⟨**권이태, 0816**⟩

Vector AutoRegression

VAR

VAR은 벡터자기회귀 모형으로, 단순히 단변량 자기회귀모형의 벡터 버전이라 보기는 어렵다. 이는 변수 간의 동적인 상호작용을 묘사한다. 특히 이는 구조방정식모형의 축소된 형태이기도 하다. 아래의 n차원 VAR 모형을 고려하여 보자.

$$X_t = \mu + \Phi_1 X_{t-1} + \cdots + \Phi_p X_{t-p} + \epsilon_t$$

여기에서 사용되는 가정은 아래와 같다.

1. $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, \cdots, \epsilon_{nt})^T$는 벡터잡음과정으로 $\epsilon_t \sim (0, \Omega)$를 따른다.

2. 정상성 조건은 특성방정식 $\Phi(z) = |I - \Phi_1 z - \cdots - \Phi_p z^p| = 0$의 근 z가 모두 단위원 밖에 있는 것이다.

시차의 결정

아래의 가설을 생각하자.

$$H_0 : p = p_0, \quad \text{v.s.} \quad H_1 : p = p_1$$

그렇다면 $p_0 < p_1$이라 할 경우 이 귀무가설과 대립가설의 차이는

$$(\Phi_{p_0+1})_{ij} = \cdots = (\Phi_{p_1})_{ij} \ \forall (i,j) \in \{1, 2, \cdots, n\}^2$$

이 제약조건으로 주어지는지 여부에 달려 있다. 이 제약에서는 총 $n^2(p_1 - p_0)$개의 변수를 조정한다.

그렇다면 가능도비 검정에서의 검정통계량은

$$LR = T(\log|\hat{\Omega}_0| - \log|\hat{\Omega}_1|) \dot{\sim} \chi^2(n^2(p_1 - p_0))$$

으로 주어진다. 이때 $\hat{\Omega}_0, \hat{\Omega}_1$은 각각 귀무가설과 대립가설에서의 잔차를 이용해 계산한 Ω의 추정량이다. 조신섭 책에서는 $p_1 = p_0 + 1$일 때에 맞추어

$$M(p_0) = -(T - n - p_0 - 2.5) \log\left(\frac{|\hat{\Omega}_1|}{|\hat{\Omega}_0|}\right) \dot{\sim} \chi^2(n^2)$$

을 제시하고 있는데, 어차피 점근적인 검정이고 T가 큰 상황이면 큰 차이가 나지는 않는다. 여기에서는 Hamilton 책의 표현을 따르기로 한다. 일반적으로는 bottom-up 방식으로 $p_1 = p_0 + 1$처럼 둔 뒤 $p_0 = 0$부터 시작하여 더이상 귀무가설을 기각하지 못할 때까지 이 검정을 수행하고 그 시점에서의 시차를 VAR 모형의 시차로 사용한다.

혹은 AIC나 SBC와 같은 정보량 기준을 사용하여줄 수도 있다. 주의할 점은 AIC나 SBC의 부호나 더해지는/곱해지는 상수가 책마다 살짝씩 다르다는 것이다. 여기에서는 제가 좋아하는 표기법대로 씁니다.

$$\begin{aligned} \text{AIC}(p_0) &= T\log(|\hat{\Omega}_0|) + 2n^2 p_0 \\ \text{SBC}(p_0) &= T\log(|\hat{\Omega}_0|) + \ln(T) n^2 p_0 \end{aligned}$$

이들이 작은 p_0를 선택하는 것 또한 하나의 방법이 될 수 있다. 혹은 시계열도나 표본부분자기회귀행렬의 형태를 봐도 된다.

부분자기회귀 행렬은 아래의 다변량 Yule-Walker 방정식으로부터 구해지는 행렬 $\Phi_{k,k}$를 의미한다.

$$\begin{pmatrix} \Gamma(0) & \Gamma^T(1) & \cdots & \Gamma^T(k-1) \\ \Gamma(1) & \Gamma(0) & \cdots & \Gamma^T(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma(k-1) & \Gamma(k-2) & \cdots & \Gamma(0) \end{pmatrix} \begin{pmatrix} \Phi_{k,1}^T \\ \Phi_{k,2}^T \\ \vdots \\ \Phi_{k,k}^T \end{pmatrix} = \begin{pmatrix} \Gamma(1) \\ \Gamma(2) \\ \vdots \\ \Gamma(k) \end{pmatrix}$$

이때 $\Gamma(k)$는 k차 교차공분산 행렬 $\mathrm{Cov}(X_t, X_{t-k})$이다. $\Gamma(k) \neq \Gamma^T(k)$인 대신 $\Gamma(k) = \Gamma^T(-k)$인 점에 유의하라. 이렇게 구한 $\Phi_{k,k}$는 VAR 모형이 참일 때

$$\Phi_{k,k} = \begin{cases} \Phi_k & k \leq p \\ O & k > p \end{cases}$$

임이 알려져 있으므로, $\Phi_{k,k}$가 특정 차수 이후 영행렬이 되는 모습이 보인다면 그 차수를 VAR 모형을 적합하기 위한 차수 p로 사용할 수 있다.

VAR 모형의 적합과 진단

시차를 결정하였다면, 주어진 자료를 해당 시차 p에 맞추어 VAR(p) 모형을 적합할 수 있다. 단변량에서와 같이, 최소제곱추정량과 최대가능도추정량을 사용한다. 이때 초기값 $X_{-p+1}, \cdots, X_0$은 주어졌다고 가정하자. 즉 조건부최소제곱추정량 또는 조건부최대가능도추정량을 얻는다고 가정하자.

1. **최소제곱추정량**

   최소제곱추정량에서 추정해야 할 모수는 절편 n개와 계수행렬의 계수 n^2p개이다. 여기에서 풀고자 하는 문제는

$$\Big\{ \text{minimize} \quad \sum_{t=1}^T \|X_t - \Phi_1 X_{t-1} - \cdots - \Phi_p X_{t-p}\|_F^2$$

   으로 주어진다. 이를 풀어내기 위해서

$$Z_t = \begin{pmatrix} 1 \\ X_t \\ \vdots \\ X_{t-p+1} \end{pmatrix} \in \mathbb{R}^{np+1}$$

$$Z = \begin{pmatrix} | & | & & | \\ Z_0 & Z_1 & \ldots & Z_{T-1} \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{(np+1)\times T}$$

$$Y = \begin{pmatrix} | & | & & | \\ X_1 & X_2 & \ldots & X_T \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{n\times T}$$

$$y = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{pmatrix} \in \mathbb{R}^{nT}$$

$$B = \begin{pmatrix} \delta & \Phi_1 & \cdots & \Phi_p \end{pmatrix} \in \mathbb{R}^{n\times(np+1)}$$

$$\beta = \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_{np+1} \end{pmatrix} \in \mathbb{R}^{n(np+1)}$$

을 정의하면, 우리의 VAR 모형이

$$Y = BZ + E$$

혹은

$$y = (Z^T \otimes I_n)\beta + \epsilon$$

처럼 써질 수 있기에 최소제곱추정량이

$$\hat{\beta} = ((ZZ^T)^{-1}Z \otimes I_n)y$$
$$\hat{\Omega}_\epsilon = \frac{1}{T-(np+1)}(Y - \hat{B}Z)(Y - \hat{B}Z)^T$$

으로 주어진다.

2. **최대가능도추정량** 정규성 가정 하에서, B와 Ω_ϵ에 대한 로그가능도함수는 아래와 같이 주어진다.

$$\ln L(B, \Omega_\epsilon) = -\frac{nT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Omega_\epsilon| - \frac{1}{2}\mathrm{tr}((Y - BZ)^T\Omega_\epsilon^{-1}(Y - BZ))$$

따라서 미분을 통해 간단하게

$$\hat{B} = YZ^T(ZZ^T)^{-1}$$
$$\hat{\Omega}_\epsilon = \frac{1}{n}(Y - \hat{B}Z)(Y - \hat{B}Z)^T$$

를 얻을 수 있다.

한편 좋은 조건 하에서, 최소제곱추정량과 최대가능도추정량은 모두 일치추정량이며, 점근적으로는

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N_{n(np+1)}(0, (\Omega_\epsilon \otimes \mathbb{E}[Z_t Z_t^T]^{-1}))$$

이다.

이렇게 추정을 마무리하고 나면, 잔차를 통해 모형을 진단할 수 있다. 진단에서는 정보량 기준을 이용할 수도 있고, 각 일변량 시계열에 대한 잔차로써 일변량 시계열의 진단을 수행해도 된다. 그러나 전 시차에서 전 시계열의 잔차의 적합성을 보기 위해서는 **다변량 포트맨토검정**을 이용할 수 있다. 잔차의 교차상관계수를 $\rho(k)$처럼 쓸 때,

$$H_0 : \rho(1) = \cdots = \rho(K) = 0, \quad \text{v.s.} \quad H_1 : \exists k \text{ s.t. } \rho(k) \neq 0$$

을 검정하기 위한 포트맨토검정통계량은

$$Q_n(K) = T^2 \sum_{k=1}^{K} \frac{1}{n-k}\mathrm{tr}(\hat{\rho}(k)^T\hat{\rho}(0)^{-1}\hat{\rho}(k)\hat{\rho}(0)^{-1})$$

으로 주어지며 이는 점근적으로 자유도가 $n^2K - n^2p$인 카이제곱분포를 따른다. 이때 K는 잔차의 형태를 보고 선택해야 하는 값이다. 일반적으로 $\ln(T)$ 따위를 사용한다.

잔차진단을 통해 모형이 유효한 것으로 나타났다면, 그 이후 예측은 일반적인 일변량 시계열에서와 같다. 너무 간단하므로 여기에서는 다루지 않는다.

Granger-Causality

VAR에서 인과성이란 아래와 같이 정의된다. X가 Y를 **그레인저 인과**하지 않는다는 것은

$$\mathbb{E}[Y_{t+s}|Y_t, Y_{t-1}, \cdots, X_t, X_{t-1}, \cdots] = \mathbb{E}[Y_{t+s}|Y_t, Y_{t-1}, \cdots]$$

처럼 X의 현재/과거값이 Y의 미래값을 예측하는 데 어떠한 도움도 되지 않는다는 것을 의미한다. 이는 Y의 미래값에 대한 정보가 X의 현재/과거값에 포함되지 않는다는 것을 의미하기도 한다.

이변량 시계열에 대한 아래의 VAR(2) 모형을 고려해 보자.

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \phi_{11}(1) & \phi_{12}(1) \\ \phi_{21}(1) & \phi_{22}(1) \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \phi_{11}(2) & \phi_{12}(2) \\ \phi_{21}(2) & \phi_{22}(2) \end{pmatrix} \begin{pmatrix} x_{t-2} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

그렇다면

$$\mathbb{E}[x_t|x_{t-1}, x_{t-2}, y_{t-1}, y_{t-2}] = \phi_{11}(1)x_{t-1} + \phi_{12}(1)y_{t-1} + \phi_{11}(2)x_{t-2} + \phi_{12}(2)y_{t-2}$$

이므로 y_t가 x_t를 그레인저 인과하지 않을 필요충분조건은 $H_0 : \phi_{12}(1) = \phi_{12}(2) = 0$이다. 동일한 이유로, x_t가 y_t를 그레인저 인과하지 않을 필요충분조건은 $H_0 : \phi_{21}(1) = \phi_{21}(2) = 0$이 된다. 즉 귀무가설은 그레인저 인과관계가 없다는 쪽이다. 그 검정은 왈드검정을 이용한다. 예를 들어 y_t가 x_t를 그레인저 인과하는지 확인할 때,

$$x_t = \phi_{11}(1)x_{t-1} + \phi_{12}(1)y_{t-1} + \phi_{11}(2)x_{t-2} + \phi_{12}(2)y_{t-2} + \epsilon_{1t}$$

을 회귀분석하여 $\phi_{12}(1) = \phi_{12}(2) = 0$인지를 검정하는 것이다. 따라서 검정통계량은 회귀분석을 진행한 뒤

$$F = \frac{(\text{SSE(RM)} - \text{SSE(FM)})/2}{\text{SSE(FM)}/(T-5)} \sim F(2, T-5)$$

임을 이용하여 검정할 수 있다. 더욱 일반적인 VAR(p) 모형의 경우에는

$$F = \frac{(\text{SSE(RM)} - \text{SSE(FM)})/p}{\text{SSE(FM)}/(T-2p-1)} \sim F(p, T-2p-1)$$

임을 이용하여 검정할 수 있다.

한편 이변량 모형을 넘어서, 더욱 일반적인 n차원 시계열 X_t를 n_1, n_2차원 X_{1t}와 X_{2t}차원 시계열로 나누어 보자. 그렇다면

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Phi_{11}(1) & \Phi_{12}(1) \\ \Phi_{21}(1) & \Phi_{22}(1) \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \Phi_{11}(p) & \Phi_{12}(p) \\ \Phi_{21}(p) & \Phi_{22}(p) \end{pmatrix} \begin{pmatrix} X_{1,t-p} \\ X_{2,t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$$

에서 그레인저 인과의 개념을 확장해 **블럭 외생성**을 고려하자. X_1이 X_2에 외생적이라는 것은, X_1의 미래값이 X_2의 과거나 현재값에 의해 영향을 받지 않는다는 것을 의미한다. 즉 X_1이 모형에서 외생변수로 작용함을 의미한다. 그렇다면 앞에서와 같이, X_2가 X_1을 그레인저 인과하지 않는다는 것은

$$H_0 : \Phi_{12}(1) = \Phi_{12}(2) = \cdots = \Phi_{12}(p) = O_{n_1 \times n_2}$$

를 검정하는 것과 동일하다. $\Phi_{12}(k)$는 $n_1 \times n_2$행렬이다. 따라서 여기에서의 왈드 검정은

$$F = \frac{(|\hat{\Omega}_{11}(0)| - |\hat{\Omega}_{11}(1)|)/pn_1n_2}{|\hat{\Omega}_{11}(1)|/(T-np-n_1)} \sim F(pn_1n_2, T-np-n_1)$$

으로 주어질 것이다. 한편 대표본 하에서

$$pn_1n_2F \dot{\sim} \chi^2(pn_1n_2)$$

이므로 이를 이용할 수도 있고, 가능도비검정에서처럼

$$LR = T(\log(|\hat{\Omega}_{11}(0)|) - \log(|\hat{\Omega}_{11}(1)|)) \dot{\sim} \chi^2(pn_1n_2)$$

임을 이용해도 된다. 주의할 것은 이러한 그레인저 인과검정은 X_t가 정상시계열일 때에만 유효하다.

충격반응함수

경제학에서는 오차항을 백색잡음으로 보기도 하지만 혁신으로 보기도 한다. 그렇다면 각 측면에서의 충격이 각종 거시경제변수에 어떠한 반응을 불러오는지 궁금할 수 있다. 이를 위해 충격반응함수를 확인하는 경우가 많다. 먼저 X_t가 VAR(p) 모형을 따르는 정상시계열이라고 하자.

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \epsilon_t$$

그렇다면 이를 VMA(∞) 모형

$$X_t = \epsilon_t + \Psi_1 \epsilon_{t-1} + \cdots$$

으로 변환시키면 과거의 충격 ϵ_{t-s}가 현재의 거시경제변수 X_t에 미치는 영향을 확인할 수 있을 것이다. 그렇다면

$$(\Psi_s)_{ij} = \frac{\partial X_{t,i}}{\partial \epsilon_{t-s,j}}$$

으로, Ψ_s의 i행 j열 원소는 X_t의 i번째 변수 $X_{t,i}$에 s기 이전의 j번째 충격 $\epsilon_{t-s,j}$가 미치는 영향을 의미하게 된다.

그러나 각 변수에 존재하는 혁신은 여러 측면에서의 충격이 혼재해 만들어진 것일 수 있다. 예를 들어 가격변수(인플레율)을 거시경제변수 중 하나로 넣어 VAR을 수행하는 경우, 공급측 충격과 수요측 충격이 모두 인플레율에 혁신을 불러올 수 있다. 올바른 분석을 위해서는 이들 변수의 ϵ_t를 근본적인, 그리고 서로 직교하는 충격 여러 개로 분해할 필요가 있다. 대표적으로

$$\epsilon_t = Au_t$$

처럼 쓸 수 있다. 이때 u_t는 직교화된 충격들이며, A는 이들이 X_t에 미치는 구조적 효과를 묘사하는 $n \times n$ 행렬이다. 우리가 $\mathrm{Var}(\epsilon_t) = \Omega$로 썼다면, Ω는 대칭인 양정치행렬이기에 $\Omega = ADA^T$처럼 분해할 수 있다. 즉 충격의 분해는 먼저 ϵ_t의 공분산 구조 Ω를 안 다음, 직교대각화를 수행하고, A를 찾은 뒤, $\mathrm{Cov}(u_t) = D$로 직교화된 $u_t = A^T\epsilon_t$를 얻음으로써 이루어진다.

그러나 여기에서 문제는 직교대각화가 유일하지 않다는 데 있다. 따라서 많은 경우에는 모형에 기반하여 적절한 제약조건을 걺으로써 A를 특정한다. 대표적으로는 Cholesky 분해를 이용하는 방법이 있다. Cholesky 분해는 유일하므로, $PP^T = \Omega$인 하삼각행렬 P를 찾을 수 있다. 대각원소를 1로 조정하기 위하여 $P = AD^{1/2}$로 쓰면, $ADA^T = PP^T = \Omega$이며, A는 하삼각행렬이 된다. 그리고 유일하다. 이때에는 $\epsilon_t = Au_t$이므로, 충격 ϵ_t가 특정한 종속구조를 가지게 되는다. $n = 3$일 때를 예로 들면,

$$\begin{aligned}
\epsilon_{t,1} &= u_{t,1} \\
\epsilon_{t,2} &= a_{21}u_{t,1} + u_{t,2} \\
\epsilon_{t,3} &= a_{31}u_{t,1} + a_{32}u_{t,2} + u_{t,3}
\end{aligned}$$

이다. 따라서 Cholesky 분해를 이용해 identification을 하는 경우에는, 다양한 충격에 의해 즉각적으로 영향을

받는 변수를 X_t에서 최대한 아래에 두어야 한다. 반대로 특정 한 충격에만 즉각적으로 반응하고 나머지에는 반응하지 않는 경우, 이를 가장 위에 두어야 한다. 이는 보통 모델 하에서 적절한 가정을 통해 해결한다. 일반적으로 연방이 결정하는 이자율 등의 경우 다양한 충격을 고려해 결정되므로 가장 아래에, 투자수요나 가격수준같이 시장 혁신에 따라 결정되는 변수들을 가장 위에 두고는 한다.

물론 순서가 다르면 Cholesky 분해를 사용하는 대신 그에 맞는 분해를 사용해도 되지만, 귀찮기 때문에 보통 이처럼 한다. Cholesky 분해를 사용하는 식별 방법을 **recursive structure**라 부른다. 추후 구조적 VAR을 다룰 때 더욱 다양한 구조화 방법에 대해 알아볼 것이다.

다시 충격반응함수로 돌아오면, 직교화된 충격(공급측 충격, 수요측 충격 등)이 각 거시경제변수에 어떤 영향을 미칠지 확인하고 싶을 수 있다. 그렇다면

$$\frac{\partial X_{t,i}}{\partial u_{t-s,j}} = \frac{\partial X_{t,i}}{\partial A_{j\cdot}^{-1}\epsilon_{t-s}} = \sum_{k=1}^{n} \frac{\partial X_{t,i}}{\partial \epsilon_{t-s,k}} \times \frac{\partial \epsilon_{t-s,k}}{\partial A_{j\cdot}^{-1}\epsilon_{t-s}} = \sum_{k=1}^{n} (\Psi_s)_{ik} \times A_{kj} = (\Psi_s A)_{ij}$$

를 얻는다. 더 나아가 표준오차를 1로 하는 충격(표준화된 충격) $v_t = D^{-1/2}u_t$에 대하여 확인하면,

$$\frac{\partial X_{t,i}}{\partial v_{t-s,j}} = (\Psi_s)_{i\cdot} A_{\cdot j}\sqrt{D_{ii}} = (\Psi_s AD)_{ij}$$

를 얻게 된다. 즉 s시점 과거에서 일어난 1-SD인 j번째 충격 $v_{t-s,j}$가 X_t의 i번째 변수 $X_{t,i}$에 미치는 영향은 위처럼 쓸 수 있는 것이다. 이를 $s = 1, 2, \cdots$,에 대해 수행한다면 우리가 흔히 IRF라 부르는 impulse-response function(curve)를 얻게 되는 것이다.

국소충격반응

충격반응함수는 전체 피리오드에 VAR을 적합한 뒤 일반적인 충격(표준화된 충격)에 대한 반응함수를 분석한다. 이는 일반적이고 평균적인 이야기이다. 그러나 우리는 이따금 특정 충격이 미친 영향 자체가 궁금할 수 있다. 곡물의 공급측 충격이 가격수준에 미치는 영향 자체를 분석하는 것은 충격반응함수로 가능하지만, 러-우 전쟁으로 인한 공급측 충격이 가격에 어떠한 영향을 미쳤는가?에 대한 답은 할 수 없다. 이때 필요한 것이 국소충격반응이다.

먼저 VAR을 적합하면, $\hat{u}_t = \hat{A}^{-1}\hat{\epsilon}_t$를 얻을 수 있을 것이다. 이때 $\hat{\epsilon}_t$는 잔차로,

$$\hat{\epsilon}_t = X_t - \hat{\Phi}_1 X_{t-1} - \cdots - \hat{\Phi}_p X_{t-p}$$

로써 얻는다. 그렇다면 VMA(∞) 표현을 수행하면

$$X_{t+s} = \hat{\epsilon}_{t+s} + \hat{\Psi}_1\hat{\epsilon}_{t+s-1} + \cdots + \hat{\Psi}_{s-1}\hat{\epsilon}_{t+1} + \hat{\Psi}_s\hat{\epsilon}_t + \hat{\Psi}_{s+1}\hat{\epsilon}_{t-1} + \cdots$$

일 것인데, 만약 이 충격이 없었더라면?

$$X_{t+s}|_{u_t=0} = \hat{\epsilon}_{t+s} + \hat{\Psi}_1\hat{\epsilon}_{t+s-1} + \cdots + \hat{\Psi}_{s-1}\hat{\epsilon}_{t+1} + 0 + \hat{\Psi}_{s+1}\hat{\epsilon}_{t-1} + \cdots$$

이었을 것이다. 따라서 u_t라는 충격의 기여는

$$X_{t+s} - X_{t+s}|_{u_t=0} = \hat{\Psi}_s\hat{\epsilon}_t = \hat{\Psi}_s\hat{A}\hat{u}_t$$

로 계산할 수 있다. 사실 이는 충격반응함수에서 1-SD를 가정함에 따라 $D^{1/2}$만을 곱했던 것과 달리, 표준편차가 $D^{1/2}$인 실제 충격 $\hat{u}_t$를 집어넣는 것이나 다름없다.

한편 이는 적당한 identification이 되었다는 가정 하에, 해당 시기에 발생한 충격 $\hat{u}_t$의 크기를 앎으로써 s기 후에 어떠한 반응이 발생하였는지를 추정한다는 점에서 특정한 사건이나 정책이 미친 영향을 평가하는 데 사용될 수 있다.

예측오차 분산분해

우리는 앞서 특정한 충격이 어떠한 변수에 미치는 영향을 보았다. 반대로 어떠한 변수의 변동에 가장 영향을 많이 미치는 충격이 무엇인지 궁금할 수도 있다. 이를 위해서는 예측오차를 구성하는 변동성분을 분해할 필요가 있다. X_{t+s}에 대한 t시점에서의 추정량은

$$\hat{X}_{t+s|t} = \mathbb{E}[X_{t+s}|I_t] = \Psi_s \epsilon_t + \Psi_{s+1}\epsilon_{t-1} + \cdots$$

으로 주어진다. 그렇다면 예측오차는

$$X_{t+s} - \hat{X}_{t+s|t} = \epsilon_{t+s} + \Psi_1 \epsilon_{t+s-1} + \cdots + \Psi_{s-1}\epsilon_{t+1}$$

으로 주어지며, 그 MSE는

$$\mathrm{MSE} = \mathrm{Var}(X_{t+s} - \hat{X}_{t+s|t}) = \Omega + \Psi_1 \Omega \Psi_1^T + \cdots + \Psi_{s-1}\Omega\Psi_{s-1}^T$$

으로 써줄 수 있다. 그런데 $\epsilon_t = Au_t$로 쓰고 $\mathrm{Cov}(u_t) = D$라고 하면,

$$\Omega = ADA^T = A_{\cdot 1} D_{11} A_{\cdot 1}^T + \cdots + A_{\cdot n} D_{nn} A_{\cdot n}^T$$

이므로,

$$\begin{aligned}
\mathrm{MSE} = & D_{11}(A_{\cdot 1}A_{\cdot 1}^T + \Psi_1(A_{\cdot 1}A_{\cdot 1}^T\Psi_1^T + \cdots + \Psi_{s-1}(A_{\cdot 1}A_{\cdot 1}^T\Psi_{s-1}) \\
& + D_{22}(A_{\cdot 2}A_{\cdot 2}^T + \Psi_1(A_{\cdot 2}A_{\cdot 2}^T\Psi_1^T + \cdots + \Psi_{s-1}(A_{\cdot 2}A_{\cdot 2}^T\Psi_{s-1}) \\
& \vdots \\
& + D_{nn}(A_{\cdot n}A_{\cdot n}^T + \Psi_1(A_{\cdot n}A_{\cdot n}^T\Psi_1^T + \cdots + \Psi_{s-1}(A_{\cdot n}A_{\cdot n}^T\Psi_{s-1})
\end{aligned}$$

처럼 쓸 수 있고, 이 중

$$D_{ii}(A_{\cdot i}A_{\cdot i}^T + \Psi_1(A_{\cdot i}A_{\cdot i}^T\Psi_1^T + \cdots + \Psi_{s-1}(A_{\cdot i}A_{\cdot i}^T\Psi_{s-1})$$

가 i번째 직교화된 충격 $u_{t,i}$가 s기 이후의 X_t를 예측하는 데 생긴 오차에 미친 기여를 뜻하게 된다. 많은 경우 한 변수에 대한 MSE를 구하고 1로 표준화한 뒤, 각 충격들이 미친 영향을 비율로써 표시하고는 한다.

국소적 예측오차 분산분해

국소충격반응에서와 마찬가지로 국소적인 분산분해를 하고플 수 있다. 앞에서와 같이

$$X_{t+s} = \hat{\epsilon}_{t+s} + \hat{\Psi}_1\hat{\epsilon}_{t+s-1} + \cdots + \hat{\Psi}_{s-1}\hat{\epsilon}_{t+1} + \hat{\Psi}_s\hat{\epsilon}_t + \hat{\Psi}_{s+1}\hat{\epsilon}_{t-1} + \cdots$$

로 쓴다면, 이번에는 t기 이후의 모든 혁신이 없다고 가정해 보자. 그러면

$$\hat{X}_{t+s|u_t,\cdots,u_{t+s}=0} = 0 + \hat{\Psi}_{s+1}\hat{\epsilon}_{t-1} + \cdots$$

이며, 예측오차는

$$X_{t+s} - \hat{X}_{t+s|u_t,\cdots,u_{t+s}=0} = \hat{\epsilon}_{t+s} + \hat{\Psi}_1\hat{\epsilon}_{t-1} + \cdots + \hat{\Psi}_s\hat{\epsilon}_t = \sum_{h=0}^{s}\hat{\Psi}_h\hat{A}\hat{u}_{t+s-h}$$

이는 t기부터 $t+s$기까지 발생한 국소적인 충격 $\hat{u}_{t+s-h}$들이 X_{t+s}의 변동에 미친 영향을 분해한다. 각각의 계수는 $\hat{\Psi}_h\hat{A}$로 주어진다. 마찬가지로 비율로써 각 충격이 특정 변수의 변동에 기여한 바를 알 수 있다.

구조적 VAR

앞의 모든 논의에서 볼 수 있듯, A의 역할이 매우 막중하다. 여기에서는 경제학적 모형에 기반하여 A를 식별하는 문제에 대해 다룰 것이다.

일반적으로 구조방정식 모형에서는 아래의 모형을 고려한다.

$$A_0X_t = G + A_1X_{t-1} + A_2X_{t-2} + \cdots + A_pX_{t-p} + Ce_t$$

여기에서 $e_t \sim (0, D)$으로, 서로 직교화된 원초적인 충격들을 의미한다. A_0는 X_t에 존재하는 변수들이 갖는 즉각적인 관계를 묘사하며, C는 직교화된 e_t가 각 거시경제변수 X_t의 충격에 어떻게 포함되는지를 묘사한다. 이제 양변에 A_0^{-1}을 곱하면

$$X_t = A_0^{-1}G + A_0^{-1}A_1X_{t-1} + \cdots + A_0^{-1}A_pX_{t-p} + A_0^{-1}Ce_t$$

를 얻고, $\mu = A_0^{-1}G$, $\Phi_i = A_0^{-1}A_i$, $\epsilon_t = A_0^{-1}Ce_t$로 정의하면 이는 우리가 그간 봐왔던 VAR 모형

$$X_t = \mu + \Phi_1X_{t-1} + \cdots + \Phi_pX_{t-p} + \epsilon_t$$

로 변환될 수 있다. 이제 문제는 VAR을 적합한 뒤 직교화된 e_t를 포함하는 SEM으로 돌아가기 위해 $\mu, \Phi_1, \cdots, \Phi_p, \epsilon_t$로부터 $G, A_0, \cdots, A_p, C, e_t$를 찾는 식별의 문제이다. 이는 A_0과 C에 적절한 조건을 가함으로써 이루어진다. 혹은 추가적으로 A_1이나 G에 제약이 주어질 수도 있다.

대표적인 제약의 예시 중 하나는 우리가 이미 보았다. 바로 Cholesky decomposition이다. Cholesky decomposition에서는 공분산행렬의 Cholesky factor P가 하삼각행렬이라는 제약을 걸었었다. 이제 여기에서는

$$P = \underbrace{A_0^{-1}C}_{A\text{ 의 역할}} \times D^{1/2} = A_0^{-1}CD^{1/2}$$

이 하삼각행렬이라는 제약이 생기게 된다. 이는 A_0와 C에 제약을 겖으로써 식별이 가능하게 된다. 단 앞서 언급하였듯 이론을 통해 변수 간에 recursive structure가 있다고 생각되는 경우에만 이러한 제약을 걸 수 있을 것이다.

더욱 일반적으로는 우리가 특정한 이론을 통해 이를 확인한 경우,

$$\begin{aligned} \Omega &= A_0^{-1}CDC^TA_0^{-T} \\ \mu &= A_0^{-1}G \\ \Phi_i &= A_0^{-1}A_i \end{aligned}$$

를 해당 제약조건 하에서 풂으로써 $A_0^{-1}C$를 특정해줄 수 있다.

예시를 보기 위해 IS-LM-PC 모형을 고려하자.

$$\begin{aligned} y &= \alpha + u^s - \sigma(i - \mathbb{E}[\Delta p_{+1}]) + u^{is} && \text{(IS equation)} \\ m - p &= \phi y - \lambda i + u^{md} && \text{(LM equation)} \\ \Delta m &= u^{ms} && \text{(money supply process)} \\ \Delta p &= \Delta p_{-1} + \beta(y - u^s) && \text{(Okun's Law + Phillips curve)} \end{aligned}$$

이때 y는 산출량의 로그값, p는 가격수준의 로그값, m은 통화량의 로그값이다. 더불어, $u^s, u^{ms}, u^{md}, u^{is}$는 각각 공급, 화폐공급, 화폐수요, IS곡선 측면에서의 충격을 의미하게 된다. 여기에 합리적 기대가 성립한다면,

$\mathbb{E}[\Delta p_{+1}] = \Delta p$이다. 따라서 SEM 모형으로 이를 쓰면,

$$\begin{pmatrix} 1 & 0 & \sigma & -\sigma \\ -\phi & 1 & \lambda & -1 \\ 0 & 1 & 0 & 0 \\ -\beta & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_t \\ m_t \\ i_t \\ p_t \end{pmatrix}$$
$$= \begin{pmatrix} \alpha \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & -\sigma \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ m_{t-1} \\ i_{t-1} \\ p_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} y_{t-2} \\ m_{t-2} \\ i_{t-2} \\ p_{t-2} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} u_t^s \\ u_t^{md} \\ u_t^{ms} \\ u_t^{is} \end{pmatrix}$$

으로 쓸 수 있을 것이다. 이는 일반적인 A_0와 C를 모두 허용하는 것이 아니라, A_0, G, A_1, A_2, C의 구조를 허용하고 $\sigma, \phi, \lambda, \beta, \alpha$만 이동할 수 있게 함으로써 5개의 변수만 결정하면 SVAR 모형이 식별될 수 있게 한다. SVAR 모형은 경제학적 모형에 맞는 VAR 모형을 적합할 수 있게 한다는 점에서 장점이 있다. 단, 제약조건의 설정이나 변수설정 등에서 단순히 통계적으로만 해석하기는 어렵고, 도메인 널리지가 충분한 상태에서 정상시계열에 대한 분석이 필요하다.

한편 SVAR의 적합은 먼저 VAR 모형의 적합을 통해 $\hat{\mu}, \hat{\Phi}_i, \hat{\Omega}$ 등을 얻은 뒤 A_0, C 등에 대한 제약조건 하에서 최대가능도법과 GMM 등을 통해 수행할 수 있다. 이는 특정한 A_0나 C 구조 하에서만 잘 밝혀져 있고 그 외에는 직접 얻어야 하는데, 복잡한 경우가 다반사이므로 문제가 나올리가 없다고 생각한다. 따라서 여기에서서는 생략한다.