시계열분석팀 3주차 교안 ⌛

[목차]

1. 2주차 복습

(1) 정상 시계열 모형(AR, MA, ARMA) 특징

1) 모형의 정상성과 가역성

	AR(p)	MA(q)	ARMA(p,q)
정상성	조건필요	자체만족	조건필요
가역성	자체만족	조건필요	조건필요

2) 모형의 ACF, PACF

	AR(p)	MA(q)	ARMA(p,q)
ACF	지수적으로 감소	q+1차부터 절단	q+1시점부터 지수적으로 감소
PACF	p+1차부터 절단	지수적으로 감소	p+1시점부터 지수적으로 감소

(2) 모형 적합과정

① 모형의 식별(model identification)단계 : 모형의 차수를 결정하는 과정

② 모수의 추정(model estimation)단계 : 식별된 모형의 모수들을 추정하는 과정

: $\phi_1, \phi_2 \dots \phi_p, \theta_1, \theta_2, \dots \theta_q$, 평균수준 μ , 백색잡음 과정의 분산 σ_ε^2

③ 모형의 진단(diagonostics)단계 : 잠정모형(tentative model)의 타당성 검토과정

④ 예측모형(forecast model)으로 선택

현실에는 정상성을 따르는 데이터보다, 비정상적인 데이터가 더 많다.

따라서 이번 주차에는 **비정상 시계열 모형**들을 다룰 것이다.

---

2. ARIMA

- 1주차에서 추세를 가지는 비정상 시계열을 회귀/평활/차분을 통해서 정상화시킬 수 있음을 배웠다.
- 만일 d차 차분된 $W_t = (1-B)^d Z_t$ 가 평균수준이 μ 인 정상 ARMA(p,q)과정을 따를 때 Z_t 는 자기회귀누적이동평균과정(ARIMA)를 따른다.
- 다시말하면, ARMA모형에서 차분까지 포함시킨 모형이 ARIMA모형이다.

(1) 표현

✓ ARIMA(p, d, q) :

$$\phi(B)(1-B)^d(Z_t-\mu) = \theta(B)\varepsilon_t$$
$$(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)(1-B)^d Z_t = (1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q)\varepsilon_t$$

- AR차수 p / 차분 차수 d / MA차수 q
- ε_t는 백색잡음 ($\varepsilon_t \sim WN(0,\sigma^2)$)

➔ 이때, I는 **누적**(Integrated)의 의미

: $\phi_B(1-B)Z_t = \phi(B)\varepsilon_t$에서

$W_t = (1-B)Z_t$이라고 한다면, $W_t = Z_t - Z_{t-1}$

$$Z_t = W_t + Z_{t-1} = W_t + (W_{t-1} + Z_{t-2}) = W_t + W_{t-1} + (W_{t-2} + Z_{t-3}) = \cdots = Z_0 + \sum_{j=1}^{t} W_j$$

따라서 Z_t부분은 W_t의 누적합이 된다.

*cf) ARIMA(p, 0, q) = ARMA(p, q)

(2) ARIMA모형 적합과정

① 시계열 그림과 ACF 그래프를 통해 비정상인지 정상인지 판단

: 만약 추세가 존재한다면, ACF가 천천히 감소

② 비정상 시계열이라고 판단이 된다면, 차분을 통해 정상화

: 일반적으로 d=1이나 d=2의 낮은 차수를 많이 사용 -> **과대차분**위험

③ 모형의 모수 추정

④ 적절한 모형인지 진단

⑤ 최종 모형 선택

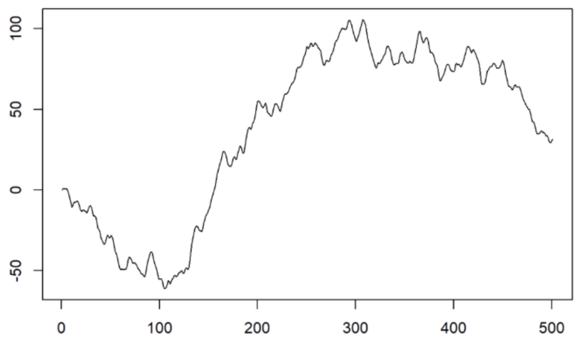
(3) 과대차분

- 정상화가 된 경우에도 차분을 시도하는 경우
- 정상적시계열을 차분해도 정상적 시계열의 선형결합이기 때문에 정상성에 문제가 없음
- 그러나, 지나친 차분은 ACF를 복잡하게 만들거나 분산을 크게 만든다는 문제점 존재

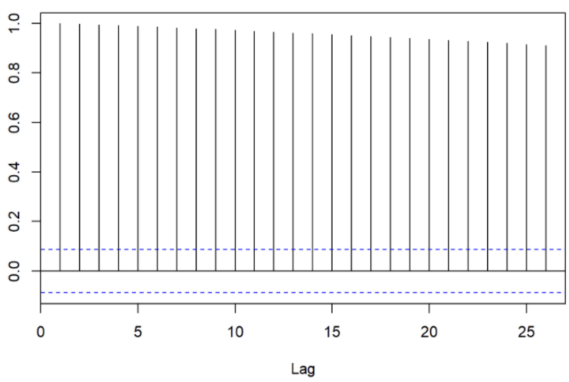
➔ 차분후에 시계열의 분산이 차분전의 시계열 분산보다 더 커지면 과대차분일 가능성이 높다고 보아 차분을 하지 않음

(4) 그래프

[ARIMA 데이터의 plot]



[ARIMA 데이터의 ACF]



- plot을 보았을 때 추세가 추세가 존재하는 것처럼 보임
- ACF가 느리게 감소

➔ ARIMA모형이 필요해 보임

3. SARIMA

1주차때 계절성을 가지는 시계열을 다음과 같이 가정했었다.

$$\mathrm{Y_t} = S_t + I_t$$

이때 계절성을 회귀, 평활 등을 통하여 계절성분 $\mathrm{S_t}$를 분해하여 제거하였다.

이렇게 계절성을 분해할 때는 '결정적'특성(disterministic)을 지닌 계절성분을 가정한다.

다시 말해, 계절 성분이 확률적이지 않고, 항상 식을 통해 도출되는 $\mathrm{S_t}$를 가진다.

그러나 이는 가정이며, 계절성이 고정되어 있지 않고 확률과정을 따를 수도 있다.

즉, 계절성이 결정적 함수($S_t = a_0 + \sum_{j=1}^{k} a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)$)가 아니라 $(1 - \mathrm{B^d})\mathrm{S_t} = X_t$를 만족시키는 확률과정(이때, $\mathrm{X_t}$는 정상적 확률과정)을 따른다.

SARIMA는 이러한 확률적 **계절성**을 고려할 수 있는 모형이다.

SARIMA에는 크게 순수SARIMA, 승법SARIMA가 있다.

➔ 다음 예시를 통해 SARIMA의 흐름을 알아보자

예시)

월단위 데이터가 있을 때, 이 데이터의 주기는 12이다.

이 데이터가 시계열이 ARMA(P, Q)과정을 따른다고 생각해보자

즉, 12개월마다 비슷한 패턴이 반복된다.

이때, 현재 관측치를 1년전 과거 관측치, 2년전 과거 관측치, … P년전 과거 관측치와 1년전 오차항, 2년전 오차항, … Q년전 오차항으로 설명할 수 있을 것이다.

이를 식으로 적어본다면,

$$\mathrm{Y_{j+12t}} - \Phi_1 Y_{j+12(t-1)} - \cdots - \Phi_P Y_{j+12(t-P)} = U_{j+12t} + \Theta_1 U_{j+12(t-1)} + \cdots + \Theta_Q U_{j+12(t-Q)}$$
$$\Phi(\mathrm{B^{12}}) Y_t = \Theta(B^{12}) U_t\,, \qquad U_t \sim WN(0, \sigma^2)$$

으로 적을 수 있다.

그러나, 이때 $\mathrm{U_t}$ 가 백색잡음이 아니라, ARMA모형을 따를 수도 있다.

$$\mathrm{U_t} \sim ARMA(p, q)$$
$$\phi(\mathrm{B})\mathrm{U_t} = \theta(B)\varepsilon_t$$
$$\mathrm{U_t} = \phi(B)^{-1}\theta(B)\varepsilon_t$$

이를 한번에 정리한다면, 다음과 같이 적을 수 있다.

$$\Phi(\mathrm{B^s}) Y_t = \Theta(B^s)\phi(B)^{-1}\theta(B)\varepsilon_t$$

이때, $\mathrm{Y_t}$가 d번 차분, D번 계절차분을 마친 상태의 시계열이라면,

$$\mathrm{Y_t} = (1-B)^d(1-B^s)^D Z_t$$

만족하고, 다음과 같이 적을 수 있다.

$$\phi(\mathrm{B})\Phi(\mathrm{B}^{12})(1-B)^d(1-B^s)^D Z_t = \theta(B)\Theta(B^{12})\varepsilon_t$$

(1) 순수SARIMA

- 시계열이 순수하게 계절형일 경우 때 적합할 수 있는 모형
- $\mathrm{ARIMA(P, D, Q)_s}$

1) 표현

$$\Phi(\mathrm{B^s})(1-B^s)^D Z_t = \delta + \Theta(B^s)u_t \quad \text{혹은} \; \Phi(\mathrm{B^s})(1-B^s)^D Z_t = \Theta(B^s)\varepsilon_t$$
$$\Phi(\mathrm{B^s}) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$$
$$\Theta(\mathrm{B^s}) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}$$

- 이때, ε_t 는 백색잡음
- s는 계절 주기

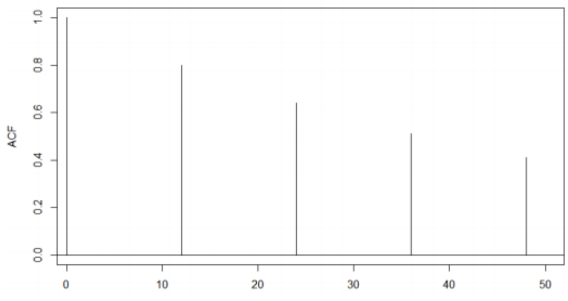
2) 특징

- 현재의 관측치를 계절성을 고려하여 $\mathrm{Z_{t-s}}, Z_{t-2s}, \ldots$들로 설명
- 해당 주기 관측치는 P개를 고려하며, 해당 주기 오차항은 Q개를 고려
- D번 계절차분
- 비계절 요소에 대한 부분을 전혀 고려하지 않기에, 사용이 제한적
- 오차가 백색잡음이므로, 계절주기에 해당되지 않는 시차의 ACF는 0

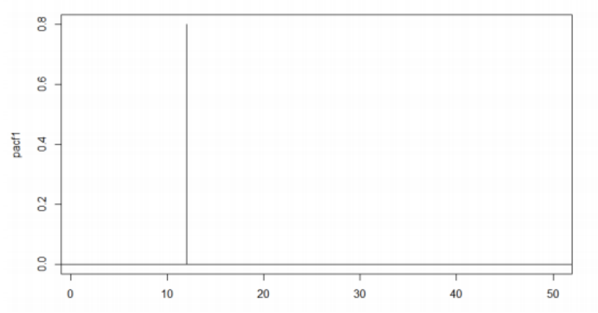
3) ACF, PACF

- ARMA 모형의 ACF와 PACF와 비슷함
- 단, 계절주기 차수에서만 자기상관성을 가짐

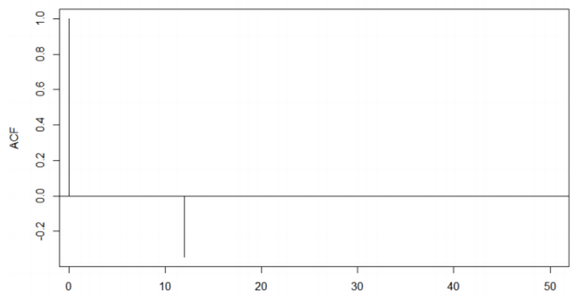
[$\text{ARIMA}(1,0,0)_{12}$ ACF]



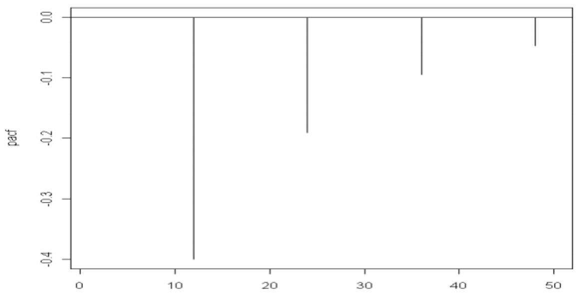
[$\text{ARIMA}(1,0,0)_{12}$ PACF]



[$\text{ARIMA}(0,0,1)_{12}$ ACF]



[$\text{ARIMA}(0,0,1)_{12}$ PACF]



(2) 승법SARIMA

- 순수SARIMA가 비계절 요소를 전혀 고려하지 않는 모형인 반면, 승법SARIMA 이를 고려하는 모형
- 순수SARIMA모형의 오차가 ARIMA를 따름
- 순수SARIMA에 ARIMA가 곱하기로 붙어서 '승법'이라는 용어 사용
- $\text{SARIMA}(\text{p},\text{d},\text{q})(\text{P},\text{D},\text{Q})_\text{s}$

1) 표현

$$\phi(\text{B})\Phi(\text{B}^s)(1-B)^d(1-B^s)^D Z_t = \theta(B)\Theta(B^s)\varepsilon_t$$

$$\Phi(\text{B}^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$$
$$\Theta(\text{B}^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}$$
$$\phi(\text{B}) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$
$$\theta(\text{B}) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

- 이때, ε_t 는 백색잡음
- s는 계절 주기

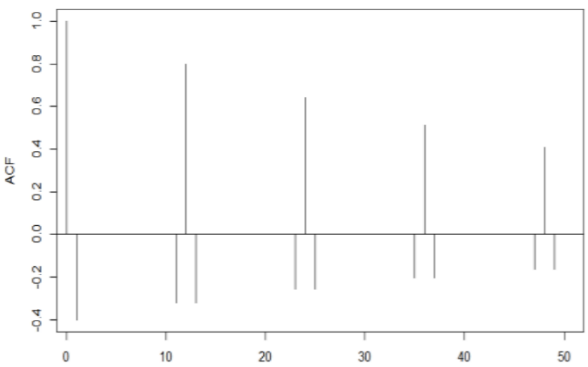
2) 특징

- 계절성과 추세를 고려한 모형

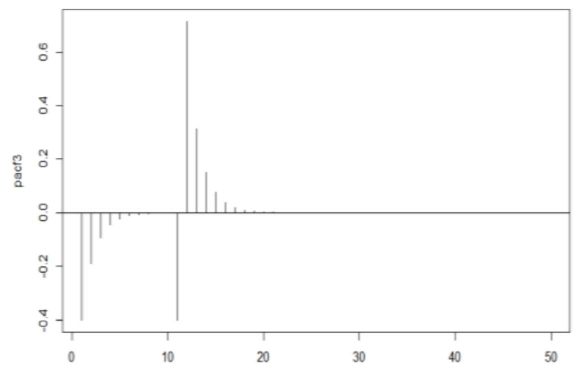
- ARIMA 모형처럼 예측시점과 가까운 과거 자료를 이용할 뿐만 아니라, 원자료의 주기적 특성을 반영하여 그 이전 주기의 자료를 추가적으로 활용
- 순수SARIMA보다 비계절 모형과 혼합된 모형이 모수절약의 원칙에 따라 더 자주 사용됨

3) ACF, PACF

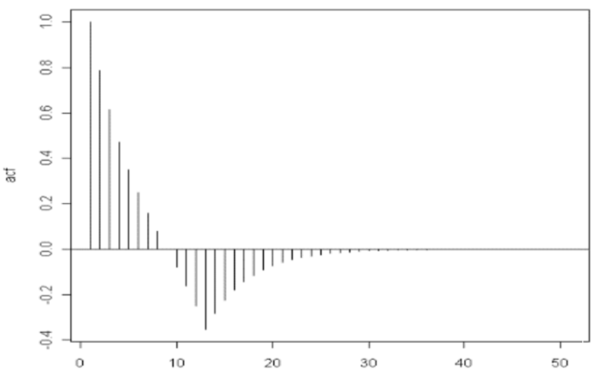
[$\text{SARIMA}(0,0,1)(1,0,0)_{12}$ ACF]



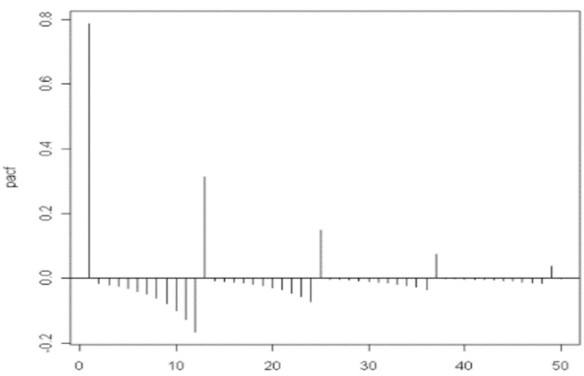
[$\text{SARIMA}(0,0,1)(1,0,0)_{12}$ PACF]



[$\text{SARIMA}(1,0,1)(0,0,1)_{12}$ ACF]



[$\text{SARIMA}(1,0,0)(0,0,1)_{12}$ PACF]



4. ARFIMA

앞에서 배운 ARMA모형의 기하급수적으로 감소하는 ACF를 가졌었다.

다시 말해 시간이 지날수록 과거 자료의 영향이 빠르게 줄어드는 모형이었다.

따라서 이는 단기억 확률과정(short memory process)라고도 부른다.

그러나, 어떤 정상 시계열 자료의 자기상관함수는 정상성을 만족함에도 불구하고 단기억 확률과정에서의 형태와 다르게 매우 느리게 0으로 접근한다.

이와 같은 경우 흔히 정상성이 만족되지 않는다고 판단하여 ARIMA를 이용하여 자료분석을 하게 되는데, 실제로

자료가 정상성 특징을 가지고 있으므로, 좋은 적합이라고 할 수 없다.

따라서 0으로 느리게 수렴하는 자기상관함수를 갖는 **장기억 확률과정**(long memory process)의 필요성이 대두되었다.

이것이 ARFIMA(Auto-regressive Fractionally integrated Moving Average)이다.

(1) 표현

- ✓ ARFIMA(p, d, q)

$$\phi(B)(1-B)^d(Z_t-\mu)=\theta(B)\varepsilon_t \quad , \text{이때 } 0<d<\frac{1}{2}$$

$$\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)(1-B)^d Z_t=\left(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q\right)\varepsilon_t$$

- 이때, ε_t 는 백색잡음

➔ ARIMA와 동일하나 d가 양의 정수가 아닌 **유리수**라는 점이 다름

(2) 특징

- d는 부분차분차수로 장기간동안의 시계열 종속성 결정
- d가 -0.5에서 0일 경우 자기상관합이 0이되는 문제가 발생

✓ 장기억 확률과정(long memory process)

정상성을 만족하는 확률과정 $\{Z_t\}$의 자기상관함수 $\rho(k)$가 $0<d<\frac{1}{2}$인 어떤 실수에 대하여

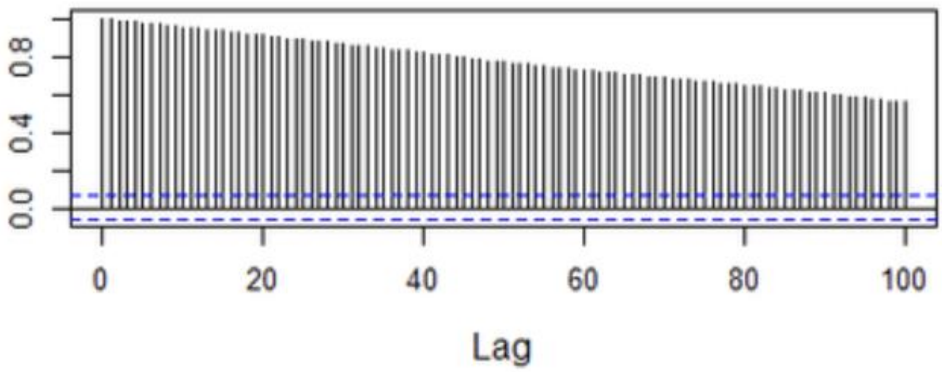
$$\rho(k) \sim Ck^{2d-1}, k\to\infty \ \left(\text{단}, C>0\right)$$

을 만족한다.

➔ ACF의 합이 무한대로 발산한다.

➔ ACF가 빠르게 0으로 수렴하지 않는다

[Long memory process의 ACF]



(3) ARFIMA의 적합

- ARFIMA에서 d는 최소제곱법, 최대우도추정법으로 찾을 수 있음
- 이후 ARMA모형과 비슷하게 나머지 모수들을 추정
- R에서는 arfima() 함수를 사용

<참고> ARIMAX

- 일종의 회귀모형
- 기존 ARIMA모형에서 시간에 따라 변화하는 외부요인(X)을 추가시켜준 모델
- 외부요인을 추가함으로써 다변량 시계열 데이터를 활용하기에 예측력을 높일 수 있다는 장점이 있음

(1) 표현

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \beta X$$

- 이때 X는 수치형일 수도 범주형일 수도 있음

(2) 적합 과정

① 관측치와 외부변수에 대한 회귀식을 적합시킴

② 적합시킨 뒤 나온 잔차에 대해 ARIMA식을 적합시켜 각 계수를 추정

✓ arima함수의 xreg인자를 사용해서 외생변수 적합

5. 이분산 시계열 모형 : ARCH, GARCH

ARIMA모형과 같은 전통적인 시계열 모형은 평균부분(1차 적률)의 움직임에 초점을 맞추며, 분산이 시간에 따라 변하지 않음(등분산; homoskedasticity)을 가정한 모형이다.

환율 주가 수익률, 선물 등의 금융관련 재무시계열 모형은 불확실성을 의미하는 분산부분(2차 적률)의 움직임에 관심이 있으며, 비선형적인 요소를 가지고 분산이 과거의 자료에 의존하는 특성을 지닌다.

따라서, 이분산(heteroskedasticity) 모형이 필요

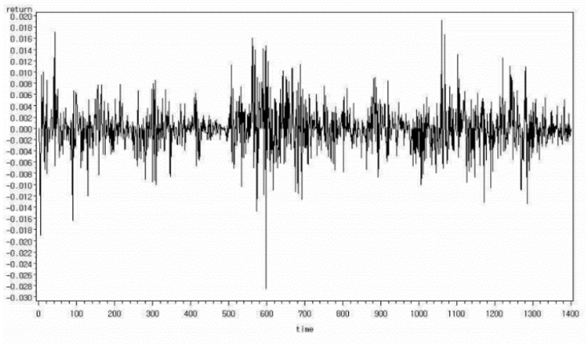
✓ **변동성(volatility)** : 위험(risk)를 측정하는 수단 = 조건부분산(conditional variance)

➔ 금융시계열에서는 위험관리가 매우 중요하기에 변동성 또는 조건부 분산은 매우 중요한 개념

(1) 변동성집중(volatility clustering) 경향

[일별 대미 원화환율]

[대미 원화환율의 로그수익율]

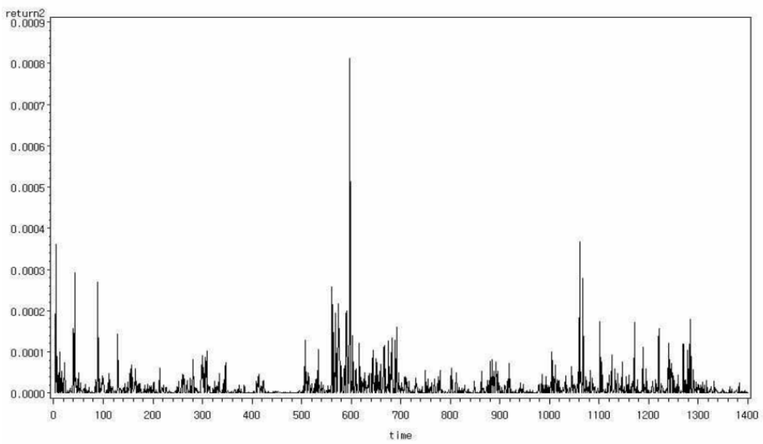


첫번째 그림은 추세를 보이므로 차분으로 정상화 후 분석을 하면 된다.

그러나 환율자료 대신 로그수익률 그래프를 보면, 확률적 추세가 없는 정상적 그림처럼 보이지만, 변동폭이 정상시계열과는 다르게 더 크게 나타나는 것을 볼 수 있다.

로그수익률을 제곱한 값(2차 적률)의 그래프를 그려보면 아래와 같다.

[대미 원화환율 로그수익율 제곱]



한번 큰 값을 기록하면 상당시간 동안은 큰상태로 지속된 후 다시 작은 값들이 상당기간동안 지속되는 현상을 볼 수 있다.

이러한 현상을 **변동성집중**(volatility clustering) 성향이라고 한다.

변동성 집중 현상은 분산σ^2에 자기상관성이 존재한다는 것을 보여준다.

오차항이 일정할 것이라는 가정 위배한다.

따라서 이분산 시계열 모형이 필요하다.

➔ 조건부 분산σ^2 시간의 함수로 모형화

(2) ARCH (Auto Regressive Conditional Heteroskedasticity)

- 변동성을 설명하기 위한 비선형 모델
- 현재 시점의 오차항의 변동성(조건부분산) σ_t^2을 과거 시점 오차항의 제곱으로 설명

1) 표현

- $\varepsilon_t \sim ARCH(p)$: 차수가 p인 자기회귀이분산 (autoregressive heteroskedastic)모형
- t시점의 오차항의 변동성을 p시점전의 오차항들의 제곱으로 설명

$$\varepsilon_t = \sigma_t \cdot \nu_t$$
$$\{\nu_t\} \sim i.i.d. N(0,1)$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2$$
$$\text{단, } \alpha_0 \geq 0, \alpha_j^2 \geq 0,\ j = 1, \dots, p$$

➔ 조건부 분산 : $E(\varepsilon_t^2 | \varepsilon_t^2, \varepsilon_t^2, \dots) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2$ -> 시간의 함수로 표현가능

2) 비선형 성질

$$\text{ARCH(1)} : \varepsilon_t = \sigma_t \cdot \nu_t,\ \ \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

$\varepsilon_t^2 = \sigma_t^2 \nu_t^2$ 이며, 이를 활용하여 계산하면

$$\varepsilon_t^2 = \sigma_t^2 \nu_t^2 = (\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)\nu_t^2 = \left(\alpha_0 + \alpha_1(\sigma_{t-1}^2 \nu_{t-1}^2)\right)\nu_t^2 = \left(\alpha_0 + \alpha_1\left((\alpha_0 + \alpha_1 \varepsilon_{t-2}^2)\nu_{t-1}^2\right)\right)\nu_t^2$$
$$= \cdots = \alpha_0 \sum_{j=0}^{n} \alpha_1^j \nu_t^2 \nu_{t-1}^2 \dots \nu_{t-j}^2 + \alpha_1^{n+1} \nu_t^2 \dots \nu_{t-n}^2$$

➔ ε_t가 ν_t의 선형적인 형태가 아니라 비선형적 형태

3) ARCH모형의 추정과 검정

귀무가설 $H_0: \alpha_1 = \cdots = \alpha_p = 0$ 에 대한 가설검정을 수행함으로써 ARCH모형의 효과를 검정할 수 있다.

① LM(largrange multiplier) 검정 : 검정통계량 $n \cdot R_\varepsilon^2$ 는 점근적으로 $\chi^2(p)$를 따른다는 성질을 활용하여 검정(이때, R_ε^2 는 다중상관계수의 제곱(squared multiple correlation))

② Ljung-Box Q 검정 : ARCH(p) 모형에서 ε_t^2가 AR(p)모형을 따름을 이용

4) 특징

- ε_t 는 시간에 따른 상관관계가 없으나(serially uncorrelated), ε_t^2 는 상관관계가 있으므로, ε_t 는 독립이 아님
- 실제로 필요한 p의 값이 상당히 클 수 있어 모수가 너무 많아지는 문제가 발생할 수 있음
- 추정해야 할 모수가 많으면 추정량의 정확도가 떨어짐
- 또한 추정해야 할 모수가 많아 비음조건(non-negative)조건을 만족하기 어려움

(3) GARCH

- ARCH모형이 추정해야 할 모수가 많은 문제를 해결하기 위해 보다 일반적인 모형의 도입 필요
- ARCH모형을 일반화한 모형으로 변동성을 설명하는데 있어 과거시점의 오차항의 제곱과 함께 과거시점의 변동성(조건부 분산)까지 고려한 모형

1) 표현

- $\varepsilon_t \sim \mathrm{GARCH}(p, q)$: 일반화 자기회귀이분산 (Generalized autoregressive heteroskedastic)모형
- t시점의 오차항의 변동성을 p시점전의 오차항들의 제곱과 q시점의 변동성으로 설명

$$\varepsilon_t = \sigma_t \cdot \nu_t$$
$$\{\nu_t\} \sim i.i.d.\, N(0,1)$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2$$
$$= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$
단, $\alpha_0 \geq 0, \alpha_j^2 \geq 0, \beta_j^2 \geq 0 \;\; j = 1, \ldots, p$

2) 추정해야 할 모수가 적은 이유

GARCH(1,1)의 예시를 이용해서 알아보자

$$\sigma_t^2 = \alpha_0 + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$
$$(1-\beta L)\sigma_t^2 = \alpha_0 + \alpha\epsilon_{t-1}^2$$
$$\sigma_t^2 = \frac{\alpha_0}{(1-\beta L)} + \frac{\alpha}{(1-\beta L)}\epsilon_{t-1}^2$$
$$= \frac{\alpha_0}{(1-\beta L)} + \alpha\sum_{j=1}^{\infty}\beta^{j-1}\epsilon_{t-j}^2$$

→ GARCH(1,1) 으로 ARCH(∞) 표현 가능

- ARCH에서 많은 모수들을 추정해야 했던 문제를 해결할 수 있음
- ARCH에서 비음조건(non-negative)조건을 만족시키기 쉬워짐

6. R실습

7. 3주차 클린업을 마치며

시계열자료는 시간에 흐름에 따라 관측되는 자료로 시간의 영향을 받는 데이터입니다. 따라서 자료 사이에 자기상관성이 있기에 시간적 특성을 고려하지 않는 일반적 분석방법으로는 제대로 된 분석을 할 수 없습니다. 우리는 3주동안 이러한 시계열 데이터를 어떻게 분석할 수 있는지 가장 기본적인 방법들을 알아보았습니다.

1주차에는 시계열데이터들의 특성들을 배웠고, 시계열 자료를 분석할 때 무한한 데이터들의 결합 확률 분포를 고려하여 분석을 하기는 어렵기에 정상성이라는 가정이 필요하다고 했습니다. 비정상 데이터들의 특징들과 비정상 데이터들을 어떻게 정상적 데이터들로 바꾸는지 배웠으며, 마지막으로 대표적인 정상 시계열인 백색잡음과 이를 검정하는 방법들을 배웠습니다. 백색잡음은 시계열에서 매우 중요한데, 많은 모형에서 백색잡음을 사용하기에 백색잡음의 특성을 정확히 알고 있는 것이 중요합니다. 2주차에는 1주차에서 배웠던 정상적 시계열 자료에 적합할 수 있는 정상 시계열 모형 AR, MA, ARMA를 배웠습니다. AR모형은 현재 관측치를 과거 관측치들로 설명하는 모형이며, MA모형은 과거 오차항들로 설명하는 모형이었습니다. 이렇게 관측치들로만, 오차들로만 데이터를 설명하기엔 차수가 커질 위험이 있으므로 이들을 결합한 ARMA모형을 사용하면 더 적은 모수들로 데이터들을 잘 설명할 수 있다고 배웠습니다. 더불어 AR, MA, ARMA모형들의 조건, 특징들을 배웠습니다. 현실에는 정상적 데이터보다는 비정상 데이터가 더 많기에, 3주차에는 비정상 데이터에 사용할 수 있는 여러 비정상 시계열 모델들을 배웠습니다. 데이터에 확률적 추세가 존재할 때 사용할 수 있는 ARIMA모형, 계절성이 존재할 때 사용할 수 있는 SARIMA모형을 배웠습니다. 이 모델들은 1주차때 배운 차분과 계절차분을 포함한 ARMA모형이었습니다. 또한, 장기억을 가지는 데이터에 일반석 ARIMA모형을 적합하면 모수가 커질 문제점이 있기에 이러한 데이터에 사

용할 수 있는 ARFIMA모형을 소개했고, 간단하게였지만 다변량 시계열 모형중 하나인 ARMAX에 대해서 소개했습니다. 지금까지 배웠던 모형들은 등분산성을 가정하였는데, 많은 데이터들이 등분산을 가지지 않습니다. 따라서 마지막으로 이분산성을 고려할 수 있는 모형인 ARCH에 대해서 배웠습니다. ARCH또한 차수가 커질 위험이 있으므로, 변동성을 같이 고려할 수 있는 GARCH모형을 배웠습니다.

시계열이 수식도 많고, 증명도 많지만 이러한 큰 흐름들을 잘 이해했으면 합니다. 시계열 자료의 특성부터 어떻게 이러한 특성들을 고려해 분석을 하는지. 그러한 모형에는 어떤 문제점이 있고, 그 문제점을 고려할 수 있는 모형들은 어떤 것이 있는지. 3주동안 여러가지들을 배웠지만, 흐름은 쭉 이어진다고 생각합니다!

지금까지 배운 것은 시계열에서 정말 기본적인 것들입니다. 다변량 회귀모형인 VAR(벡터자기회귀모형)부터 스펙트럴 분석, 개입분석 등 시계열분석에는 많은 모형들과 분석방법들이 있습니다. 또한, 지금까지 배운 시계열 모형들은 선형 모형을 가정하기 때문에 현실에는 어려움이 많습니다. 그래서 최근에는 머신러닝, 딥러닝을 통하여 시계열 데이터를 분석한다고 합니다. 전통적인 통계적 분석방법과 똑같이 머신러닝 또한 손실함수를 최소화하는 방법으로 예측의 정확도를 향상시킵니다. 다만, 다른점은 비선형 알고리즘을 사용한다는 것입니다. 딥러닝 분야에서는 RNN종류의 LSTM이 좋은 성능을 낸다고 합니다. 아직 배운 것은 정말 기초중의 기초지만, 가장 기본적인 모형들을 다루었기에 배웠던 것들을 바탕으로 앞으로 시계열분석을 공부한다면, 수월하게 공부할 수 있을 것이라고 생각합니다!

드디어 길었다면 길었고 짧았다면 짧았던 3주의 클린업이 끝났습니다~!! 시계열자료분석팀이라는 어려운 팀에 들어와서 3주동안 어려웠을텐데, 잘 따라와줘서 너무 고마워요ㅎㅎ 다른팀에 비해 수식도 많고 증명도 많았는데, 재미가 없을 까봐 걱정을 많이 했어요ㅠㅠ 우리가 살아가는 세계에는 시간이라는 특성이 당연히 개입될 수 밖에 없기에, 많은 데이터들이 시계열 데이터입니다. 앞으로 데이터 분석 할 때 클린업에서 배운 것들이 많은 도움이 되기를 바랍니다~! (시계열 모델쓰면 멋져보이는건 덤-ㅎㅎ) 열정적이고, 멋진 팀원들 덕분에 저도 자극 받아서 열심히 할 수 있었던 것 같아요!-! 3주동안 너무 고생했다는 말 하고싶었고, 부족한 팀장 믿고 따라와줘서 고맙다는말 하고 싶었습니다!! 주제분석때도 파이팅해봐요오 😛