비정상적 시계열

시계열의 비정상성 (nonstationarity)

- 시계열에 추세 또는 계절성이 포함되는 경우 정상성을 만족하지 못한다.

- 비정상성 판단 방법
  - 시계열의 시간에 대한 그래프를 보고 시각적으로 판단
  - ACF가 시차에 대하여 매우 서서히 감소하는 패턴
  - 단위근 검정

단위근이 있다는 가설 기각 → 정상적 시계열

기각 X → 비정상적 시계열

단위근이 있으면 비정상적 시계열

* 단위근: $x^n = 1$을 만족하는 근 (복소수 포함)

$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + \varepsilon_t$

↳ 특성방정식

$m^p - m^{p-1} a_1 - m^{p-2} a_2 - \cdots - a_p = 0$

$\mathrm{abs}(m) = 1 \Rightarrow$ 단위근이 있다.

- 비정상성 대응 방안
  - 차분 (differencing)을 통해 정상적 시계열로 변환
  - 함수변환을 통하여 분산 안정화
  - 분해법으로 추세 및 계절성 제거

* 원시계열 : 분산 증가 패턴 ⇒ 로그 함수 변환

ARIMA 모형

차분 (differencing)

$Z_{t-1} = BZ_t$

- 1차 차분: 인근한 두 값의 차이를 산출

$$\Delta Z_t = Z_t - Z_{t-1} = (1-B)Z_t$$

- 2차 차분

$$\Delta^2 Z_t = \boxed{\Delta(\Delta Z_t)} = \Delta(Z_t - Z_{t-1}) = (Z_t - Z_{t-1}) - (Z_{t-1} - Z_{t-2}) = Z_t - 2\underbrace{Z_{t-1}}_{BZ_t} + \underbrace{Z_{t-2}}_{B^2 Z_t}$$
$$= (1 - 2B + B^2)Z_t = (1-B)^2 Z_t$$

- d차 차분 ($d = 1,2,\ldots$)

$$\Delta^d Z_t = (1-B)^d Z_t$$

ARIMA 모형

- 차수 d 누적시계열 (integrated process of order d)
  - d차 차분 후 시계열이 처음으로 정상적일 때, 원 시계열을 차수 d 누적시계열이라 하고 $I(d)$ 로 표기

누적.

- ARIMA (autoregressive integrated moving average) 모형
  - d차 차분한 시계열이 정상적 ARMA(p,q)모형을 따를 때, 원 시계열이 ARIMA(p,d,q)모형을 따른다고 함.
  - 차분시계열 $W_t = (1-B)^d Z_t$
  - 차분시계열이 ARMA(p,q): $\phi_p(B)W_t = \theta_q(B)a_t$
  - 즉, 다음이 성립

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)a_t$$

원시계열 $\xrightarrow{\text{d차 차분}}$ 차분시계열

Z_t ~ ARIMA(p,d,q)　　W_t ~ ARMA(p,q)

※ $Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \cdots - \phi_p Z_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$

$\Rightarrow \Phi_p(B) Z_t = \Theta_q(B) a_t$

ARIMA 모형

(예) ARIMA(1,1,1) 모형

- 차분 시계열: $W_t = (1-B)Z_t \sim ARMA(1,1) \Rightarrow W_t = \phi_1 W_{t-1} + a_t - \theta_1 a_{t-1}$
- 원 시계열: $Z_t \sim ARIMA(1,1,1) \Rightarrow (1-\phi_1 B)(1-B)Z_t = (1-\theta_1 B)a_t$

$$\Rightarrow Z_t = (1+\phi_1)Z_{t-1} - \phi_1 Z_{t-2} + a_t - \theta_1 a_{t-1}$$

(예) ARIMA(0,1,1) 또는 IMA(1,1)

- 차분 시계열: $W_t = (1-B)Z_t \sim MA(1) \Rightarrow W_t = a_t - \theta_1 a_{t-1}$
- 원 시계열: $Z_t \sim ARIMA(0,1,1) \Rightarrow Z_t = Z_{t-1} + a_t - \theta_1 a_{t-1}$

$(1-\phi_1 B) = \phi_p(B)$ $\quad (1-\theta_1 B) = \theta_1(B)$

$-(1+\phi_1)Z_{t-1} \quad + \phi_1 Z_{t-2}$

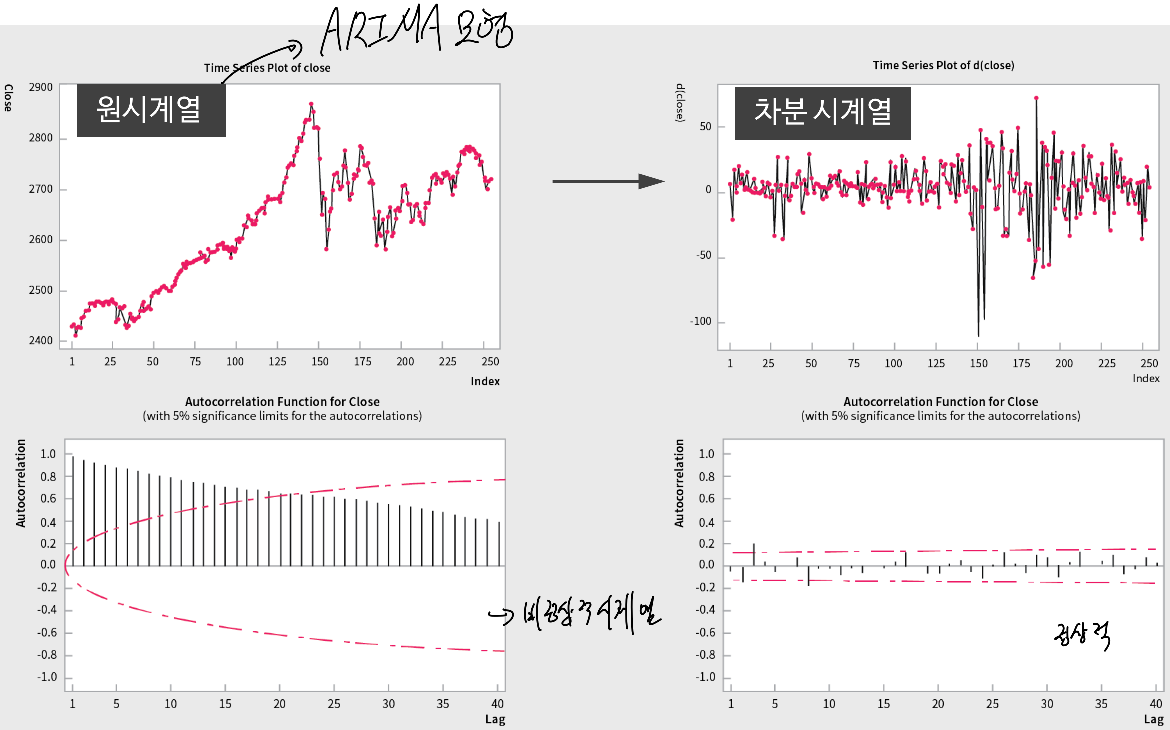
$Z_t - \phi_1 B Z_t - B Z_t + \phi_1 B^2 Z_t$

$= a_t - \theta_1 \underbrace{B a_t}_{a_{t-1}}$

ARIMA 모형

예 (S&P 200 지수) 아래 그림은 2017년 7월 3일 부터 2018년 6월 29일 까지 미국 S&P 200 지수의 종가를 나타낸다.

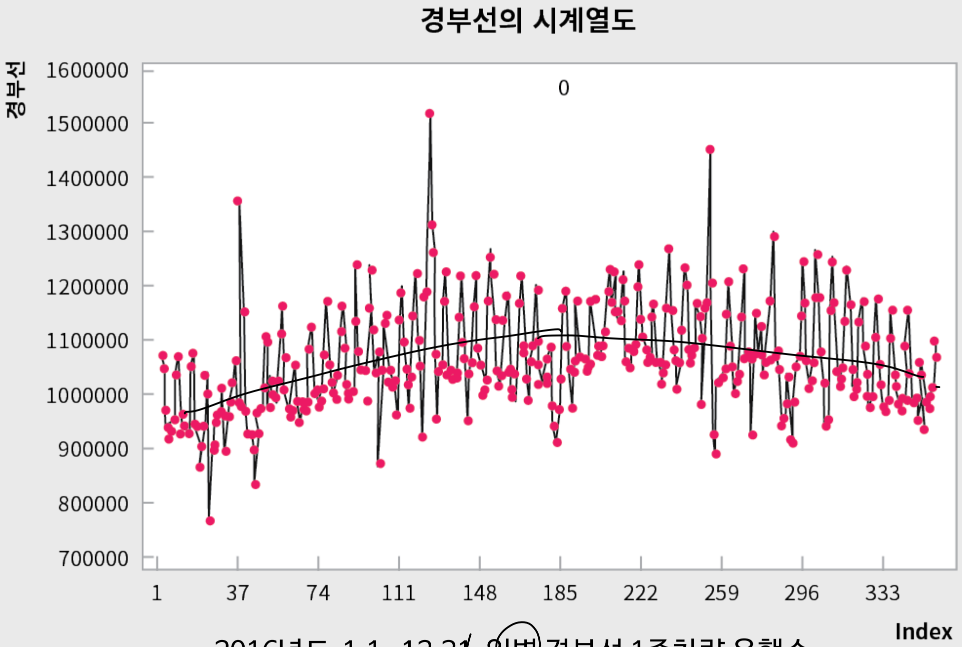
원 시계열과 1차 차분 시계열에 대한 ACF를 살펴보자



#1

계절성 시계열

- 일반적 시계열에는 추세 와 계절성이 동시에 존재하는 경우가 많다.
- 추세는 차분으로 제거될 수 있으나 계절성은 여전히 남을 수 있다.
- 이전 ARIMA모형은 비계절성에 대한 것이며 계절성은 별도로 처리하여야 한다.
- 즉 일반적 시계열은 비계절성 ARIMA모형과 계절성 ARIMA모형이 복합된 형태이다.



2016년도 1.1~12.31 일별 경부선 1종차량 운행수

일주일

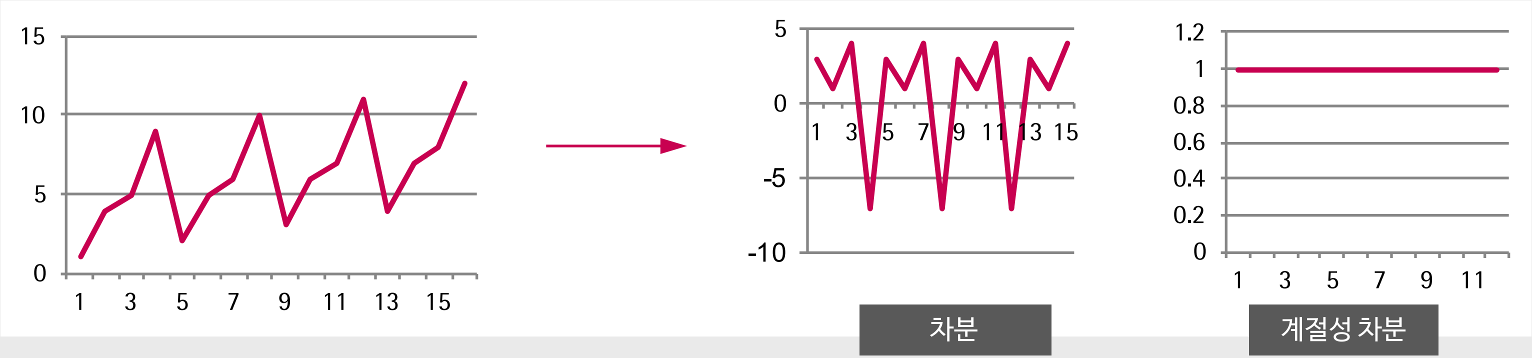
계절성 ARIMA 모형

계절성 차분 (seasonal differencing)

- 계절성 주기 s (월별데이터: s=12; 분기별데이터: s=4)
- 계절성이 있는 경우 단순 (비계절성) 차분으로는 정상화가 되지 않음
- 1차 계절성 차분: 인근한 두 계절 값의 차이를 산출

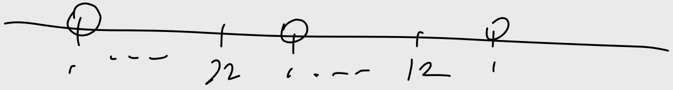
$$\Delta_s Z_t = Z_t - Z_{t-s} = (1 - B^s) Z_t$$

추세만 없어짐



계절성 ARIMA 모형

계절성 ARIMA모형의 유도 예



- 주기 s=12를 갖는 추세없는 월별 시계열 고려
- 매년 1월 데이터들만 볼 때 MA(1)모형을 따른다 하자.
  즉, $Z_t = (1 - \Theta B^{12})\alpha_t$
  여기서 α_t 들은 오차항으로 $\alpha_t, \alpha_{t-12}, \alpha_{t-24}$ 등은 서로 상관관계가 없다.
  매년 2월 데이터들도 MA(1)을 따른다 하면 위와 동일한 모형이 된다.
  그러나 인근 월의 오차항간에는 상관관계가 있으므로 새로운 모형이 필요하다.
- 오차항 α_t이 다시 MA(1)모형을 따른다 하자. '즉, $\alpha_t = (1 - \theta B)a_t$ (여기서 a_t는 백색잡음)
  위의 두 MA(1)모형을 결합하면 다음과 같다.

$$Z_t = (1 - \theta B)(1 - \Theta B^{12})a_t$$

- 이 모형은 비계절성 MA(1)과 계절성 MA(1) 모형이 결합된 형태로 계절성 $ARMA\ (0,1) \times (0,1)_{12}$ 또는 계절성 $\underbrace{ARIMA\ (0,0,1)}_{\text{비계절성}} \times \underbrace{(0,0,1)_{12}}_{\text{계절성}}$ 라 함.

계절성 ARIMA 모형

(예) 계절성 $ARIMA\ (0,1,1)\times(0,1,1)_{12}$

비계절성 1차 차분 시계열이 MA(1)을 따르며 주기 12의 계절성 1차 차분 시계열이 MA(1)을 따름

$W_t = (1-B)(1-B^{12})Z_t$: 비계절성 1차 차분 및 계절성 1차 차분 시계열

$$W_t = \boxed{(1-\theta B)}(1-\Theta B^{12})a_t$$

$$(1-B)(1-B^{12})Z_t = (1-\theta B)(1-\Theta B^{12})a_t$$

$$(1-B-B^{12}+B^{13})Z_t = (1-\theta B-\Theta B^{12}+\theta\Theta B^{13})a_t$$

$$Z_t = Z_{t-1}+Z_{t-12}-Z_{t-13}+a_t-\theta a_{t-1}-\Theta a_{t-12}+\theta\Theta a_{t-13}$$

- 일반적 계절성 ARIMA 모형
  - 표기: 계절성 $ARIMA\ (p,d,q)\times(P,D,Q)_s$
  - 비계절성 부분 d차 차분시 $\boxed{ARMA\ (p,q)}$
  - 주기 s의 계절성 부분 D차 차분시 $ARMA\ (P,Q)$
  - $W_t = (1-B)^d(1-B^s)^D Z_t$ 라 할 때

$$\phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)a_t$$

계절성 ARIMA 모형

모형의 식별 및 추정

(단계 1) 시계열도를 그려보고 추세 및 계절성 존재여부를 판단
(단계 2) 아래사항을 고려하여 적절히 차분 실시

- 추세는 없고 계절성이 있는 경우: 해당 주기에 대한 계절성 차분
- 추세가 있고 뚜렷한 계절성이 없는 경우: 선형추세가 있는 경우 1차 차분, 곡선형태의 추세가 있는 경우 차분 전에 함수 변환 시도
- 추세와 계절성이 있는 경우: 우선 계절성 차분을 실시하고 추세를 다시 검토; 추세가 여전이 남아있는 경우 1차 차분 추가 실시

(단계 3) 차분 시계열에 대한 ACF 와 PACF를 바탕으로 p, q,P,Q를 결정

- 비계절성 계수인 p, q는 ARMA모형의 경우와 동일한 요령으로 결정
- 계절성 계수인 P, Q는 주기의 배수에서 나타나는 ACF 와 PACF의 패턴을 보고 결정

*여러가지 대안 모형 세우고 잔차검정 통해 적절한 모형 찾아내자

(단계 4) 모형 파라미터 추정
(단계 5) 잔차 검정 실시

계절성 ARIMA 모형

ACF 산출 예

- 계절성 $ARIMA\ (0,1,1)\times(0,1,1)_{12}$
- 차분시계열 산출 $W_t = (1-B)(1-B^{12})Z_t$
- 모형: $W_t = (1-\theta B)(1-\Theta B^{12})a_t = a_t - \theta a_{t-1} - \Theta\, a_{t-12} + \theta\Theta\, a_{t-13}$
- W_t 분산

  $\gamma(0) = Var[W_t] = (1+\theta^2+\Theta^2+\theta^2\Theta^2)\sigma_a^2 = (1+\theta^2)(1+\Theta^2)\,\sigma_a^2$
- 시차 k의 자기공분산

  $$\gamma(k) = E[W_t W_{t-k}] = E[(a_t - \theta a_{t-1} - \Theta\, a_{t-12} + \theta\Theta\, a_{t-13})(a_{t-k} - \theta a_{t-k-1} - \Theta\, a_{t-k-12} + \theta\Theta\, a_{t-k-13})]$$
- ACF

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{-\theta(1+\Theta^2)\sigma_a^2}{(1+\theta^2)(1+\Theta^2)\,\sigma_a^2} = \frac{-\theta}{1+\theta^2}$$

$$\rho(11) = \frac{\gamma(11)}{\gamma(0)} = \frac{\theta\Theta\sigma_a^2}{(1+\theta^2)(1+\Theta^2)\,\sigma_a^2} = \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)} = \rho(13)$$

$$\rho(12) = \frac{\gamma(12)}{\gamma(0)} = \frac{-\Theta(1+\theta^2)\sigma_a^2}{(1+\theta^2)(1+\Theta^2)\,\sigma_a^2} = \frac{-\Theta}{1+\Theta^2}$$

$1 - \theta B a_t - \Theta B^{12} a_t + \theta\Theta B^{13} a_t.$

$\gamma(1) = E[(a_t - \theta a_{t-1}\ \Theta a_{t-12} + \theta\Theta a_{t-13})(a_{t-1} - \theta a_{t-2} - \Theta a_{t-13} + \theta\Theta a_{t-14})]$

$= -\theta E[a_{t-1}a_{t-1}] - \theta\Theta^2 E[a_{t-13}a_{t-13}]$

$= \sigma_a^2 \times (-\theta)(1+\Theta^2)$

* MA → ACF cut off.

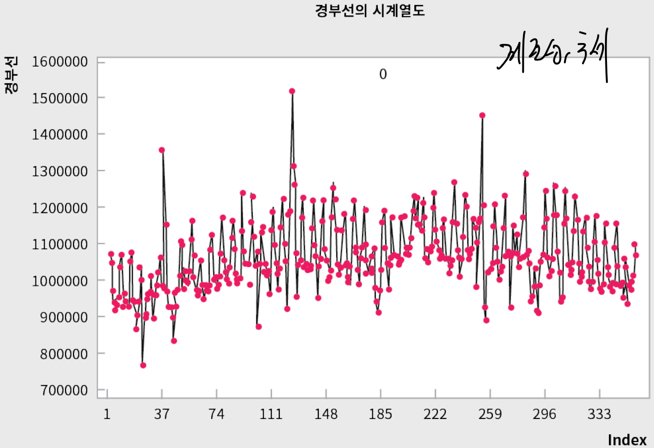
11, 12, 13 값 가짐

모형의 특성은 모형 식별에 이용

계절성 ARIMA 모형

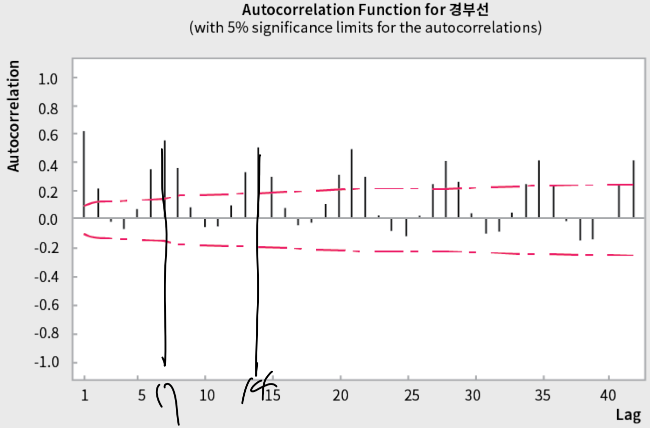
(예 - 일별 경부선 차량운행수) 2016년 1월 1일부터 2016년 12월 31일 1년간 일별 경부선 1종 차량운행수를 나타내고 있다

단계 1



주기 7의 계절성, 약간의 추세

ACF



ACF가 주기 7 부근에서 나타나며 서서히 감수

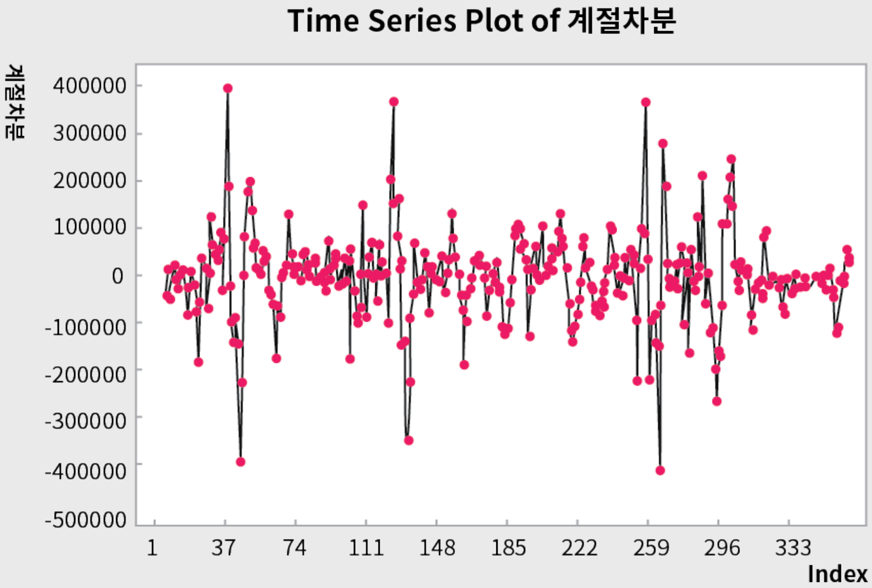
#1

계절성 ARIMA 모형

(예 계속 – 일별 경부선 차량운행수)

추세 없어지고 정상적 => 추가 차분 X

단계 2



계절 (주기 7) 차분 시계열

차분 시계열의 경우 추세와 계절성이 보이지 않으므로 정상적으로 간주됨

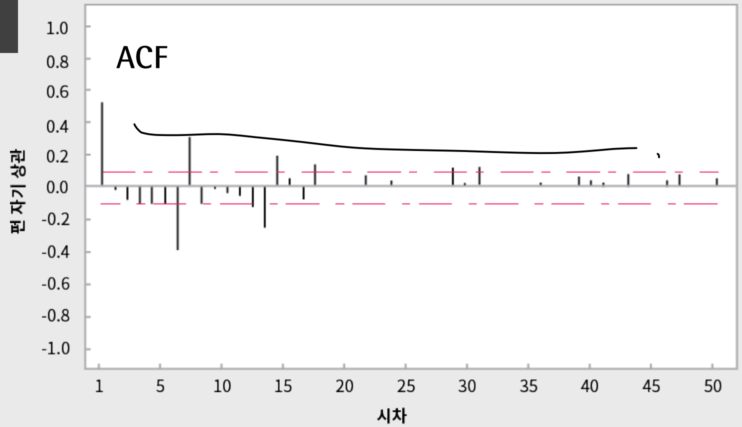
#1

계절성 ARIMA 모형

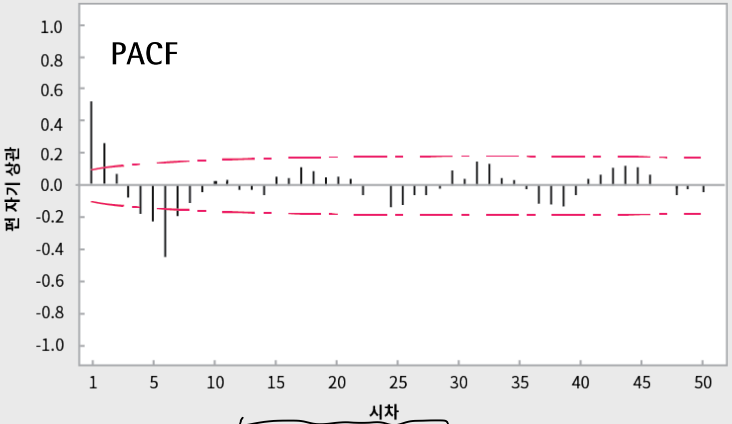
(예 계속 - 일별 경부선 차량운행수)

단계 2

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



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



#1

ACF를 볼 때 비계절성 부분은 지수적으로 감소하는 AR(1)형태

PACF를 볼 때 주기 7의 배수에서 지수적으로 감소하므로 계절성 부분은 MA(1)형태

따라서 계절성 $ARIMA\ (1,0,0)\times(0,1,1)_7$

계절성 ARIMA 모형

비계절성 계절성

(예 계속 - 일별 경부선 차량운행수) 모형: $(1-\phi B)(1-B^{7})Z_t = (1-\Theta B^{7})a_t$

단계 4

	계수추정치	표준오차	T값	p값
AR1 ϕ	0.6052	0.0428	14.15	0.000
SMA12 Θ	0.8721	0.0264	32.99	0.000

단계 5 (잔차검정)

시차	12	24	36	48
카이제곱통계량	16.6	36.4	49.9	59.9
p-값	0.084	0.028	0.039	0.081

#1

5%에서 유의하지 않다

=> 잔차는 백색잡음

단위근 검정

formal 하게

- 단위근 검정 (unit root test)는 통계적 검정을 통하여 시계열의 정상성 여부를 판정
  - 대표적인 단위근 검정은 ADF (augmented Dickey-Fuller) 검정
  - Dickey and Fuller (1979)가 AR(1)모형에 대해 제안
  - ADF 검정은 Said and Dickey (1984)가 ARMA모형으로 확장한 것
  - 모든 정상적 시계열은 고차원의 AR모형으로 근사될 수 있다고 가정
  - AR(p)모형: $\phi_p(B)Z_t = a_t$

  ⇒ (단위근 포함) $\phi_p(B) = (1-B)\varphi_{p-1}(B)$

  ⇒ $(1-B)(1-\varphi_1 B - \cdots - \varphi_{p-1}B^{p-1})Z_t = a_t$

  ⇒ $(1-B)Z_t = (1-B)(\varphi_1 B - \cdots - \varphi_{p-1}B^{p-1})Z_t + a_t$

  ⇒ $Z_t - Z_{t-1} = \sum_{j=1}^{p-1} \varphi_j (Z_{t-j} - Z_{t-j-1}) + a_t$

단위근 검정

ADF 검정 (형태 1)

- 다음 모형을 고려

$$Z_t = \phi Z_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta Z_{t-j} + a_t \quad (\Delta Z_{t-j} = Z_{t-j} - Z_{t-j-1})$$

- 가설

$$\boxed{H_0: \phi = 1}$$

- 검정통계량

$$\mathrm{T} = \frac{\hat{\phi} - 1}{se(\hat{\phi})}$$

완전한 T분포 아님

  - 위의 통계량 분포는 브라운운동과 관련된 복잡한 형태이나 누적확률분포표가 만들어져 있음

- 판정
  - 가설이 기각되면 단위근이 없다고 할 수 있으므로 시계열이 정상적으로 간주
  - 가설이 채택되면 단위근이 있으므로 차분을 취한 시계열을 추후 분석에 활용

단위근 검정

ADF 검정 (형태 2)

$$Z_t = \phi Z_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta Z_{t-j} + a_t$$ 양변에 $-Z_{t-1}$

$$\Rightarrow \Delta Z_t = (\phi - 1) Z_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta Z_{t-j} + a_t$$

$$\Rightarrow \Delta Z_t = \phi^* Z_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta Z_{t-j} + a_t \quad \underline{(\phi^* = \phi - 1)}$$

- 가설

$$H_0: \phi^* = 0$$

- 검정통계량

$$\mathrm{T} = \frac{\widehat{\phi^*}}{se(\widehat{\phi^*})}$$

단위근 검정 예

(예 – 금값) 그림은 1978 ~ 2017년 평균 금값 (미화)을 나타낸다. 단위근 검정을 실시하면 아래와 같다.

Null Hypothesis: US_DOLLAR has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=9)

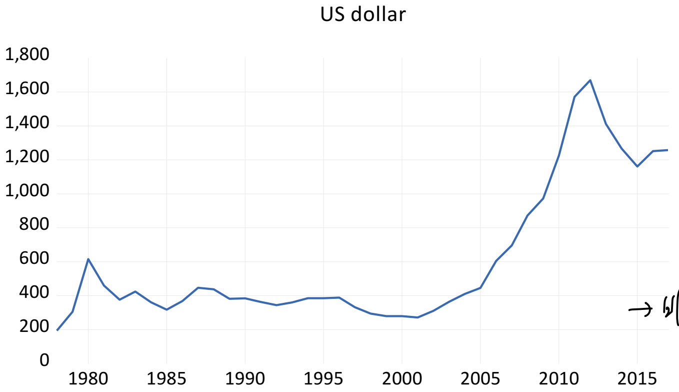
ADF

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.942313	0.7635
Test critical values:	1% level	-3.615588	
	5% level	-2.941145	
	10% level	-2.609066	

*MacKinnon (1996) one-sided p-values.

가설 기각 X

unit root가 있다.



→ 비정상적

#1

단위근 검정 예

(예 계속) ADF 추정식은 아래와 같다.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(US_DOLLAR)
Method: Least Squares
Date: 07/24/19 Time: 15:19
Sample (adjusted): 1980 2017
Included observations: 38 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z_{t-1} US_DOLLAR(-1)	-0.043415	0.046073	-0.942313	0.3525
ΔZ_{t-1} D(US_DOLLAR(-1))	0.414490	0.157314	2.634789	0.0125
C	39.54695	32.36196	1.222019	0.2299

R-squared	0.167961	Mean dependent var	25.06500
Adjusted R-squared	0.120416	S.D. dependent var	117.7672
S.E. of regression	110.4493	Akaike info criterion	12.32265
Sum squared resid	426967.0	Schwarz criterion	12.45193
Log likelihood	-231.1303	Hannan-Quinn criter.	12.36865
F-statistic	3.532673	Durbin-Watson stat	1.813049
Prob(F-statistic)	0.040042		

ADF 추정식:

$$\Delta US_t = 39.55 - 0.0434 US_{t-1} + 0.4145 \Delta US_{t-1}$$

$\phi^* = 0$

#1

단위근 검정 예

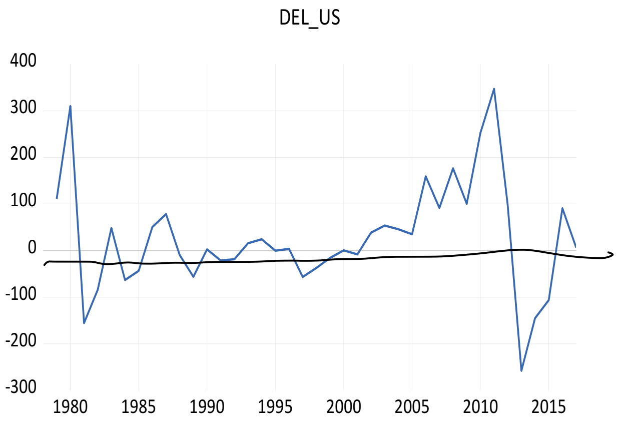
(예 계속) 금값의 1차 차분 시계열에 대한 단위근 검정

Null Hypothesis: D(US_DOLLAR) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.049396	0.0032
Test critical values:	1% level	-3.615588	
	5% level	-2.941145	
	10% level	-2.609066	

*MacKinnon (1996) one-sided p-values.

unit root X



#1