2$ 분산이 일정하며,
- $Cov(X_t, X_{t+h}) = 0$ 자기공분산이 존재하지 않는다

✓ $IID(0, \sigma^2)$는 백색잡음이지만, 그 역은 성립하지 않는다

자기공분산함수, 자기상관함수

추세, 계절성 제거 → 정상성을 만족하는 오차만 남아야 함

시간에 따른 상관 정도를 나타내는

자기공분산함수(autocovariance function) 또는 **자기상관계수(autocorrelation function)**을 확인

1. 자기공분산함수 ACVF
(Autocovariance Function)

2. 표본자기공분산함수 SACVF
(Sample Autocovariance Function)

3. 자기상관함수 ACF
(Autocorrelation Function)

4. 표본자기상관함수 SACF
(Sample Autocorrelation Function)

● 자기공분산함수, 자기상관함수

자기공분산함수 ACVF

: Autocovariance Function

$$\gamma_h = Cov(X_t, X_{t+h}) \; = \; E[(X_t - \mu\,)(X_{t+h} - \mu)]$$

이 때, 데이터는 샘플이므로 **추정한 값**을 이용

표본자기공분산함수 SACVF

: Sample Autocovariance Function

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

● 자기공분산함수, 자기상관함수

자기상관함수 ACF

: Autocorrelation Function

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

이 때, 데이터는 샘플이므로 **추정한 값**을 이용

표본자기상관함수 SACF

: Sample Autocorrelation Function

$$\widehat{\rho_k} = \frac{\widehat{\gamma_k}}{\widehat{\gamma_0}}\ ,\ \hat{\rho}(0) = 1$$

백색잡음 검정

만약 !

추세, 계절성 제거 → 남은 오차항이 **WN 또는 IID**

$\gamma(0)$만 추정해주면 됨
즉, 공분산 행렬 추정 불필요!

그러나 !

추세, 계절성 제거 → 남은 오차항이 **WN** 또는 **IID** 불만족

여전히 존재하는 의존성 설명을 위해
복잡한 시계열 모델 필요

백색잡음 검정

독립성 검정

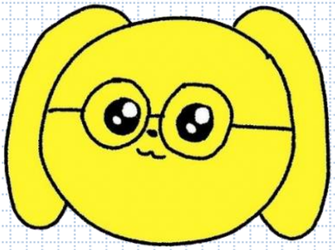
: Correlated vs Uncorrelated

만약 !

오차가 백색잡음 → $\hat{p} \approx \mathcal{N}(0, \frac{1}{n})$

$$H_0 : p(h) = 0 \; vs \; H_1 : p(h) \neq 0$$

$|\hat{\boldsymbol{p}}(\boldsymbol{h})|$가 $\mathbf{1.96}/\sqrt{\boldsymbol{n}}$의 범위 내에 있다면, **오차 간 상관관계가 없다**고 판단



고심이는 오차 간 상관관계를 확인하고 싶어~

1. ACF 그래프를 통해 시각적으로 확인

2. Portmanteau test, Ljung-Box test, McLeod-Li test

백색잡음 검정

정규성 검정

H_0 : 정규성 존재 vs H_1 : 정규성 존재 X

친절한 고심이와 함께 알아보는 정규성 확인 방법~

1. QQ plot
2. Kolmogorov – Sminorv test
3. Jarque – Bera test

왜도, 첨도가 정규분포로 보기에 적합한지에 대한 적합도를 검정

● 백색잡음 검정

정상성 검정

정상성 검정의 3가지 방법

kpss test	ADF test	PP test
H0: **정상성**을 가진다	H0: **비정상성**을 가진다	H0: **비정상성**을 가진다
단위근 검정 방법 중 하나	단위근 검정 방법 중 하나	이분산의 경우에도 사용 가능

이 때, 각 방법의 귀무가설(H0)이 모두 다르므로 주의!

5

다음주 예고입니다~

모형의 필요성

● 다음주 예고

오차항 Y_t의 공분산 행렬

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

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

백색잡음 과정에서는 주대각선의 $\gamma(0)$ (분산) 제외 모두 0

∴ 분산만 추정

즉, 공분산 행렬 추정 불필요!

● 다음주 예고

오차항 Y_t의 공분산 행렬

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

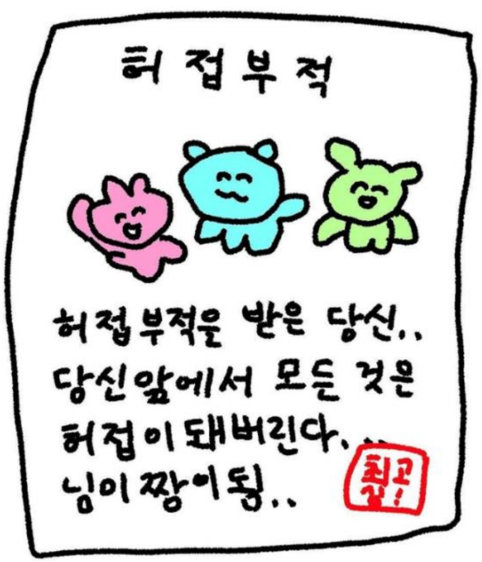
$$= \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix}$$

백색잡음 과정이 아닌 경우 분산 이외의 요소까지 모두 추정 필요

↓

이 때, 특정 모형을 이용하여 추정 가능

THANK YOU



님이 시계열 짱이 됨...