제13장 시계열분석 개요

1. 시계열자료
2. 시계열분석
3. 예측시스템
4. 확률과정의 종류
5. 안정성 및 Wold Decomposition
6. 기술통계량

1. 시계열자료

(1) 시계열자료란?

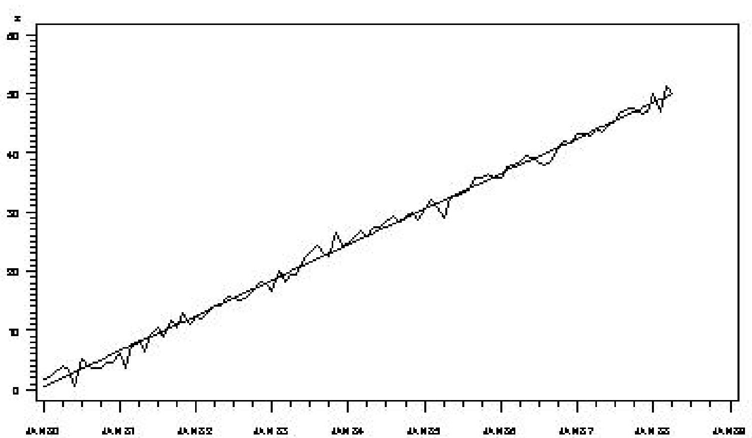
- 다른 시점에서 관측된 값의 계열을 말하며 시계열이라고도 함
- 시간의 흐름에 따라 관측된 자료

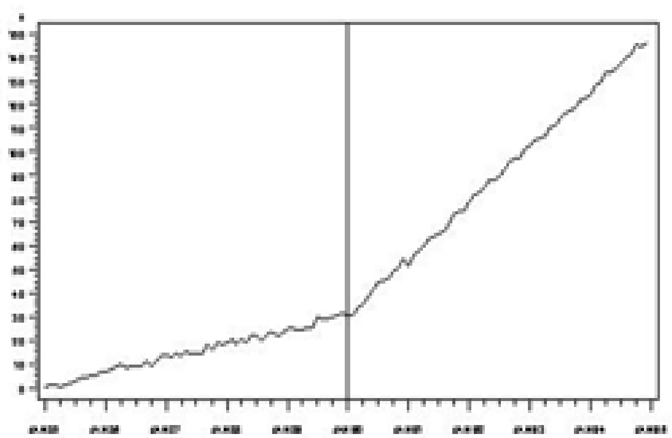
(2) 시계열자료의 예

- 국민총생산, 물가지수, 주가지수 등 경제활동과 관련된 시계열
- 일일 강수량, 기온, 연간 지진 발생 수 등 물리적 현상과 관련된 시계열
- 상품판매량, 상품광고액, 상품재고량 등 경영활동과 관련된 시계열
- 총인구, 농가 수, 인구증가율 등 인구와 관련된 시계열
- 품질관리 등 생산관리와 관련된 시계열
- 월별 교통사고 건수, 월별 범죄발생 수 등 사회생활과 관련된 시계열

(3) 시계열자료의 특징

- 추세성(trend) : 일정기간 같은 방향으로 직선이나 곡선을 따라 상승하거나 하강하는 경향





(참고) 추세는 확정적 추세(deterministic trend)와 확률적 추세(stochastic trend)로 구분

- 확정적 추세 : 시계열의 값이 시간의 함수로 주어짐

  추세를 제거하는 방법은 모수에 관한 선형 또는 비선형모형을 이용(회귀모형)

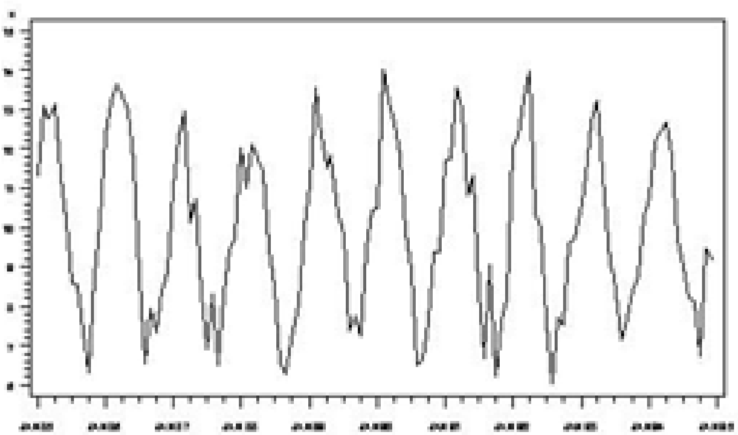
- 확률적 추세 : 시계열의 성장이나 감소율이 자기 과거 값들의 종속관계로 설명됨

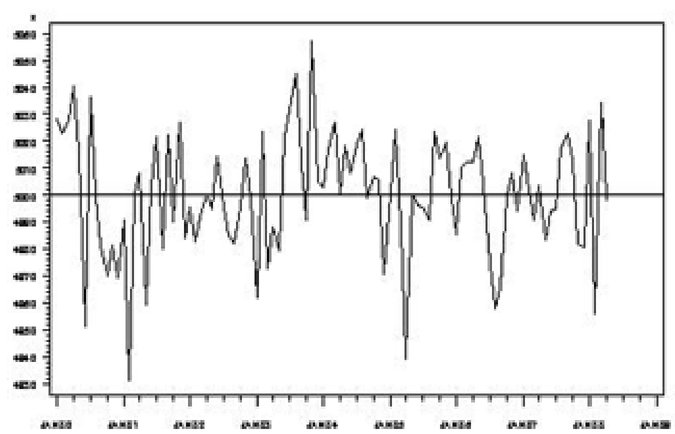
  추세를 제거하는 방법은 적절한 차수의 차분을 이용(차분모형)

– 계절성(seasonality) : 계절에 따른 기온의 변화, 수요와 공급의 변화 등 자연 및 경제현상의 변화에 기인

  계절요인의 제거는 계절차분을 이용

– 불규칙성(irregularity) : 시간에 따른 불규칙적인 움직임이 나타나는 형태

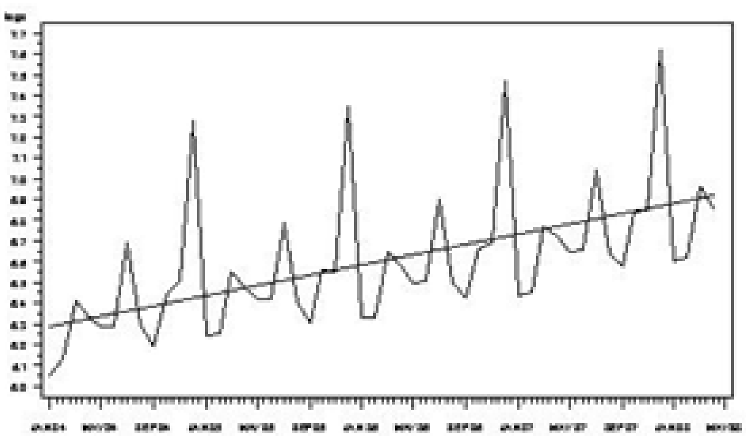




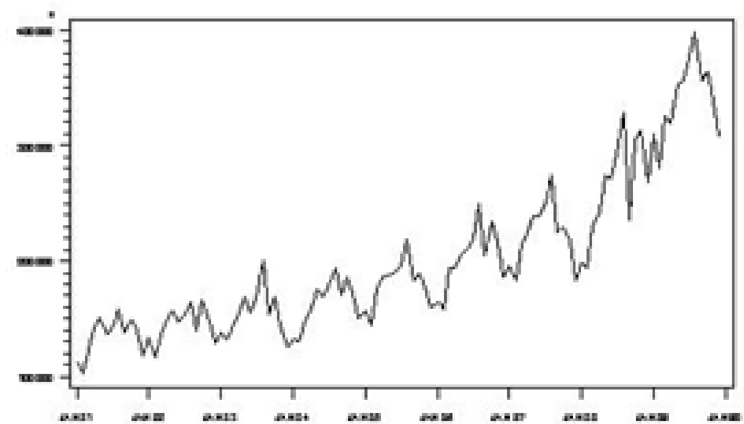
- 변동성(volatility)
  - 시계열의 분산이 시간 추이에 따라 변하는 성질
  - 금융이나 주식시장에서 자산 가치에 영향을 주는 외적 사건의 출현에서 주어지는 시계열의 패턴
  - 경제 또는 금융현상과 관련된 위험(risk)을 설명하는 수단이 됨

- 비선형성(nonlinearity) : 시간에 따라 변하는 값들 사이에 존재하는 종속관계나 인과관계가 자신의 값들 사이의 비선형관계로 설명되는 경우

추세성분과 계절성분을 가진 경우



추세성분과 계절성분을 갖고 시간의 변화에 따라 변동폭이 커지는 경우



- 시계열 자료 = 체계적 성분+불규칙성분 = 추세성분(T_t)+순환성분(C_t)+계절성분(S_t)+불규칙성분(I_t)

$$X_t = T_t + C_t + S_t + I_t$$

$$X_t = T_t \times C_t \times S_t \times I_t$$

· 추세성분 : 관측값이 지속적으로 증가 하거나 감소하는 추세를 갖는 경우의 변동

· 순환성분 : 주기적인 변화를 가지나 변화가 계절에 의한 것이 아니고, 주기가 긴 경우의 변동

· 계절성분 : 계절의 변화에 따른 주기적인 변동

· 불규칙성분 : 시간에 다른 규칙적인 움직임과는 무관하게 랜덤 원인에 의한 변동

2. 시계열분석

(1) 목적

- 시계열 자료가 가지고 있는 시간에 따른 자기종속구조를 파악
- 이 종속구조를 효과적으로 기술하는 모형을 개발하여 미래의 값을 예측

(2) 분석방법의 종류

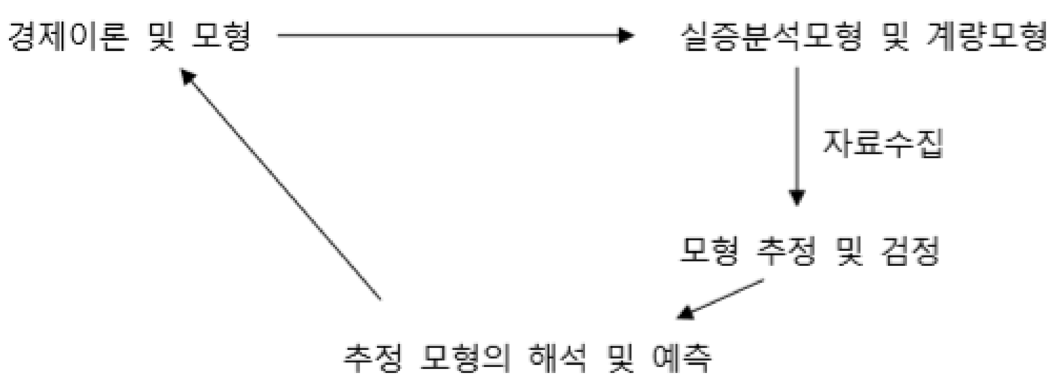
- 시계열분석방법에는 이동평균법, 평활법, 분해법, ARIMA모형 등 다양한 방법이 있음
- 이동평균법: 평균을 취하면서 이동
- 평활법 : 과거의 관측 값들을 가중평균하여 예측 값으로 이용하는 방법
- 분해법 : 시계열을 체계적 성분(추세, 순환, 계절) 및 불규칙 성분으로 분해하여 각 변동을 추정하여 해석
- ARIMA모형 : Box-Jenkins가 제안한 3단계 절차를 이용하는 방법
  - 1단계 : 모형의 식별(model identification)
  - 2단계 : 모형의 추정(model estimation)
  - 3단계 : 모형의 진단(model diagnostic checking)

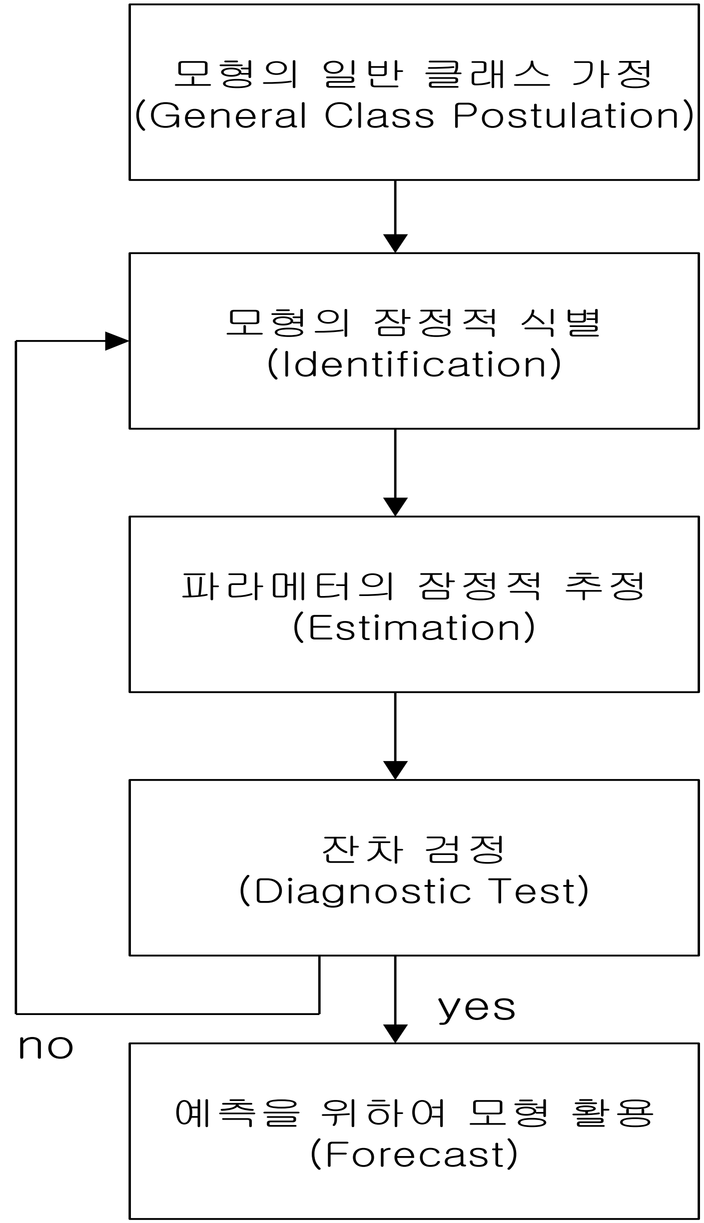
(3) 절차

- 시계열 그래프 작성 : 정상시계열 vs. 비정상시계열, 계절성여부
- 시계열 데이터의 변환 : 비정상시계열(차분), 계절성이 있을 경우(계절조정 또는 계절 차분),
  변동성이 클 경우(대수 변환)
- 자기상관 및 편자기상관 그리기 : AR vs. MA vs. ARMA

3. 예측시스템

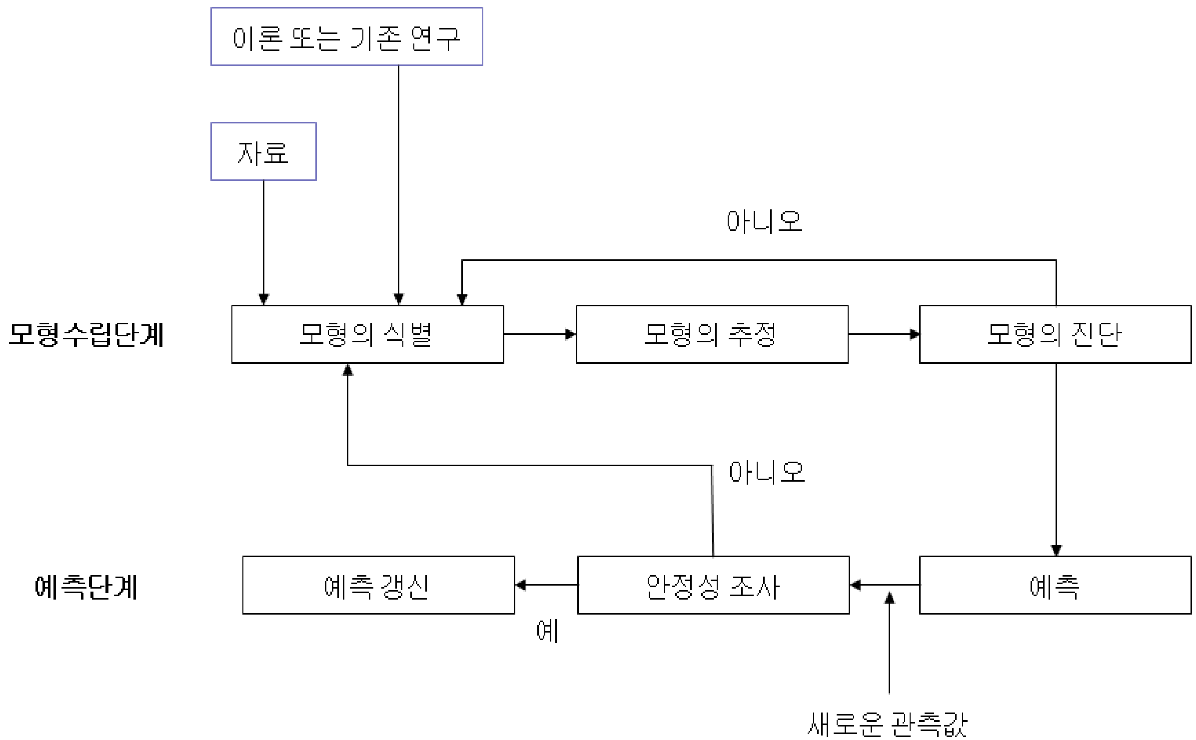
- 예측시스템은 모형수립단계(model building stage) 및 예측단계(forecasting stage)로 구성됨
- 모형수립단계는 모형적합의 3단계와 동일
  - 1단계 : 모형의 식별(model identification)
  - 2단계 : 모형의 추정(model estimation)
  - 3단계 : 모형의 진단(model diagnostic checking)
- Box-Jenkins 5단계
  - 모형의 일반 클래스 가정
  - 모형의 잠정적 식별
  - 모형의 잠정적 추정
  - 잔차 검정
  - 모형을 이용한 예측

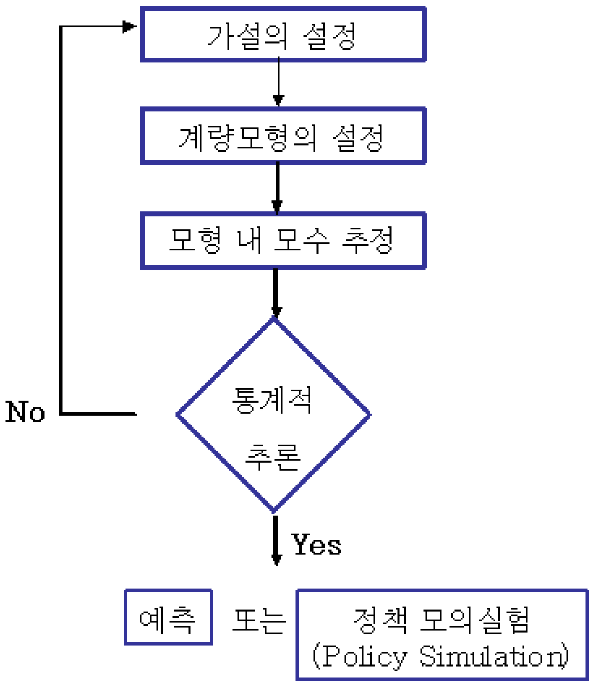


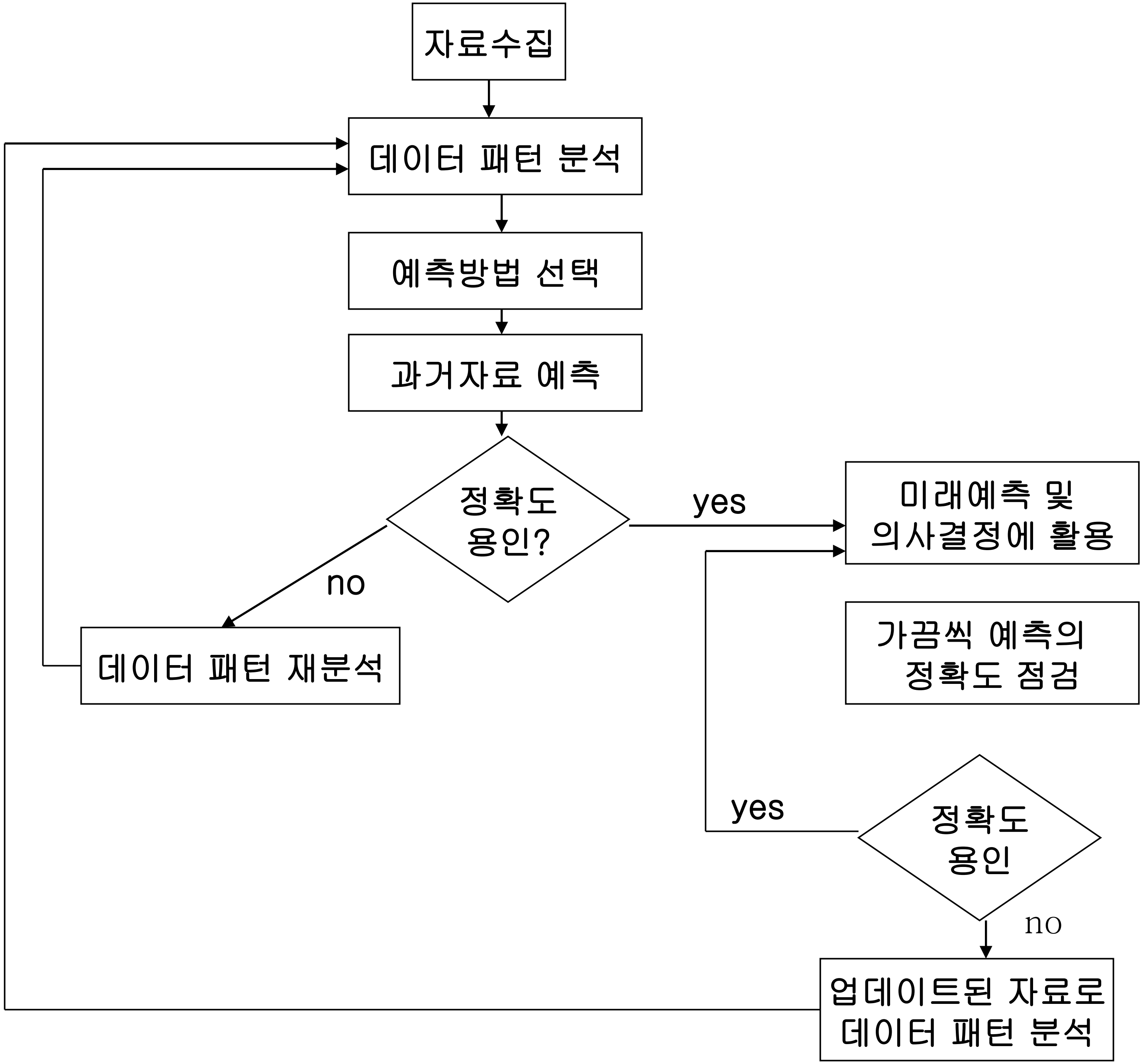


– 예측단계는

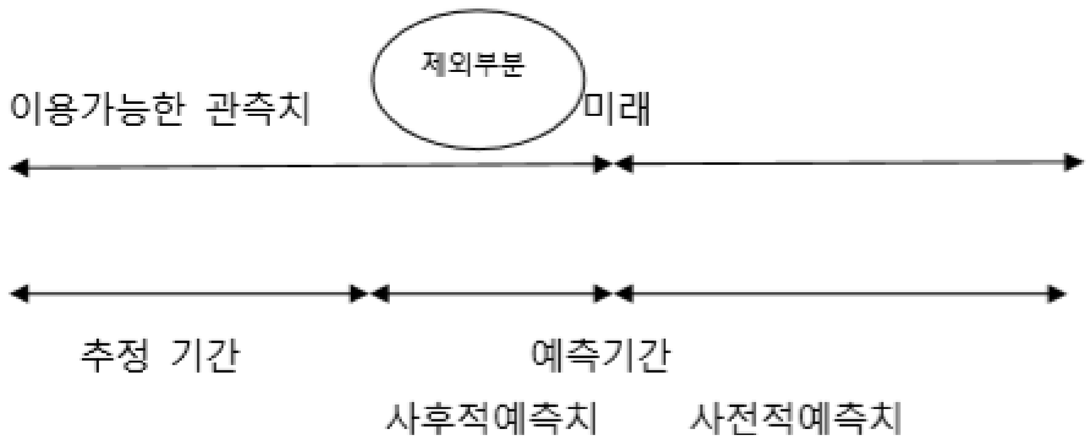
- 모형수립단계를 거쳐 최종적으로 선택된 모형을 이용하여 예측을 한 후,
- 새로운 자료가 관측될 때마다 이를 예측 값과 비교
- 예측시스템 자체의 변화 여부를 통해 안정성을 조사한 후
- 새로운 관측 값과 이전에 구한 예측 값을 이용하여 미래의 예측 값을 갱신

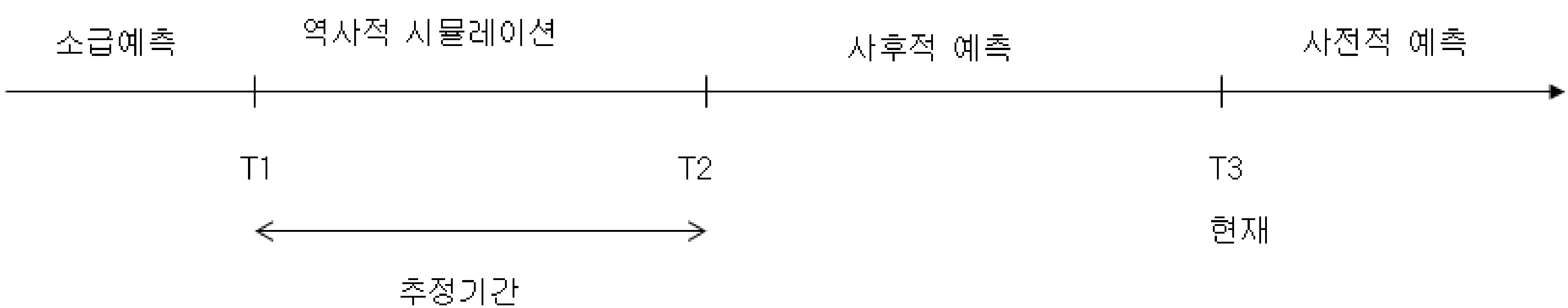





자료수집
데이터 패턴 분석
예측방법 선택
과거자료 예측
정확도 용인?
yes
no
미래예측 및 의사결정에 활용
데이터 패턴 재분석
가끔씩 예측의 정확도 점검
yes
정확도 용인
no
업데이트된 자료로 데이터 패턴 분석

- 미래 예측치의 정확도에 대한 사전평가가 불가능하므로 추정된 예측모형을 이용하여 실험적 예측치 도출
- 사용할 수 있는 관측치 중 마지막 일부를 제외하고 모형 추정, 즉, 실제값을 알 수 있는 기간에 대한 사후 예측치(Ex post Forecast) 도출
- 실제값과 예측치 비교하여 적합성 기준으로 최적 모형 선택
- 최적 모형을 이용하여 사전 예측치(Ex ante Forecast) 도출 → 이후 시간이 경과한 후 실제값과 비교 → 예측 모형에 대한 재평가 → 필요한 경우 새로운 예측 모형 선택





– 예측시 발생하는 예측오차는 예측방법의 신뢰성을 평가하는 기준으로 활용

– 평균오차(ME: Mean of Errors) : 실제값과 예측값의 차이의 합을 예측기간 수로 나눔

$$ME = \frac{\sum_{t=1}^{k} \varepsilon_t}{k} = \frac{\sum_{t=1}^{k}(Y_t - \hat{Y}_t)}{k}$$

– 평균절대오차(MAE: Mean of Absolute Errors) : 실제값과 예측값의 절대값 차이의 합을 예측기간 수로 나눔

$$MAE = \frac{\sum_{t=1}^{k} |\varepsilon_t|}{k} = \frac{\sum_{t=1}^{k} |Y_t - \hat{Y}_t|}{k}$$

– 평균절대값퍼센트오차(MAPE: Mean of Absolute Percentage Errors) : 실제값에 대한 MAE 비율

$$MAPE = \frac{\sum_{t=1}^{k} \frac{|Y_t - \hat{Y}_t|}{Y_t} * 100}{k}$$

– 평균자승오차(MSE: Mean of Squared Errors) : 실제값과 예측값의 차이의 제곱을 합하여 예측기간으로 나눔

$$MSE = \frac{\sum_{t=1}^{k} \varepsilon_t^2}{k}$$

– 평균자승오차제곱근(RMSE: Root Mean of Squared Errors) : MSE의 제곱근

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{t=1}^{k} \varepsilon_t^2}{k}}$$

– Theil의 U통계량 : 분모의 예측값을 현재값 그대로 사용하는 단순예측치와 비교한 상대적인 예측의 정확도

$$\text{Theil's } U = \sqrt{\frac{\sum_{t=1}^{k}\left(\frac{Y_{t+1} - \hat{Y}_{t+1}}{Y_t}\right)^2}{\sum_{t=1}^{k}\left(\frac{Y_{t+1} - Y_t}{Y_t}\right)^2}}$$

U=0(완전예측), U=1(단순예측치 수준 정확도), U<1(단순예측보다 정확), U>1(단순예측보다 부정확)

(1) 확률과정

- T개의 관측치(시계열자료) $y_1,\ y_2, \cdots,\ y_T$ 는 확률변수 $Y_1,\ Y_2, \cdots,\ Y_T$의 실현된 값이며, 확률변수 $Y_1,\ Y_2, \cdots,\ Y_T$는 무한계열의 일부분으로 볼 수 있는데 이 무한한 계열을 확률과정이라고 함

(2) 확률과정의 종류

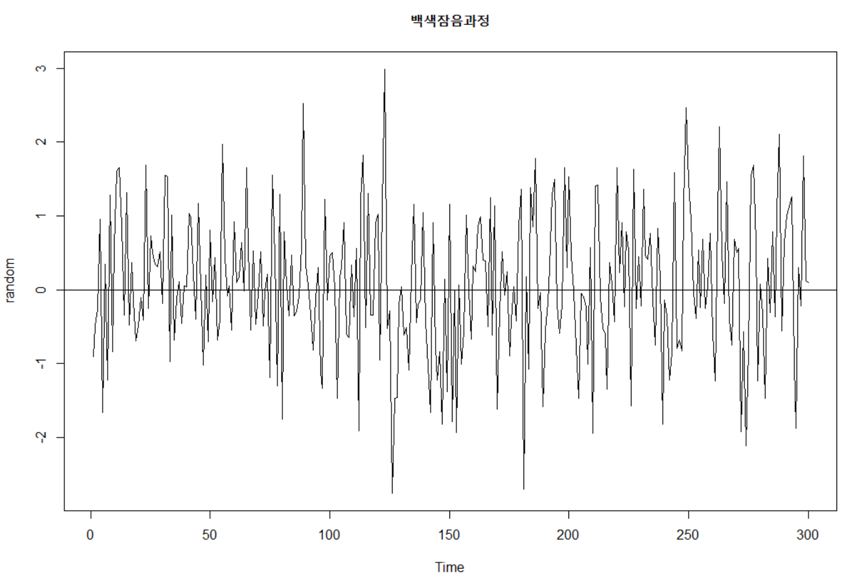
① 백색잡음과정(white noise process)

- $e_1,\ e_2, \cdots, e_T$ 가 있을 경우 다음이 성립하는 확률과정

$E(e_t)=0$ for $\forall\ t$ (zero mean)

$E(e_t e_s)=0$ for $\forall\ t \neq s$ (uncorrelated)

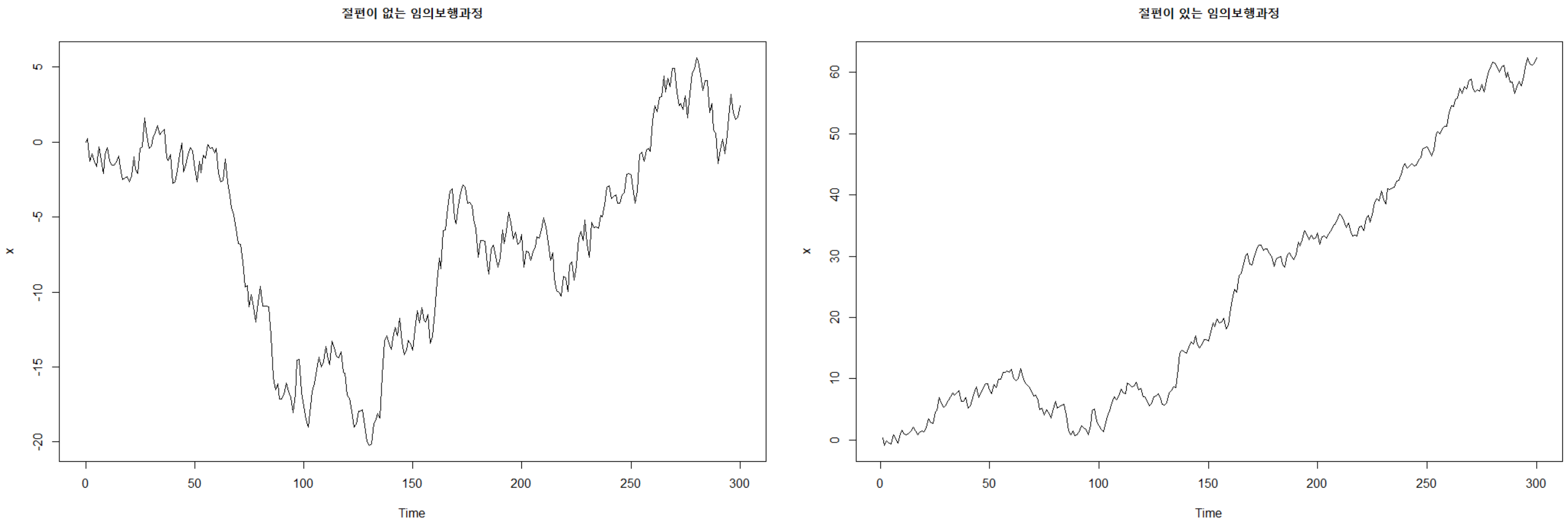
$E(e_1^2)= \ldots E(e_t^2)=\sigma_e^2 < \infty$ (finite variance)



② 임의보행과정(random walk process)

- 절편이 없는 임의보행과정 : $y_t = y_{t-1} + e_t$, $e_t \sim WN\ N(0,1)$

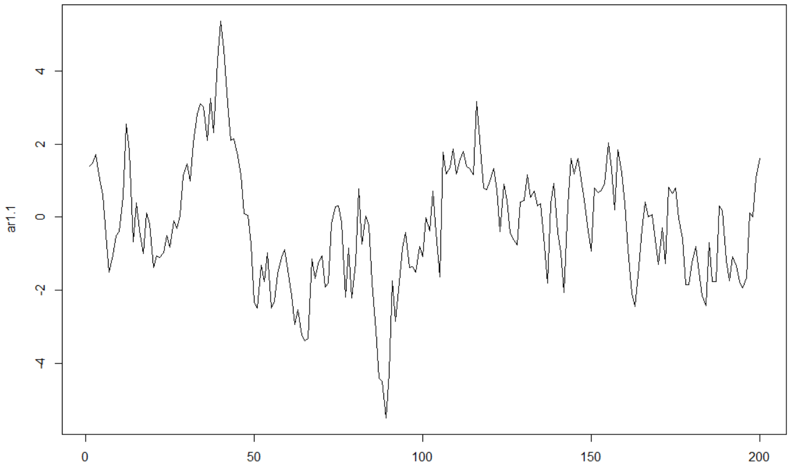
- 절편이 있는 임의보행과정: $y_t = 0.2 + y_{t-1} + e_t$, $y_0 = 0, e_t \sim WN\ N(0,1)$



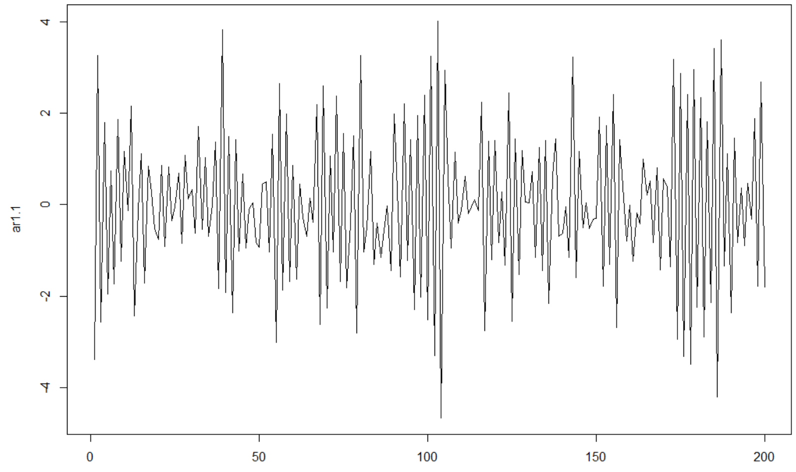
③ 자기회귀과정(AutoRegressive process)

$$y_t = \Phi y_{t-1} + e_t \quad (\mathrm{AR}(1))$$

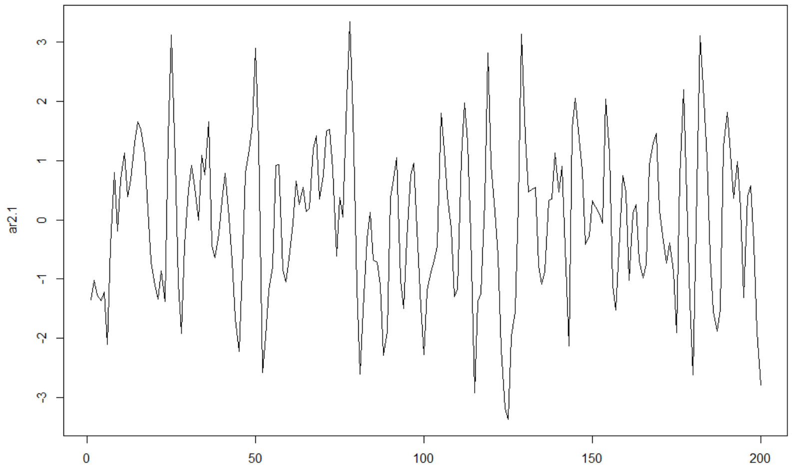
$$y_t = \Phi_1 y_{t-1} + \ldots\ldots + \Phi_p y_{t-p} + e_t \quad (\mathrm{AR}(p))$$



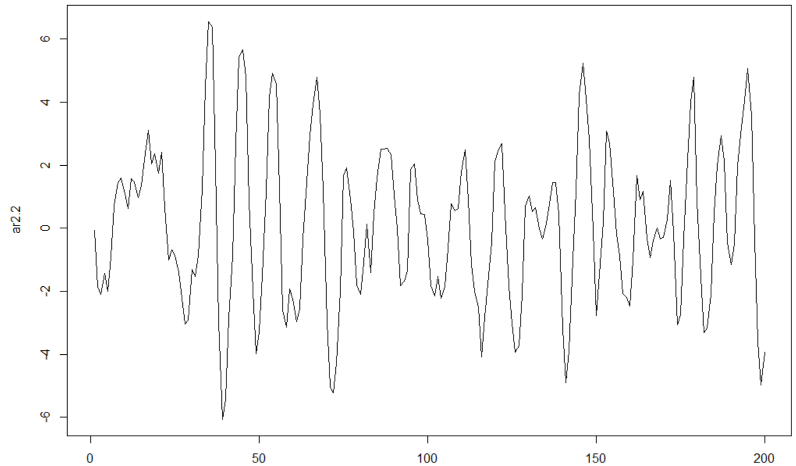
AR(1) : y(t)=0.9*y(t-1)+e(t)



y(t)=-0.9*y(t-1)+e(t)



AR(2) : y(t)=0.9*y(t-1)-0.5*y(t-2)+e(t)

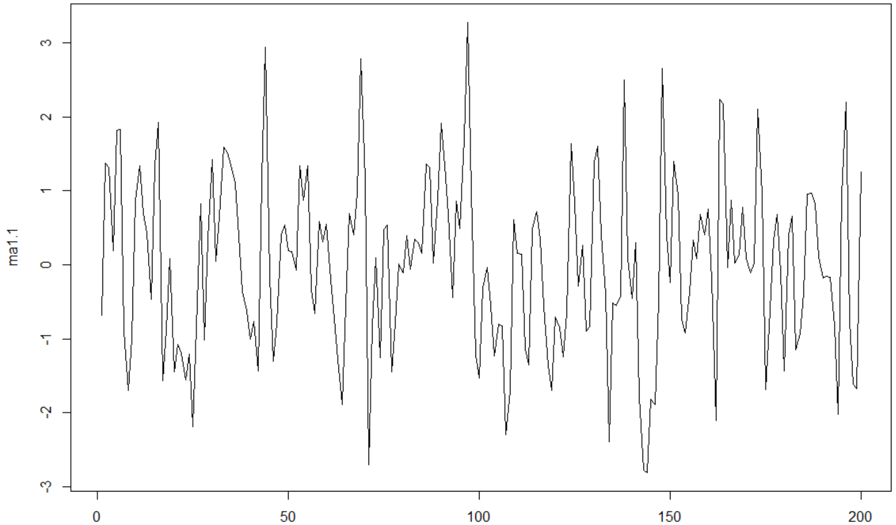


AR(2) : y(t)=1.4*y(t-1)-0.6*y(t-2)+e(t)

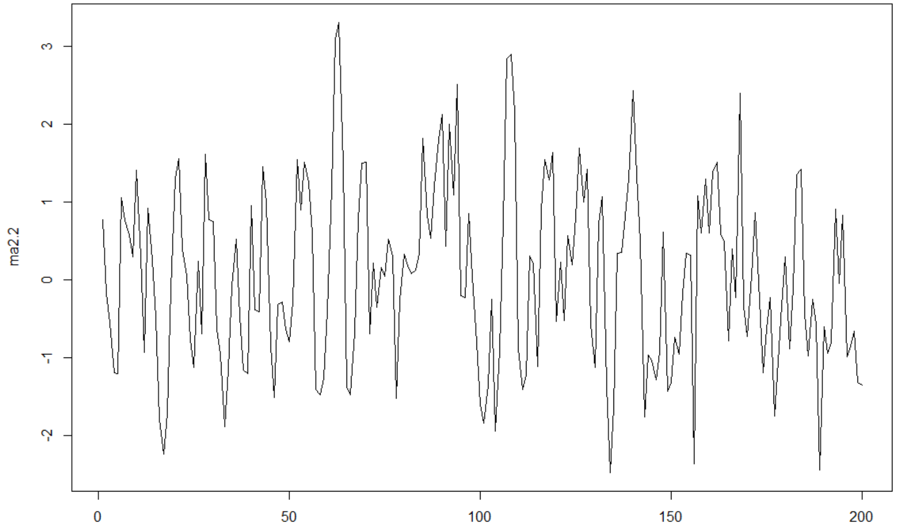
④ 이동평균과정(Moving Average process)

$$y_t = e_t - \theta e_{t-1} \quad \text{(MA(1))}$$

$$y_t = e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q} \quad \text{(MA(q))}$$



MA(1) : y(t)=e(t)+0.8*e(t-1)



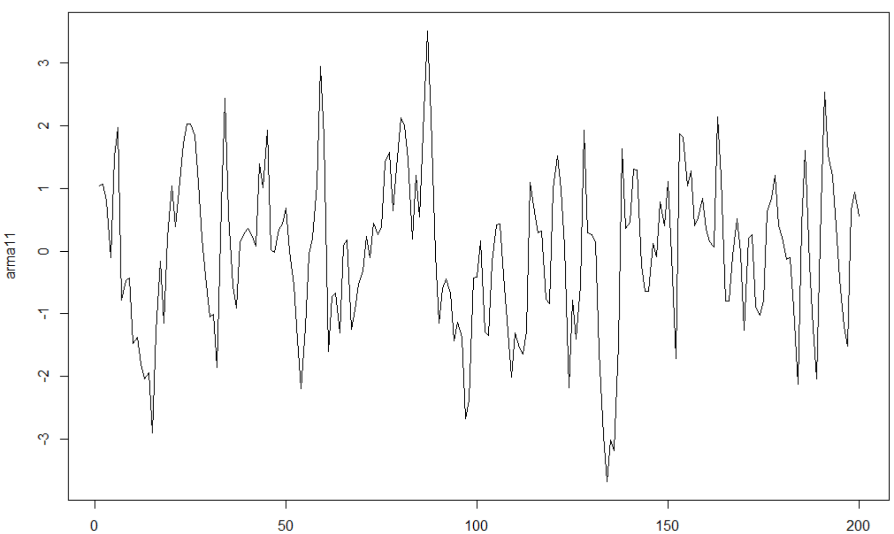
MA(2) : y(t)=e(t)+0.6*e(t-1)+0.3*e(t-2)

⑤ 자기회귀이동평균과정(ARMA process)

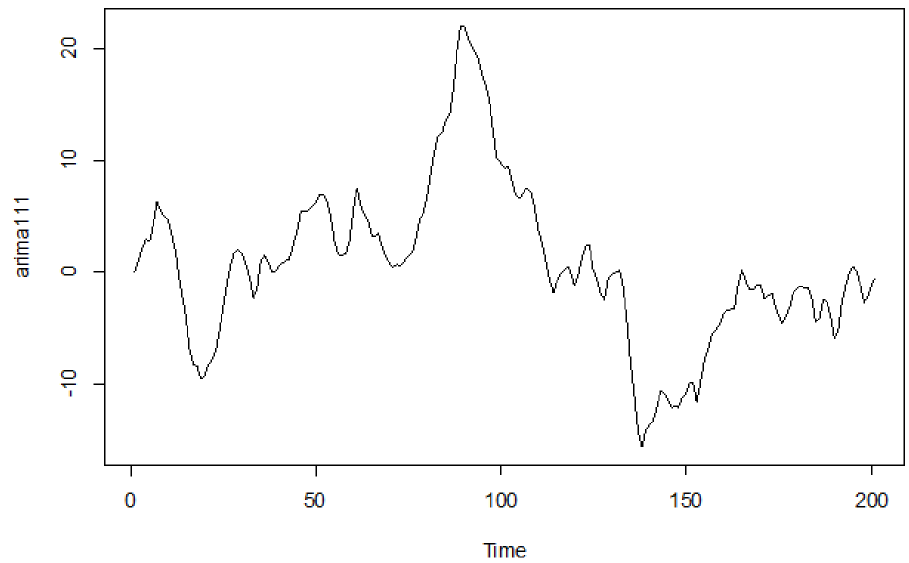
$$y_t - \phi_1 y_{t-1} + \ldots\ldots - \phi_p y_{t-p} = e_t + \theta_1 e_{t-1} \ldots\ldots + \theta_q e_{t-q} \quad (\mathrm{ARMA}(p,q))$$

⑥ 자기회귀적분이동평균과정(ARIMA process)

$$\Delta y_t - \phi_1 \Delta y_{t-1} + \ldots\ldots - \phi_p \Delta y_{t-p} = e_t + \theta_1 e_{t-1} \ldots\ldots + \theta_q e_{t-q} \quad (\mathrm{ARIMA}(p,d,q))$$



ARMA(1,1) : y(t)=0.5*y(t-1)+0.4*e(t-1)+e(t)



ARIMA(1,1,1) : $\Delta y_t = \Delta 0.5 y_{t-1} + 0.4 e_{t-1} + e_t$

5. 안정성 및 Wold Decomposition

(1) 안정성

- 다음의 3조건을 만족하는 확률과정, y_t 를 안정적 시계열(covariance-stationary)이라 함
  - 시계열의 평균이 일정 즉, $E(y_t) = E(y_{t-s}) = \mu, \text{for } \forall\, t, t-s$
  - 시계열의 분산이 일정 즉, $E[(y_t - \mu)^2] = E[(y_{t-s} - \mu)^2] = \sigma_y^2$
  - 시계열의 공분산이 일정 즉, $E[(y_t - \mu)(y_{t-s} - \mu)] = E[(y_{t-j} - \mu)(y_{t-j-s} - \mu)] = \gamma_s$
- 시계열의 평균이 일정하지 않을 경우를 homogeneous non-stationarity라 하며, 이러한 경우는 원계열을 차분(differencing)을 해주면 안정시계열이 됨
  - 차분=1:원계열의 수준이 시간에 따라 다를 경우/차분=2 : 원계열의 수준과 기울기가 시간에 따라 다를 경우
- 시계열의 분산이 일정하지 않을 경우를 non-stationary variance라 하며, 이러한 경우 원계열에 자연대수(natural logarithm)를 취하면 됨
- 일반적으로 성장률의 개념으로 생각하는 시계열(예 : GDP, Money, Price)은 자연대수를 취하는데 이러한 시계열은 자연대수를 취한 후 차분여부를 결정해야 함

(예) 임의보행과정의 불안정성

$$y_t = y_{t-1} + e_t$$
$$= y_{t-2} + e_{t-1} + e_t$$
$$\cdot$$
$$y_0 + \sum_{j=1}^{t} e_j$$

만약에 y_0를 주어진 것으로 가정하면, 즉 $y_0 = 0$ 이라 하면

$$E(y_t) = 0$$
$$E(y_t^2) = E[(e_t + e_{t-1} \ldots + e_1)(e_t + e_{t-1} \ldots + e_1)] = t\sigma_e^2 \quad \text{(dependent on t)}$$
$$E(y_t y_{t-1}) = E[(e_t + e_{t-1} \ldots + e_1)(e_{t-1} + e_{t-2} \ldots + e_1)] = (t-1)\sigma_e^2 \quad (\;''\;)$$

(2) Wold Decomposition

- 모든 안정적인 확률과정은 다음과 같이 확정적 요소(deterministic component)와 확률적 요소(stochastic component)의 합으로 나타낼 수 있음

  $\boldsymbol{Y_t} = \boldsymbol{\mu} + \boldsymbol{e_t} + \boldsymbol{\Psi_1}\, \boldsymbol{e_{t-1}} + \boldsymbol{\Psi_2} \boldsymbol{e_{t-2}} + \cdots$ 단, $\boldsymbol{e_t}$ ~i.i.d(0,$\boldsymbol{\sigma^2}$)

  -- ----------------------------------

  (평균) (현재 및 과거의 충격)

- 현재의 $\boldsymbol{Y_t}$는 평균($\boldsymbol{\mu}$)과 현재 및 과거의 충격(shock)으로 나타낼 수 있는데 이를 Wold Representation이라고 함
- 충격의 백색잡음과정의 가정 외에 $Var(\boldsymbol{Y_t}) = \boldsymbol{\sigma^2} + \boldsymbol{\psi_1^2}\, \boldsymbol{\sigma^2} + \cdots = \boldsymbol{\sigma^2} \sum_{i=0}^{\infty} \boldsymbol{\Psi_i^2} < \infty$ (finite variance)를 가정

(1) 자기상관함수

– 안정적인 확률과정 y_t의 평균, 분산, 자기공분산, 자기상관, 편자기상관은 다음과 같이 추정

· 표본평균 : $\hat{\mu} = \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$

· 표본분산 : $\hat{\gamma}_0 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2$

· 표본자기공분산 : $\hat{\gamma}_k = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})$

· 표본자기상관함수 : $\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$

· 표본편자기상관함수 :

$$\hat{\phi}_{11} = \hat{\rho}_k$$

$$\hat{\phi}_{kk} = \frac{\hat{\rho}_k - \sum_{j=1}^{k-1}\hat{\phi}_{k-1,j}\hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1}\hat{\phi}_{k-1,j}\hat{\rho}_j} \quad , \ k=2, 3, ...$$

단, $\hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk}\hat{\phi}_{k-1,k-j}$, k=3, 4, ... , j= 1, 2, 3, ...k-1

(2) 편자기상관함수

- k 이외의 모든 시차를 갖는 관측치(y_{t-1}, $y_{t-2}, \cdots, y_{t-k+1}$)의 영향력을 배제한 가운데 특정의 두 관측치, y_t 및 y_{t-k}가 얼마나 관련이 있는 지 나타내는 척도로 회귀계수가 편자기상관함수가 됨
- 즉, $\phi_{kk} = corr(y_t, y_{t-k} \mid y_{t-1}, y_{t-2}, \cdots, y_{t-k+1})$
- 시차 k의 편자기상관계수는 다음 식에서 k번째 회귀계수 ϕ_{kk}를 의미함

$$y_t = \phi_{k1} y_{t-1} + \phi_{k2} y_{t-2} + \ldots + \phi_{kk} y_{t-k} + e_t$$

즉, AR(1): $\phi_{11} = \phi_1$

AR(2): $\phi_{22} = \phi_2$

- 표본편자기상관함수 :

$$\widehat{\phi_{11}} = \widehat{\rho_k}$$

$$\widehat{\phi_{kk}} = \frac{\widehat{\rho_k} - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_j}, \quad k=2, 3, \ldots\ldots$$

단, $\hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk} \hat{\phi}_{k-1,k-j}$, k=3, 4, . . . , j= 1, 2, 3, . . . k-1

(3) 교차상관계수 및 시차상관계수

- 교차 및 시차상관계수는 t기의 특정(기준)변수 x의 값(x_t)과 t+k기에 관찰된 y값(y_{t+k}) 간의 상관관계의 정도를 나타냄
- k=0인 경우 즉, γ_0인 경우를 교차상관계수(cross correlation coefficient)라고 하고, k≠0인 경우 를 시차상관계수(leads and lags correlation라고도 함

$$\gamma_k = \frac{\sum_{t=1}^{N-k}(x_t-\bar{x})(y_{t+k}-\bar{y})}{\sqrt{\sum_{t=1}^{N}(x_t-\bar{x})^2\sum_{t=1}^{N}(y_t-\bar{y})^2}}, \quad k=0,\pm1,\pm2,\ldots,\pm k$$

- 교차상관계수 해석
  - $\gamma_0 > 0$: 두 변수들이 서로 같은 방향으로 변화(pro-cyclical:경기순응)
  - $\gamma_0 < 0$: 두 변수들이 서로 반대 방향으로 변화(counter-cyclical:경기역행)
  - $\gamma_0 = 0$: 두 변수들이 서로 경기중립적
- 시차상관계수 해석
  - γ_k의 값이 최대가 되는 시차 k가 양(+)이면 해당변수 y_t는 x_t의 후행지표
  - γ_k의 값이 최대가 되는 시차 k가 음(-)이면 해당변수 y_t는 x_t의 선행지표
  - γ_k의 값이 최대가 되는 시차 k가 0이면 해당변수 y_t는 x_t와 동행지표

연도	2001	.	.	2010
GDP(x_t)	x_1	.	.	x_{10}
해당변수(y_t)	y_1	.	.	y_{10}

  - 만약에 k=2에서 γ_k의 값이 최대이면, $\gamma_2 = x_1y_3 + x_2y_4 + \cdots + x_8y_{10}$ (y가 GDP 뒤따라 변하는 후행지표)
  - 만약에 k=-2에서 γ_k의 값이 최대이면, $\gamma_{-2} = x_3y_1 + x_4y_2 + \cdots + x_{10}y_8$ (y가 GDP보다 먼저 변하는 선행지표)

(실증분석 예 : 우리나라 경기변동의 실증분석)

- 표준편차 : 변동성(volatility)측정
- 상대적 표준편차 : GDP에 대한 상대적 변동성 측정
- 자기상관계수(autocorrelation) : 경기변동의 지속성 측정
- 교차상관계수(cross correlation): GDP와의 동행성(comovement) 측정
  - · 경기순응적(pro-cyclical) / 경기중립적(a-cyclical) / 경기역행적(counter-cyclical)
- 시차상관계수(leads and lags correlation):GDP에 대한 선행성 측정
  - · 경기선행적(leading) / 경기동행적(coincident) / 경기후행적(lagging)

○대상기간 : 1970: I ~1997:III (111분기)

주요변수의 변동폭 및 시차상관계수

변 수	표준편차	상대적 표준편차	GDP와의 시차상관계수								
			-4	-3	-2	-1	0=γ_0	+1	+2	+3	+4
G D P	2.19	1.00	0.20	0.33	0.51	0.68	1.00	0.68	0.51	0.33	0.20
소 비	1.01	0.46	0.01	0.16	0.30	0.50	0.70	0.66	0.59	0.41	0.27
투 자	8.01	3.65	0.12	0.33	0.53	0.67	0.70	0.62	0.51	0.40	0.20
수 출	7.17	3.27	0.27	0.37	0.44	0.43	0.35	0.24	0.06	-0.11	-0.25
수 입	5.97	2.72	-0.02	0.18	0.44	0.65	0.66	0.58	0.51	0.29	0.22
무역수지	2.83	1.29	0.40	0.37	0.31	0.19	0.12	0.00	-0.13	-0.22	-0.37
자본스톡	0.88	0.40	0.15	0.20	0.16	0.15	0.09	0.05	0.02	-0.02	-0.05
노동투입	1.59	0.72	0.14	0.17	0.23	0.29	0.35	0.30	0.27	0.15	0.00
교역조건	5.61	2.56	0.36	0.41	0.50	0.52	0.48	0.36	0.25	0.08	-0.03