시계열 분석 기법과 응용

Week 6. 벡터자기회귀모형
6-1. 벡터자기회귀모형의 식별 및 추정

전치혁 교수
(포항공과대학교 산업경영공학과)

벡터자기회귀모형

- **벡터 시계열**
  - 여러 시계열을 동시에 고려하여 상호 연관성을 분석
  - 예 (3개의 시계열)

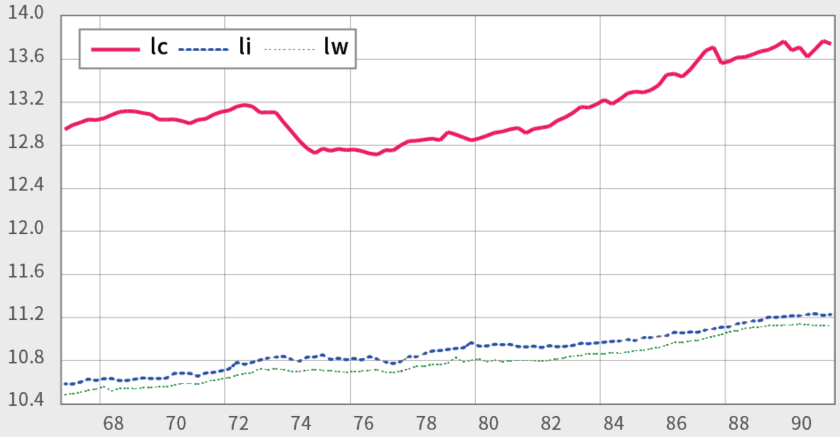
    시계열 1: Z_{1t} – 분기별 소비

    시계열 2: Z_{2t}– 분기별 소득

    시계열 3: Z_{2t}– 분기별 자산
  - 벡터시계열의 표현:

$$\boldsymbol{z}_t = \begin{pmatrix} Z_{1t} \\ Z_{2t} \\ Z_{3t} \end{pmatrix}, t = 1,2, \dots$$

  - 벡터자기회귀 (vector autoregressive; VAR)모형이 주로 사용



벡터자기회귀모형

- 벡터자기회귀 (VAR)모형
  - 한 시계열에 대한 AR(1) 모형

$$Z_t = \phi_1 Z_{t-1} + a_t$$

  - 두 시계열에 대한 VAR(1) 모형

$$Z_{1t} = \phi_{11} Z_{1,t-1} + \phi_{12} Z_{2,t-1} + a_{1t}$$
$$Z_{2t} = \phi_{21} Z_{1,t-1} + \phi_{22} Z_{2,t-1} + a_{2t}$$

  여기서 $a_{1t} \sim Nor(0, \sigma_1^2), a_{2t} \sim Nor(0, \sigma_2^2)$ 이며, $Cov[a_{1t}, a_{2t}] = \sigma_{12}$ ⇒ (벡터로 표현)

$$\boldsymbol{z}_t = \begin{pmatrix} Z_{1t} \\ Z_{2t} \end{pmatrix}, \boldsymbol{a}_t = \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}, \Phi_1 = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$$

$$\boldsymbol{z}_t = \Phi_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t, \boldsymbol{a}_t \sim MVN(\boldsymbol{0}, \Sigma), \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

VAR 모형

- m차원 시계열의 VAR(1) 모형

$$\boldsymbol{z}_t = \begin{pmatrix} Z_{1t} \\ \vdots \\ Z_{mt} \end{pmatrix}; \boldsymbol{z}_t = \Phi_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t$$ (여기서 Φ_1는 mxm 행렬)

- m차원 시계열의 VAR(p) 모형

$$\boldsymbol{z}_t = \Phi_1 \boldsymbol{z}_{t-1} + \Phi_2 \boldsymbol{z}_{t-2} + \cdots + \Phi_p \boldsymbol{z}_{t-p} + \boldsymbol{a}_t$$

$$\Phi(x) = I - \Phi_1 x - \cdots - \Phi_p x^p \Rightarrow \Phi(B)\boldsymbol{z}_t = \boldsymbol{a}_t$$

VAR(1)형태의 표현

$$\boldsymbol{y}_t = \begin{pmatrix} \boldsymbol{z}_t \\ \boldsymbol{z}_{t-1} \\ \vdots \\ \boldsymbol{z}_{t-p+1} \end{pmatrix}, F = \begin{bmatrix} \Phi_1 \Phi_2 & & \cdots & \Phi_p \\ I & 0 & & 0 \\ & \ddots & & \\ 0 & & I & 0 \end{bmatrix}, \boldsymbol{v}_t = \begin{pmatrix} \boldsymbol{a}_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Rightarrow \boldsymbol{y}_t = F \boldsymbol{y}_{t-1} + \boldsymbol{v}_t$$

VAR 모형

- 정상성 조건

VAR(1)모형 의 정상성 조건

- 모형: $\boldsymbol{z}_t = \Phi_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t$
- 계수행렬 Φ_1의 고유치들의 크기가모두 1보다 작아야 한다.
- $det\Phi(x) = \det(I - \Phi_1 x) = 0$ 의 근들의 크기가 모두 1보다 커야 한다.

VAR(p)모형 의 정상성 조건

- 모형: $\boldsymbol{z}_t = \Phi_1 \boldsymbol{z}_{t-1} + \Phi_2 \boldsymbol{z}_{t-2} + \cdots + \Phi_p \boldsymbol{z}_{t-p} + \boldsymbol{a}_t$

  $\Phi(B)\boldsymbol{z}_t = \boldsymbol{a}_t, \Phi(x) = I - \Phi_1 x - \cdots - \Phi_p x^p$
- $det\Phi(x) = \det\left(I - \ \Phi_1 x - \cdots - \ \Phi_p x^p\right) = 0$ 의 근들의 크기가 모두 1보다 커야 한다.
- $\boldsymbol{y}_t = F\boldsymbol{y}_{t-1} + \boldsymbol{v}_t$의 형태에서 행렬 F의 고유치들의 크기가 모두 1보다 작아야 한다.

VAR 모형

- 정상성 조건

(예) 다음의 두 시계열에 대한 VAR(2)모형이 정상적인지 알아보자.

$$\begin{pmatrix} Z_{1t} \\ Z_{2t} \end{pmatrix} = \begin{bmatrix} 0.3 & -0.1 \\ -0.05 & 0.25 \end{bmatrix} \begin{pmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{pmatrix} + \begin{bmatrix} 0.2 & -0.75 \\ -0.1 & 0.4 \end{bmatrix} \begin{pmatrix} Z_{1,t-2} \\ Z_{2,t-2} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}, \ \Sigma = \begin{bmatrix} 0.6 & 0.05 \\ 0.05 & 1 \end{bmatrix}$$

(풀이) $\boldsymbol{y}_t = F\boldsymbol{y}_{t-1} + \boldsymbol{v}_t$ 형태로 표시하면 행렬 F는 다음과 같다.

$$F = \begin{bmatrix} 0.3 & -0.1 & 0.2 & -0.75 \\ -0.05 & 0.25 & -0.1 & 0.4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

- 고유치는 다음과 같다. $\lambda_1 = 0.9627, \lambda_2 = 0.2376, \lambda_3 = -0.0356, \lambda_4 = -0.6148$

- 즉, 고유치 크기가 모두 1보다 작으므로 위의 VAR(2)모형은 정상적이다.

VAR 모형

- 모형의 식별 – 시차 p 결정

방법 1: 이론적 ACF/PACF를 구하고 표본 ACF/PACF와 비교

- 교차상관관계 등이 존재하므로 쉽지 않음
- 개별 시계열에 대한 표본 ACF/PACF를 관찰하고 충분한 시차 p를 추정할 수도 있음

방법 2: 정보기준 (information criteria) 사용

- 실제적으로 널리 사용되고 있음
- 여러 시차에 대해 AIC, BIC, HQ 등을 산출하고 정보기준값이 최소인 시차 선택
- $AIC(p) = ln|\hat{\Sigma}| + \frac{2m^2 p}{N}$: Akaike information criteria
- $BIC(p) = ln|\hat{\Sigma}| + \frac{m^2 p lnN}{N}$: Schwarz (Bayesian) information criteria
- $HQ(p) = ln|\hat{\Sigma}| + \frac{2m^2 p ln lnN}{N}$: Hannan –Quinn information criteria

방법 3: 우도비 검정

VAR 모형의 식별

(예) 다음은 세가지 벡터 시계열에 대한 여러 시차에 따른 VAR모형을 추정한 후 정보기준값을 산출한 결과이다.

적절한 시차 p를 선택하라.

- SC 기준으로는 시차 1이 가정 적절하다고 볼 수 있으나 대부분 기준은 시차 2가 적절하다고 선언하므로 결과적으로 시차 2가 최적이라 하겠다.
- 기준마다 서로 다른 최적 시차를 제시하는 경우에는 가장 작은 시차를 최종적으로 선택하는 것도 방법이라 하겠다.

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-1415.48	NA	5.05e+11	35.46	35.55	35.49
1	-1070.11	656.19	1.13e+08	27.05	27.41*	27.19
2	-1055.28	27.07*	9.74e+07*	26.90*	27.53	27.16*
3	-1046.40	15.53	9.79e+07	26.91	27.80	27.27
4	-1039.98	10.75	1.05e+08	26.97	28.14	27.44
5	-1038.50	2.37	1.28e+08	27.16	28.59	27.73

LR: sequential modified LR test statistic (each test at 5% level)
FPE: final prediction error