WEEK4

1. **ARIMA 모형**
   - 비정상적 시계열: 시계열이 추세나 계절성을 포함하면 정상성을 만족하지 못함
   1) 시간에 따른 시계열 데이터의 그래프를 통해 시각적으로 판단
   2) ACF가 매우 천천히 감소하는 패턴
   3) 단위근 검정

   - 해결법
   1) 차분을 통해 변환
   2) 비선형이거나 분산을 안정화할 필요가 있을 때 → 함수변환(로그 변환)
   3) 분해법으로 추세 및 계절성 제거

- 차분
  - 1차 차분: 인접한 두 값의 차이를 산출

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

B는 후향 연산자

- d차 차분:

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

- 차수 d 누적 시계열: d차 차분 후, 시계열이 처음으로 정상적일 때, 원 시계열을 일컫는 표현-I(d)

- ARIMA 모형의 표현 → d차 차분을 한 시계열이 정상적 ARMA(p,q)를 따를 때, 원 시계열은 ARIMA(p,d,q)를 따르게 됨

$$B: \text{후향연산자(Backshift operator) i.e. } B^b y_t = y_{t-b}$$

$$\delta_p = (1 - \phi_1 - \phi_2 - \cdots - \phi_p)\mu$$

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$$

$$\theta_q(B) = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)$$

$$\Phi_p(B^s) = (1 - \Phi_1 B^s, -\Phi_2 B^{2s} - \cdots - \Phi_P B^{PS})$$

$$\Theta_Q(B^s) = (1 + \Theta_1 B^s + \Theta_B^{2s} + \cdots + \Theta_Q B^{Qs})$$

후향 연산자 notation

- 따라서, 다음이 성립함

- 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

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

(ex)

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

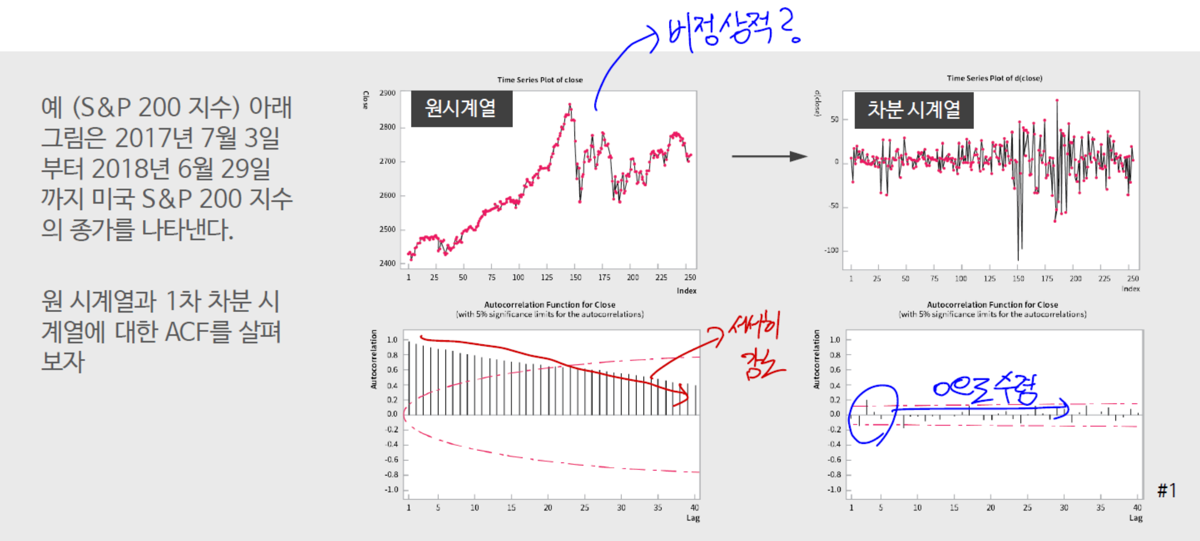
$$\Rightarrow Z_t = (1+\phi_1)Z_{t-1} - \phi_1 Z_{t-2} + a_t - \theta_1 a_{t-1}$$

p=0, d=1, q=1

$$\theta_1 B) a_t$$

$$(1 - B - \phi_1 B + \phi_1 B^2) z_t$$
$$= a_t - \theta_1 B \cdot a_t$$
$$\Leftrightarrow z_t = z_{t-1} + \phi_1 z_{t-1}$$
$$- \phi_1 z_{t-2} + a_t - \theta_1 \cdot a_{t-1}$$
(using 후향연산자)

- 원 시계열 vs 차분 시계열의 ACF

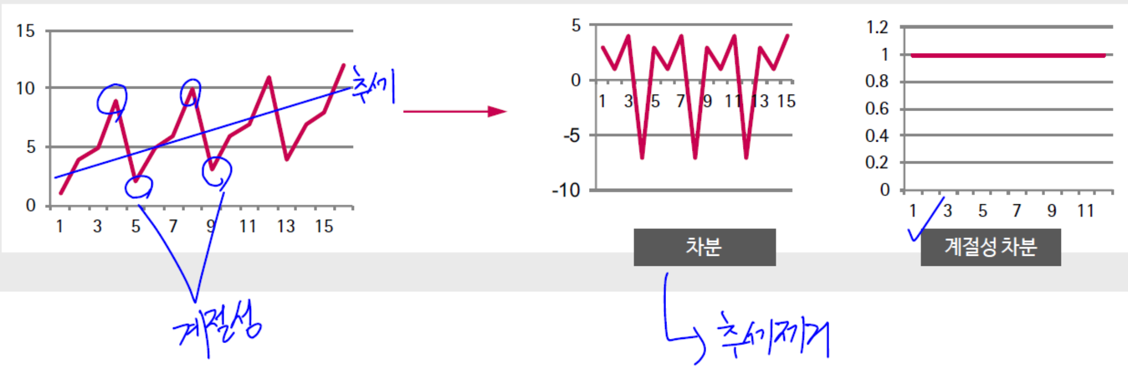


2. **계절성 ARIMA 모형(SARIMA)**

- 비계절성 ARIMA 모형에 계절성 항을 추가하여 표현함
- 일반적인 시계열에는 추세와 계절성이 모두 포함되기 때문에 둘을 모두 처리해야함

- 계절성 차분(주기=s)
- 1차 계절성 차분: 인근한 두 계절 값의 차이를 산출

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



예1) SARIMA 모형 유도
- 매년 1월 데이터만 추출하였을 때, MA(1)을 따름(주기 = 12)
- 오차항 a_t 역시 MA(1)을 따름
- 위의 두 MA(1) 모형을 결합한 것이 비계절성 MA(1)과 계절성 MA(1)을 결합한 것임

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

(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

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

차분

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

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

예2) 분기별 시계열 모형
- 상수가 없는 ARIMA$(1,1,1)(1,1,1)_4$ 모형 생각(y_t는 최초 시계열 데이터의 target)

$$(1-\phi_1 B)\ (1-\Phi_1 B^4)(1-B)(1-B^4)y_t = (1+\theta_1 B)\ (1+\Theta_1 B^4)\varepsilon_t$$

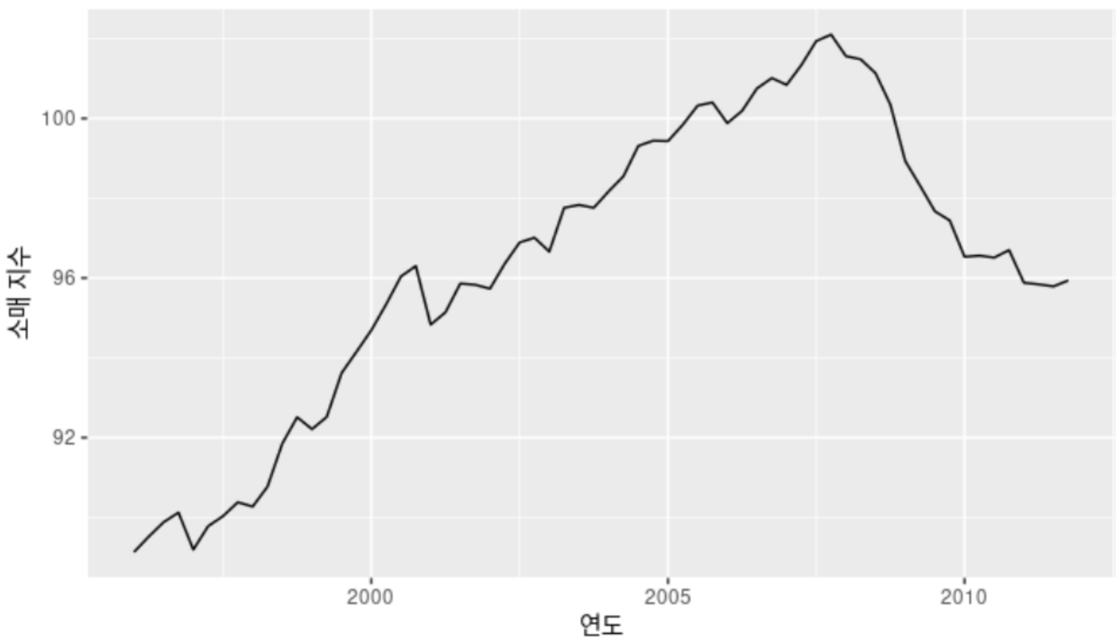
- 모형의 식별 및 추정

(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

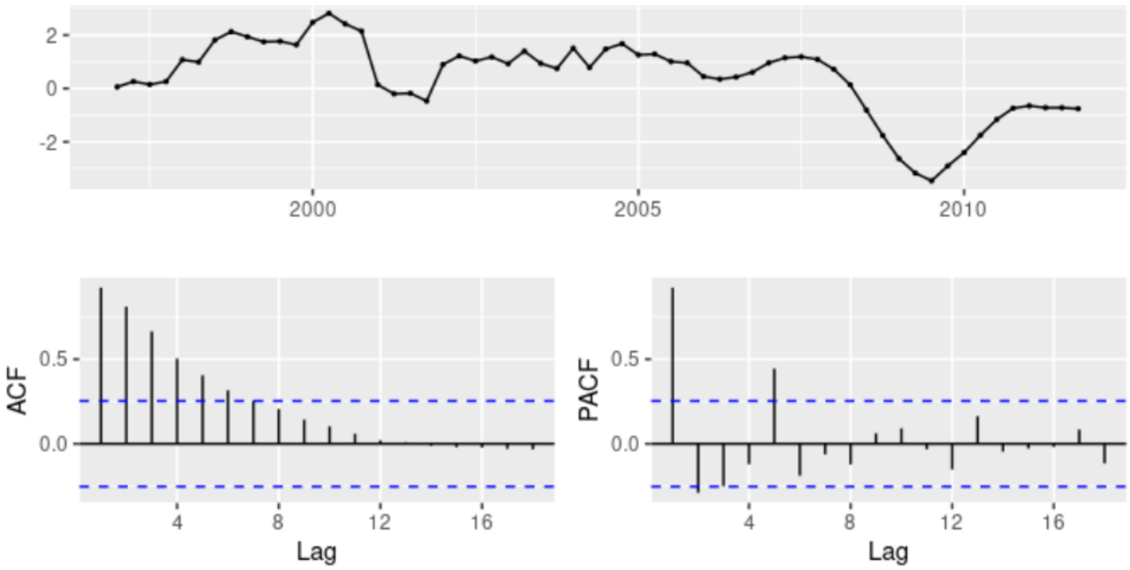
(단계 3) 차분 시계열에 대한 ACF 와 PACF를 바탕으로 p, q,P,Q를 결정
- 비계절성 계수인 p, q는 ARMA모형의 경우와 동일한 요령으로 결정
- 계절성 계수인 P, Q는 주기의 배수에서 나타나는 ACF 와 PACF의 패턴을 보고 결정

(단계 4) 모형 파라미터 추정
(단계 5) 잔차 검정 실시

(예시)

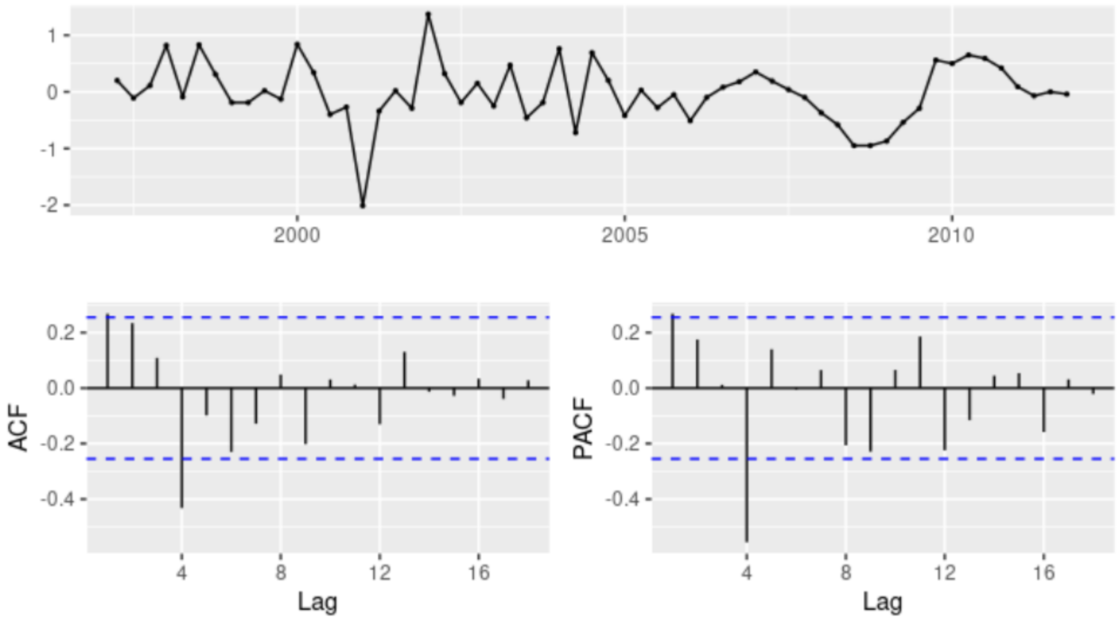


→ 데이터가 정상성을 나타내지 않음을 확인할 수 있고, 약간의 계절성도 포함되어 있어보임
→ 계절성 차분을 구함



→ ACF를 보면 서서히 감소하는 형태를 띠어, 이 역시 정상성을 띈다고 보기에는 어려움

→ 1차 차분을 한 번 더 구해봄



→ 이를 활용해 적절한 ARIMA 예측 모형을 찾기(parameter 찾기)

→ ACF에서 시차 1의 막대가 비-계절성 MA(1)을 암시함

→ ACF에서 시차 4의 유의미하게 뾰족한 막대는 계절성 MA(1) 성분을 암시함

→ 이를 통해, 1차 차분과 계절성 성분을 나타내는 ARIMA$(0,1,1)(0,1,1)_4$ 모델로 시작함

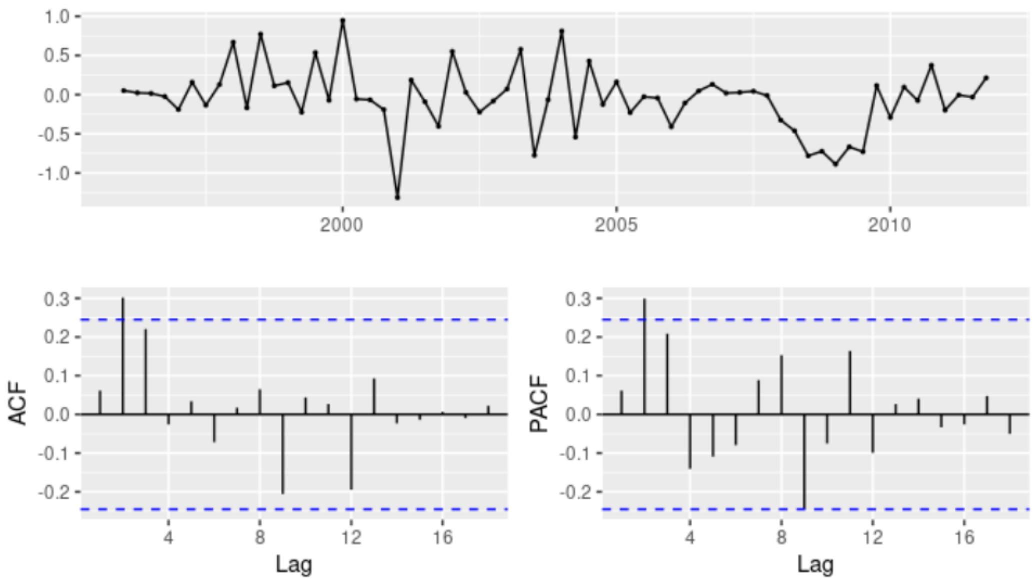


Figure 8.20: 유럽 소매 거래 지수 데이터에 대해 맞춘 ARIMA(0,1,1)(0,1,1)$_4$ 모델의 잔차.

→ ACF와 PACF 모두 시차 2에서 유의미한 막대가 나타나고, 시차 3에서 역시 유의하진 않지만 상대적으로 뾰족한 막대가 나타남을 확인할 수 있음
→ 몇몇의 추가적인 비-계절성 항을 모형에 추가하는 것이 합리적
→ ARIMA$(0,1,2)(0,1,1)_4$ 모형의 AIC 값과 ARIMA$(0,1,3)(0,1,1)_4$ 모형의 AIC 값을 비교
→ AIC 값이 더 작은 모형인 후자를 선택 → 해당 모형으로 예측 수행

- 추정

계절성 ARIMA 모형

추정

(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

백색잡음인지 확인

유의 X
∴ 잔차는 백색잡음임

3. **단위근 검정**

- 통계적 가설검정을 통해 시계열의 정상성 여부를 판단
- 대표적인 검정 방법: ADF
- 모든 정상적 시계열은 고차원의 AR 모형으로 근사될 수 있다고 가정

$$y_t = \alpha + \theta y_{t-1} + \varepsilon_t$$

theta = 1이면 비정상적 시계열

- 따라서, 정상성 판단을 위해, OLS 추정값을 활용하여 'H0: theta=1'를 검정하는 단위근 검정을 함
- 그런데, 위 식을 단순 OLS 추정하여 귀무가설을 검정하면 y가 비정상적 시계열인 경우 가성회귀 문제가 발생 → 식을 아래와 같이 변형하여 'H0: theta-1=0'을 검정

$$\Delta y_t = \alpha + (\theta - 1) y_{t-1} + \varepsilon_t$$

cf) 가성회귀: 실제로 두 변수 사이에 상관관계가 없다고 할지라도 그저 시간이 지남에 따라 증가하는 추세로 인하여, 검정통계량이 커져서 회귀 결과가 잘못 도출됨

- 그러나, 오차항에 자기상관이 존재하면 단위근 검정의 설명력을 낮춤
- 따라서, 종속변수 *Δyt의 과거값을 모형에 포함하여 모형을 추정하고 귀무가설을 검정함 = ADF 검정*

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

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

- 가설

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

- 검정통계량

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

→ 가설이 기각되면, 단위근이 없다고 할 수 있으므로 시계열이 정상적이라고 판단할 수 있음

(예시)

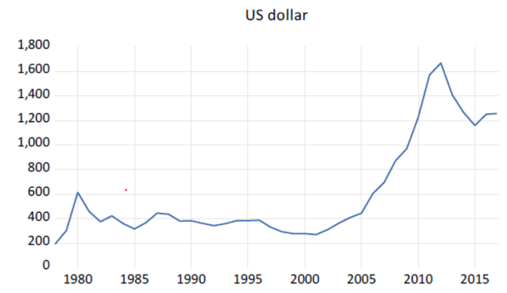
(예 - 금값) 그림은 1978 ~ 2017년 평균 금값 (미화)을 나타낸다. 단위근 검정을 실시하면 아래와 같다.

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

		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.

단위근 있음



→ 귀무가설을 채택하므로, 단위근이 있다고 보아야 함

(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.
US_DOLLAR(-1)	-0.043415	0.046073	-0.942313	0.3525
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:

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

→ 마찬가지로, ADF 추정식에서 US_{t-1}의 회귀계수에 대한 p값이 유의하지 않으므로 귀무가설 채택(≠0)

→ 차분을 해줘야 함

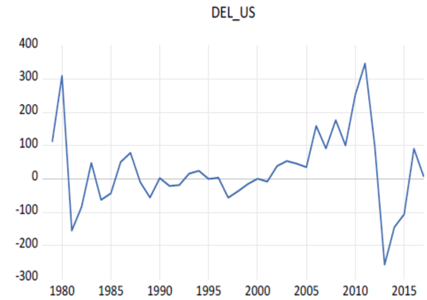
(예 계속) 금값의 1차 차분 시계열에 대한 단위근 검정

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	

reject H_0

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



→ ADF 검정 결과, 귀무가설을 기각함 ⇒ 시계열이 정상적으로 변환됨