클린업 3주차

5팀 시계열자료분석팀

염예빈
한유진
이재현
박세령
이정우

ARIMA

Auto-Regressive Integrated Moving Average

정상화 방법 중 차분 후

ARMA(p,q) = ARIMA(p,d,q)

즉, 차분과정까지 포함한 모델!!

Trend(추세) + ARMA(p,q)형태가 있을 때 사용!

ARIMA의 표현

ARIMA에서 I는 무슨 뜻인가요?

I는 누적(integrated)의 의미

$$\phi(B)(1-B)^{d}(Z_t-\mu)=\theta(B)\varepsilon_t$$

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

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

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

$$Z_t=W_t+Z_{t-1}=W_t+(W_{t-1}+Z_{t-2})=W_t+W_{t-1}+(W_{t-2}+Z_{t-3})$$

$=\cdots=Z_0+\sum_{j=1}^{t}W_j \quad \therefore \quad Z_t$부분은 W_t의 누적합이 된다.

ARIMA 모형의 적합

일반적으로 d=1 or 2 낮은 차수를 많이 사용.
과대 차분 위험때문!!!

차분을 통한 정상화

차분한 시계열이 ARMA(p,q) 과정을 따르는지 살펴보고
ARIMA(p,d,q)에 적합!

순수 SARIMA

시계열이 순수하게 계절형일 경우 때 적합할 수 있는 모형

$$\mathbf{ARIMA}(\mathbf{P},\mathbf{D},\boldsymbol{Q})_{\boldsymbol{s}}$$

$$\Phi(\mathrm{B}^s)(1-B^s)^D Z_t = \delta + \Theta(B^s)u_t \quad \text{혹은} \quad \Phi(\mathrm{B}^s)(1-B^s)^D Z_t = \Theta(B^s)\varepsilon_t$$

$$\Phi(\mathrm{B}^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$$

$$\Theta(\mathrm{B}^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}$$

- 현재의 관측치를 계절성을 고려하여 $\mathrm{Z}_{\mathrm{t-s}}, Z_{t-2s}, \dots$들로 설명
- 해당 주기 관측치는 P개를 고려하며, 해당 주기 오차항은 Q개를 고려
- 오차 ε_t는 백색잡음 , D번 계절차분

승법 SARIMA

$$\mathrm{Y}_{\mathrm{j+12t}} - \Phi_1 Y_{j+12(t-1)} - \cdots - \Phi_P Y_{j+12(t-P)} = U_{j+12t} + \Theta_1 U_{j+12(t-1)} + \cdots + \Theta_Q U_{j+12(t-Q)}$$

$$\Phi(\mathrm{B}^{12}) Y_t = \Theta(B^{12}) U_t , \qquad U_t \sim WN(0, \sigma^2)$$

$\mathrm{U_t}$ 가 백색잡음이 아니라 ARMA 모형을 따를 수도 있음!

$$\mathrm{U_t} \sim ARMA(p, q)$$

$$\phi(\mathrm{B})\mathrm{U_t} = \theta(B)\varepsilon_t, \qquad \mathrm{U_t} = \phi(B)^{-1}\theta(B)\varepsilon_t$$

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

승법 SARIMA

$$\mathrm{U_t} \sim ARMA(p,q)$$

$$\phi(\mathrm{B})\mathrm{U_t} = \theta(B)\varepsilon_t, \quad \boxed{\mathrm{U_t} = \phi(B)^{-1}\theta(B)\varepsilon_t}$$

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

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

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

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

승법 SARIMA

순수 SARIMA와 다르게, 승법 SARIMA는 비계절적 요소를 고려하는 모형
순수 SARIMA모형의 오차가 ARIMA를 따름

$$ARIMA(P,D,Q)_s \ \& \ U_t \sim ARIMA(p,d,q)$$

순수SARIMA에 ARIMA가 곱하기로 붙어서 '승법'이라는 용어 사용

$$\boldsymbol{SARIMA(p,d,q)(P,D,Q)_s}$$

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

$$\Phi(\mathrm{B}^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}, \Theta(\mathrm{B}^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}$$

$$\phi(\mathrm{B}) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \theta(\mathrm{B}) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

승법 SARIMA

$$SARIMA(p,d,q)(P,D,Q)_s$$

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

- 계절성과 추세를 고려한 모형
- ARIMA 모형처럼 예측시점과 가까운 과거 자료를 이용할 뿐만 아니라, 원자료의 주기적 … 자료를 추가적으로 활용
- 순수SARIMA보다 비계절 모형과 혼합된 모형이 모수절약의 원칙에 따라 더 자주 사용됨

순수 SARIMA

$$ARIMA(P,D,Q)_s$$

$$\Phi(B^s)(1-B^s)^D Z_t = \Theta(B^s)u_t$$

비계절적 요소를 고려하지 않음

승법 SARIMA

$$SARIMA(p,d,q)(P,D,Q)_s$$

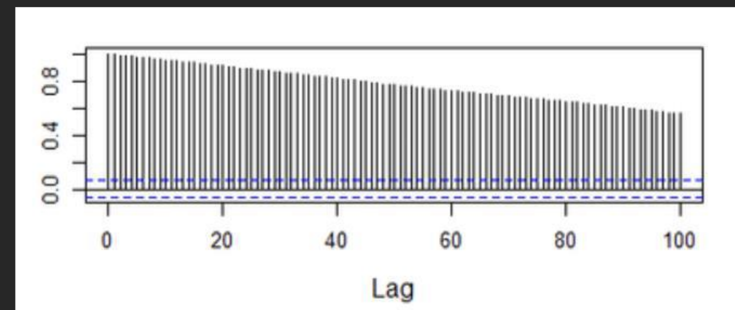
$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D Z_t = \theta(B)\Theta(B^s)\varepsilon_t$$

오차항 $u_t \sim ARIMA(p,d,q)$

비계절적 요소를 고려함

ARFIMA

ARFIMA



MA

ARIMA

그 한계점을 보완한 모형이 **ARFIMA**!

ARFIMA 모형의 ACF는

0으로 느리게 수렴

→ 장기억 확률과정

ARFIMA

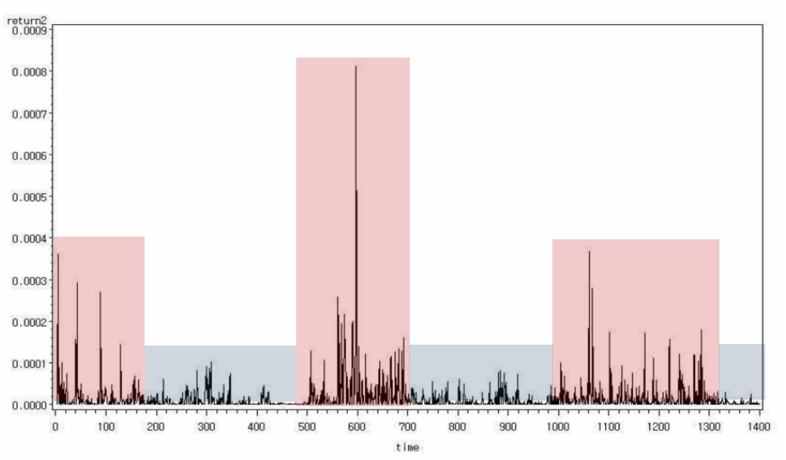
Auto Regressive Fractionally Integrated Moving Average

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

$$\Rightarrow (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(1-B)^d Z_t$$

$$= (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^p)\,\varepsilon_t$$

이분산 시계열 모형의 필요성



- **변동성 집중 현상** : 한번 큰 값을 기록하면 상당시간 동안 큰 상태로 지속된 후 다시 작은 값들이 상당기간동안 지속되는 현상.
- 변동성 집중 현상은 **분산에 자기상관성이 존재함**을 보여줌.

ARCH

[표현: ε_t ~ ARCH(p)]

$\{\nu_t\}$ ~ i.i.d.N(0,1)

$\alpha_0 \geq 0,\ \alpha_j \geq 0,\ j = 1, \cdots, p$

$$\varepsilon_t = \sigma_t \cdot \nu_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2$$

t시점의 오차항의 변동성을

P시점 전의 오차항들의 제곱으로 설명

ARCH

ARCH모형의 문제점

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + \cdots$$

추정해야 할 모수가 많음

- 추정량의 정확도가 떨어짐
- 비음조건(non-negative)을 만족하기 어려움

GARCH

일반화 자기 회귀 이분산 모형
= Generalized Auto Regressive Conditional Heteroskedastic

ARCH 모형을 일반화한 모형

현재 시점의 오차항 변동성 ← 설명 ← 과거 시점의 오차항 제곱, 과거 시점의 변동성