시계열 분석 기법과 응용

Week 4. 비정상적 시계열

ARIMA , 이를 정상적 시계열로 변환하는 방법,
계절성을 동반하는 시계열의 처리
* 추세 or 계절성이 포함된 비정상적 시계열은 적절한 차분을 통해 정상적 시계열로 바꿀 수 있음

4-1. ARIMA 모형

전치혁 교수
(포항공과대학교 산업경영공학과)

비정상적 시계열

시계열의 비정상성 (nonstationarity)

- 시계열에 추세 또는 계절성이 포함되는 경우 정상성을 만족하지 못한다.

- 비정상성 판단 방법
  - ① 시계열의 시간에 대한 그래프를 보고 시각적으로 판단
  - ② ACF가 시차에 대하여 매우 서서히 감소하는 패턴
  - ③ 단위근 검정 (통계적 검정)

    └→ 가설 검정, 단위근 有 = 비정상적 시계열

    * H_0: 단위근 존재 vs H_a: $\sim H_0$

- 비정상성 대응 방안 → 정상적 시계열로 변환!
  - 차분 (differencing)을 통해 정상적 시계열로 변환
  - 함수변환을 통하여 분산 안정화 ex) 원시계열에 분산이 증가하는 패턴이 있을 때 log 변환 → 분산 안정화
  - 분해법으로 추세 및 계절성 제거 이후 정상적 시계열로 간주

ARIMA 모형

차분 (differencing)

- 1차 차분: 인근한 두 값의 차이를 산출

  └ ① 1차차분

  결과: 정상적인 시계열 → 끝

  비정상, 추세有 → 2차차분 ..

  후방 연산자

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

- 2차 차분 → ΔZ_t에 대해 다시 차분

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

- d차 차분 $(d = 1,2, \dots)$

  *현실적: 2차차분에서 대부분 해결됨

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

비정상적 시계열의 모형화

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

AR, MA, 그리고 누적 시계열이 관련있다

- 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

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

이를 식별하면 원시계열 파악 가능

이게 정상적일 때, 원시계열

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

= 차수 d누적시계열

Z_t ~ ARIMA(p,d,q)

W_t ~ ARMA(p,q)

ARIMA 모형

p d q

(예) ARIMA(1,1,1) 모형

1차차분 (d=1)

- 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: $Z_t \sim ARIMA(1,1,1) \Rightarrow (1-\phi_1 B)(1-B)Z_t = (1-\theta_1 B)a_t \longrightarrow$

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

p d q

(예) ARIMA(0,1,1) 또는 IMA(1,1)

=

AR 부분없음

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

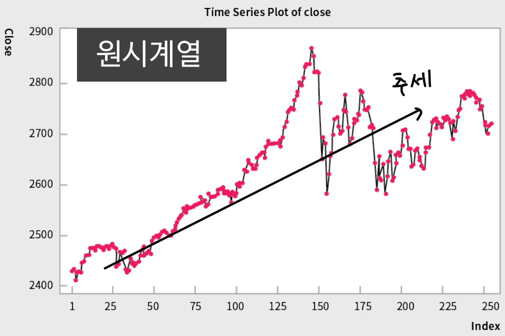
= IMA(1,1)

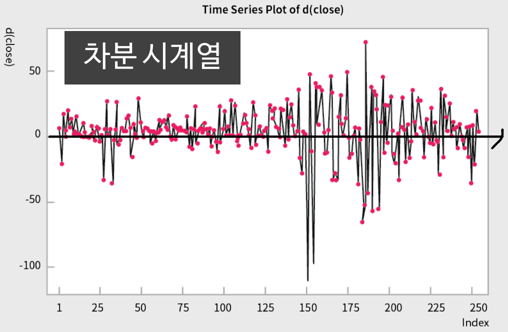
ARIMA 모형

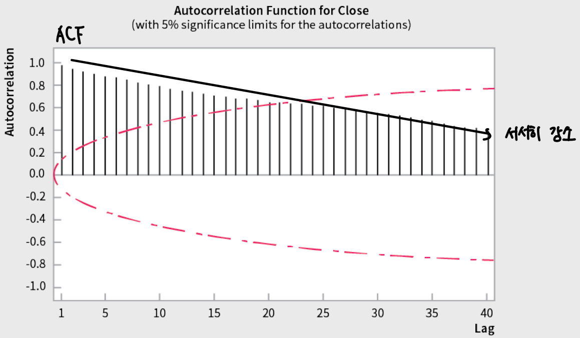
예 (S&P 200 지수) 아래 그림은 2017년 7월 3일 부터 2018년 6월 29일 까지 미국 S&P 200 지수의 종가를 나타낸다.

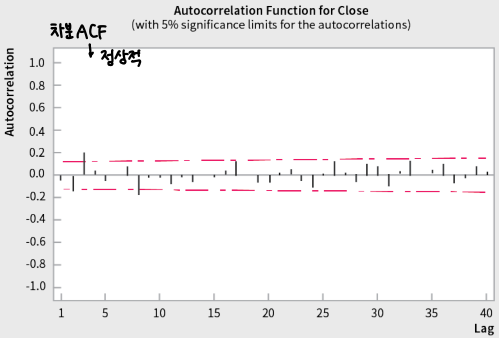
매일

원 시계열과 1차 차분 시계열에 대한 ACF를 살펴보자









#1

Reference

#1. Yahoo Finance https://finance.yahoo.com/quote/%5EGSPC/history/ 2019.12

시계열 분석 기법과 응용

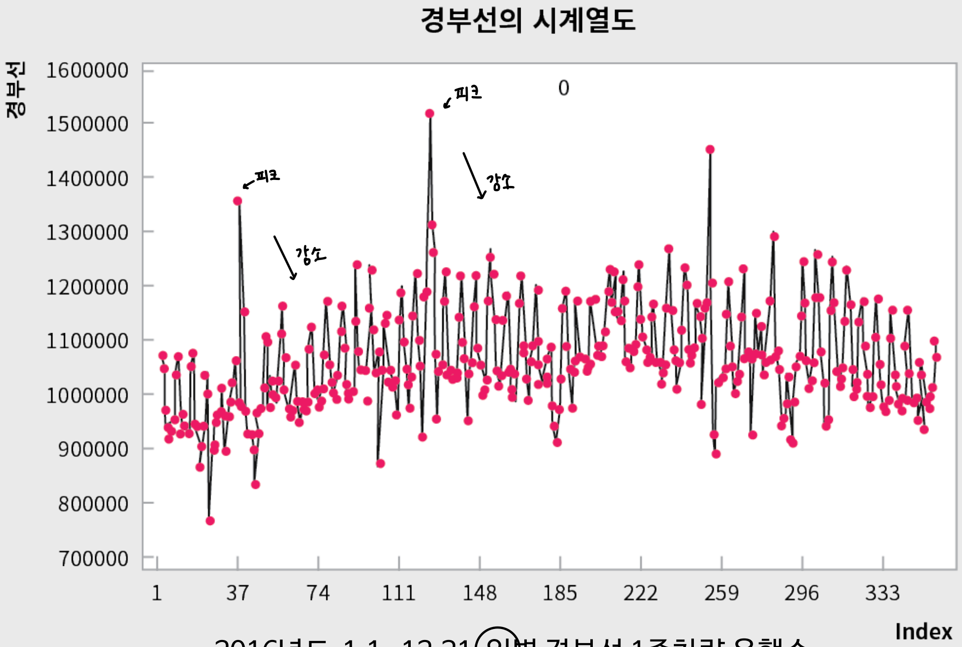
Week 4. 비정상적 시계열
4-2. 계절성 ARIMA모형

전치혁 교수
(포항공과대학교 산업경영공학과)

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.

계절성이 보임



2016년도 1.1~12.31 일별 경부선 1종차량 운행수

계절성 ARIMA 모형

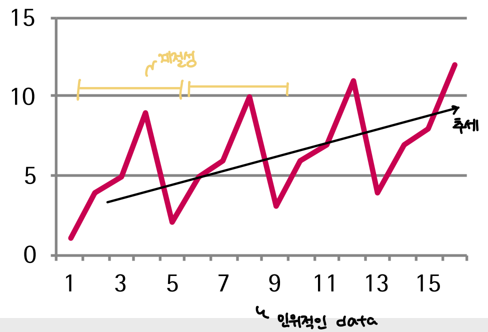
계절성 제거를 위함

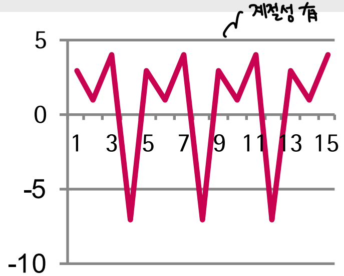
계절성 차분 (seasonal differencing)

- 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

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

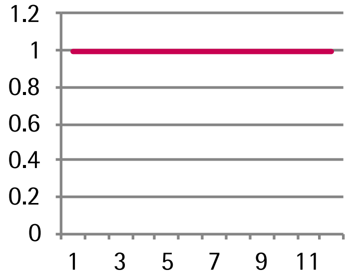
ex) s=12
이전 해의 월 (t-12)를 빼는것





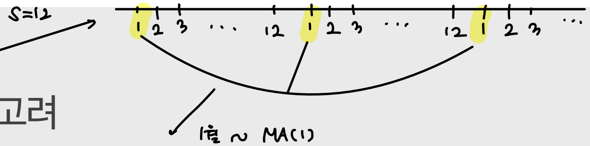
차분

추세제거



계절성 차분

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.

  계절성모형

  즉, $Z_t = (1 - \Theta \underbrace{B^{12}}_{\text{1월만 보므로 '12'간격!}})\alpha_t$ 오차항

  여기서 α_t 들은 오차항으로 $\alpha_t, \alpha_{t-12}, \alpha_{t-24}$ 등은 서로 상관관계가 없다. (α는 독립이므로)

  매년 2월 데이터들도 MA(1)을 따른다 하면 위와 동일한 모형이 된다.

  그러나 인근 월의 오차항간에는 상관관계가 있으므로 새로운 모형이 필요하다. ∴ α에 대한 새로운 모형이 필요함

- 비계절성 모형 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. 비계절성 MA 계절성 MA

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

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

  차분없으므로 d=0

계절성 ARIMA 모형

0이므로 AR없음 비계절성 0이므로 AR없음 계절성

(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

비계절성, 계절성 차분 고려한 것

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

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

소프트웨어 활용

(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 ① 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 아 여러 모형을 세워 추정 → 잔차검정 → 적절모형 선택

(단계 4)모형 파라미터 추정

(단계 5)잔차 검정 실시

계절성 ARIMA 모형

차분한 시계열에 대해 ACF를 구하면 됨

ACF 산출 예

- 계절성 $ARIMA\ (0,1,1)\times(0,1,1)_{12}$
- ① 차분시계열 산출 $W_t = (1-B)(1-B^{12})Z_t$
- 모형: $W_t = (1-\theta B)(1-\Theta B^{12})a_t = a_t - \theta a_{t-1} - \Theta\, a_{t-12} + \theta\Theta\, a_{t-13}$
- ② W_t 분산

  $$\gamma(0) = Var[W_t] = (1+\theta^2+\Theta^2+\theta^2\Theta^2)\sigma_a^2 = (1+\theta^2)(1+\Theta^2)\,\sigma_a^2$$
- ③ 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

자기상관계수

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

$\gamma(0)$ 분산

원래 MA는 시차 1일 때만 있음 (MA(1)의 ACF)
비계절성 x 계절성 복합 → 시차 11, 12, 13에도 값을 가짐

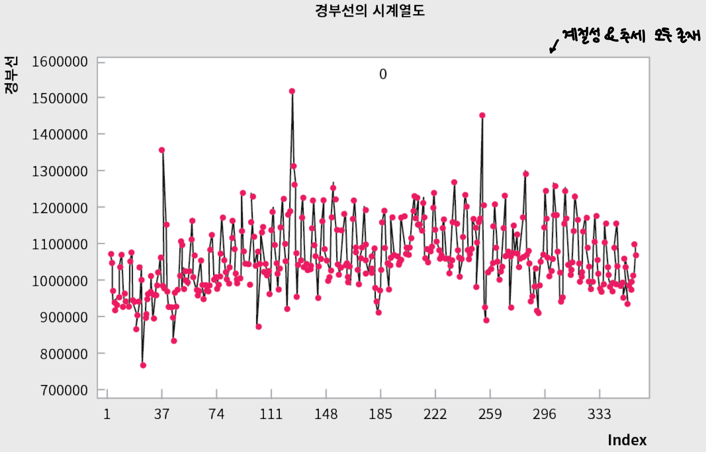
$$\rho(11) = \frac{\gamma(11)}{\gamma(0)} = \frac{\theta\Theta\sigma_a^2}{(1+\theta^2)(1+\Theta^2)\,\sigma_a^2} = \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)} = \rho(13)$$

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

계절성 ARIMA 모형

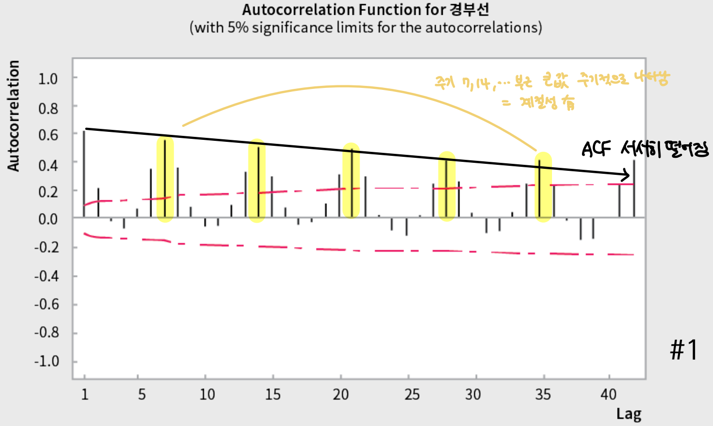
(예 – 일별 경부선 차량운행수) 2016년 1월 1일부터 2016년 12월 31일 1년간 일별 경부선 1종 차량운행수를 나타내고 있다

단계 1



주기 7의 계절성, 약간의 추세

ACF



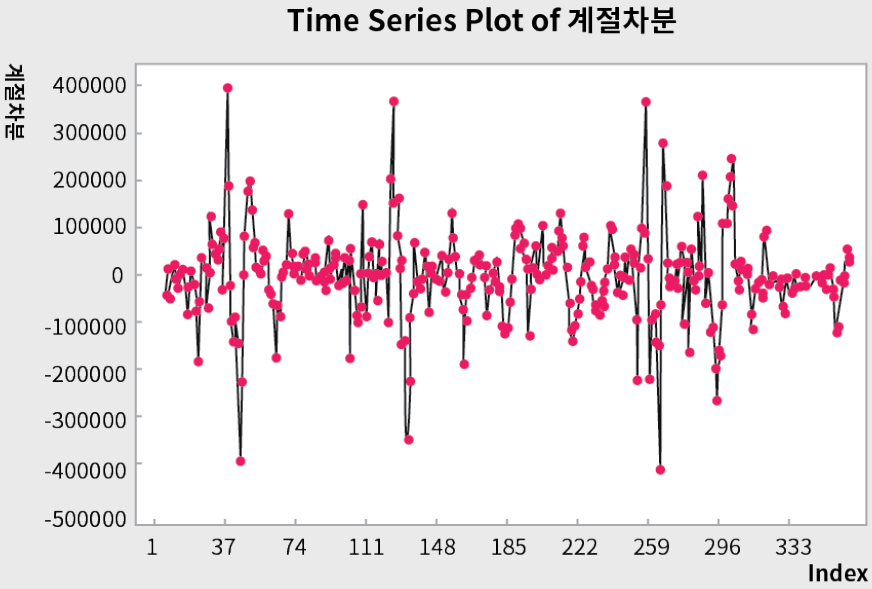
ACF가 주기 7 부근에서 나타나며 서서히 감수

#1

계절성 ARIMA 모형

(예 계속 - 일별 경부선 차량운행수)

단계 2



계절 (주기 7) 차분 시계열

차분 시계열의 경우 추세와 계절성이 보이지 않으므로 정상적으로 간주됨

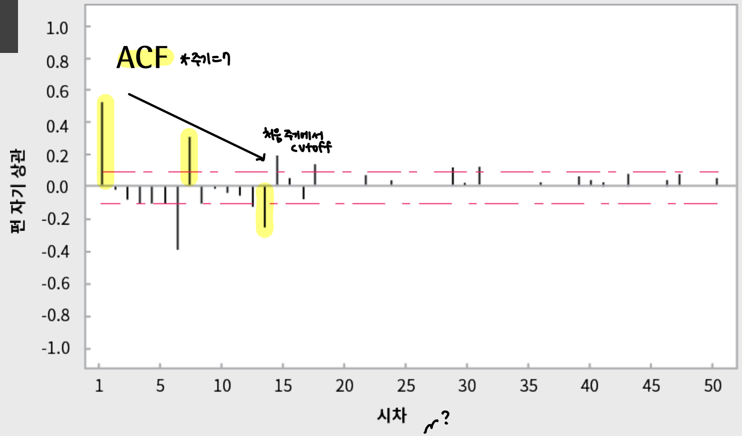
#1

계절성 ARIMA 모형

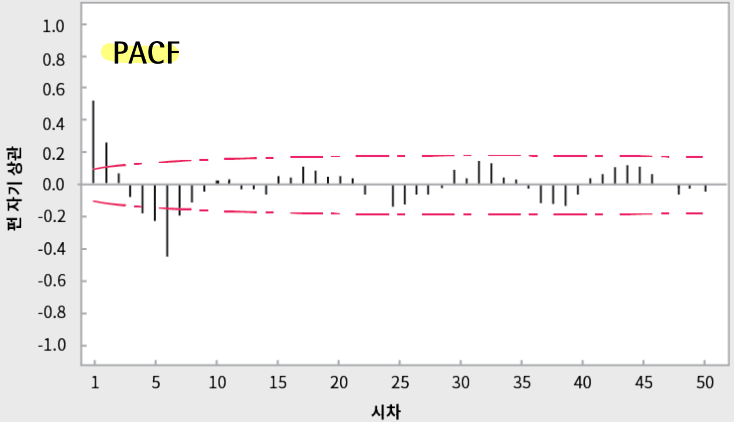
(예 계속 - 일별 경부선 차량운행수)

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

*비계절성 → 차분X : AR(1)
계절성 → 차분o : MA(1)

계절성 ARIMA 모형

계절성 부분안 차분항

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

비계절성: AR(1) ϕ 추정해야함

MA의 parameter

* EVIEWS or MINITAB 사용 추정 결과

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

유의

잔차가 백색잡음인가?

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

유의 → 의심할 필요

큰 차수에서 유의 X

∴ 식별 제대로 되었다고 볼 수 있음

Reference

#1. KOSIS 국가통계포털 http://kosis.kr 2019.12

시계열 분석 기법과 응용

Week 4. 비정상적 시계열
4-3. 단위근 검정

전치혁 교수
(포항공과대학교 산업경영공학과)

단위근 검정

- 단위근 검정 (unit root test)는 통계적 검정을 통하여 시계열의 정상성 여부를 판정
  - 대표적인 단위근 검정은 $\boxed{\text{ADF (augmented Dickey-Fuller) 검정}}$
  - 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 ∴ AR모형만 고려
  - AR(p)모형: $\phi_p(B)Z_t = a_t$ ← white noise

다항식

$\Rightarrow$ (단위근 포함) $\phi_p(B) = (1-B)\varphi_{p-1}(B)$ ~ 즉, AR에 단위근 포함

→ (1-B)의 인수 분해 가능 ⇒ $(1-B)\varphi_{p-1}(B)=0$ 즉, B=1이 됨!

∴ 단위근이라 부름

$\Rightarrow (1-B)(1-\varphi_1 B - \cdots - \varphi_{p-1}B^{p-1})Z_t = a_t$ ← (1-B)를 포함하는 AR 다항식 모형

이항

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

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

일반화: φ 2더블

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

$$H_0: \phi = 1$$ → 단위근 존재

- 검정통계량

$$\mathrm{T} = \frac{\hat{\phi} - 1}{se(\hat{\phi})}$$ → t분포는 아니지만 t값과 유사하게 유도함

- 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)

EViews나 다른 소프트웨어에서 사용하는 형태

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

일반화 : φ 지워볼

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

양변에 $-Z_{t-1}$

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

차분한 형태

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

- 가설

H_0: $\phi^* = 0$

- 검정통계량

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

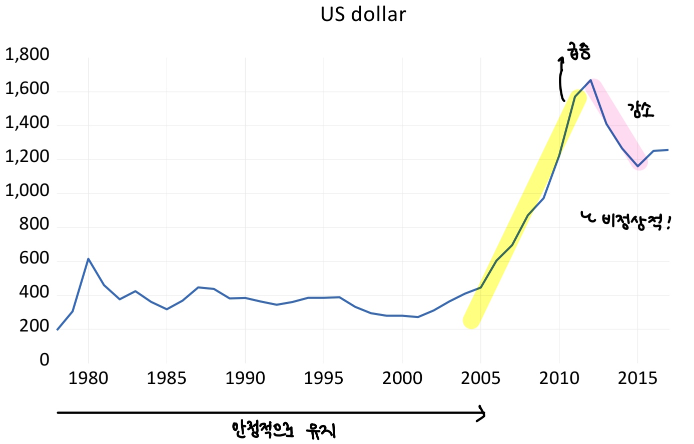
단위근 검정 예

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

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

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic ADF검정		-0.942313	0.7635 > 0.05 ∴ don't reject H_0
Test critical values:	1% level	-3.615588	
	5% level	-2.941145	
	10% level	-2.609066	

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



#1

단위근 검정 예

(예 계속) ADF 추정식은 아래와 같다.

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

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

유의 X : 기각 X
∴ $\phi = 0$
⇓
unit root 존재

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:

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

단위근 검정 예

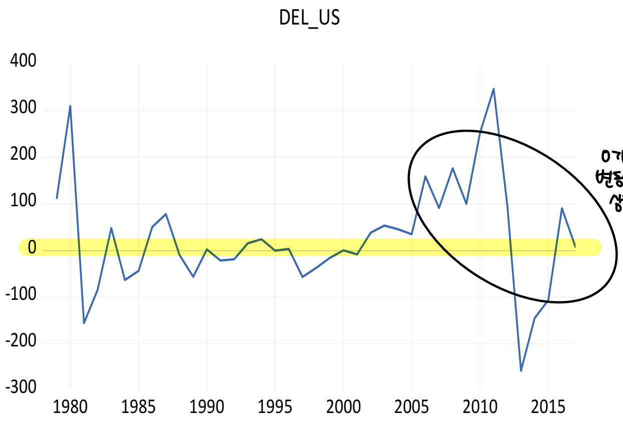
(예 계속) 금값의 1차 차분 시계열에 대한 단위근 검정

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.

↳ reject H_0
∴ unit root 없음
⇒ 1차 차분 시계열
: 정상적이다!



0기준
변동이 특히
상향

#1

Reference

#1. Reuters Datastream, LBMA, World Gold Council