13주차 2차시 : 시차분포모형(추정방법)

2. 시차분포모형 추정방법

2.시차분포모형 추정방법

- 시차분포모형을 보통최소자승법(OLS)으로 추정할 경우 잘못된 통계적 추론을 가져올 수 있음

- 왜냐하면, 시차변수를 설명변수로 포함시키면 자유도가 작아지고(자유도는 n-k이므로), 자유도가 작아지면 귀무가설을 기각할 영역이 작아지므로 통계적으로 유의한 변수도 유의하지 않은 것으로 결론을 내릴 수 있게 됨

- 따라서, 시차분포모형은 OLS로 추정하지 않고 다른 방법으로 추정하는데 대표적인 추정방법으로는 Koyck과 Almon의 방법이 있음

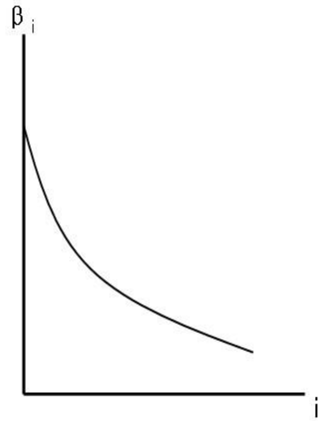
① Koyck의 추정방법

- Koyck의 추정방법은 시차의 수를 가능한 줄이려는 의도에서 출발

- 다음 식과 같이 먼 과거로 거슬러 올라 갈수록 y에 대한 시차의 영향이 점점 작아진다는 가정을 도입

$$\beta_k = \beta_0 \lambda^k$$

- 내생시차변수모형(endogenous lagged variable model) 이라고 불림



시차계수 구조

- 시차계수 구조를 시차분포모형에 대입하면 같음

$$y_t = \alpha + \beta_0 X_t + (\beta_0\lambda)X_{t-1} + \cdots + (\beta_0\lambda^{k-1})X_{t-k+1} + (\beta_0\lambda^k)X_{t-k} + u_t \quad ⓐ$$

- ⓐ식에서 시차를 하나 뒤로 하면 ⓑ과 식과 같게 됨

$$y_{t-1} = \alpha + \beta_0 X_{t-1} + (\beta_0\lambda)X_{t-2} + \cdots + (\beta_0\lambda^{k-1})X_{t-k} + u_{t-1} \quad ⓑ$$

- ⓑ의 모형에 λ를 곱하면 ⓒ식이 됨

$$\lambda y_{t-1} = \lambda\alpha + (\beta_0\lambda)X_{t-1} + (\beta_0\lambda^2)X_{t-2} + \cdots + (\beta_0\lambda^k)X_{t-k} + \lambda u_{t-1} \quad ⓒ$$

- ⓐ식에서 ⓒ식을 빼면 ⓓ식이 되고, 이를 다시 정리하면 ⓔ식이 됨

$$y_t - \lambda y_{t-1} = \alpha(1-\lambda) + \beta_0 X_t + u_t - \lambda u_{t-1} \quad ⓓ$$

$$y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda y_{t-1} + v_t \ (\text{단, } v_t = u_t - \lambda u_{t-1}) \quad ⓔ$$

- ⓔ식은 다중공선성의 문제가 없으므로 OLS로 추정하면 $\hat{\alpha}, \hat{\beta}_0, \hat{\lambda}$을 구할 수 있고, 따라서 $\hat{\beta}_k = \hat{\beta}_0 \hat{\lambda}^k$를 이용하여 $\hat{\beta}_k$를 구할 수 있음

- OLS 추정량은 불편추정량 뿐만 아니라 효율적인 추정량도 되지 못함

- 시차분포모형에서 u_t가 자기상관이 없다고 가정하더라도 ⓔ식의 v_t는 자기상관이 있어 OLS 추정량은 효율적인 추정량이 되지 못함

$$E(v_t, v_{t-1}) = \mathrm{E}[(u_t - \lambda u_{t-1})(u_{t-1} - \lambda u_{t-2})] = -\lambda E(u_{t-1}^2) = -\lambda \sigma_u^2 \neq 0$$

- 또한 v_t와 y_{t-1}가 서로 독립이 아니므로 (즉, 직교조건이 성립하지 않으므로) OLS 추정량은 불편추정량이 되지 못함

$$E(v_t, y_{t-1}) = \mathrm{E}[(u_t - \lambda u_{t-1})(y_{t-1})] = -\lambda E(u_{t-1}^2) = -\lambda \sigma_u^2 \neq 0$$

- 이에 대한 해결책으로 ⓔ식의 교란항 ν_t와 독립이고 설명변수 y_{t-1}과 상관관계가 있는 변수(이를 수단변수라 함)를 이용하는 수단변수(Instrumental Variable : IV) 추정방법이 있음

- 여기서는 X_{t-1}이 수단변수로 이용될 수 있는데 그 이유는 X_{t-1}이 이러한 조건을 만족시키는 좋은 변수이기 때문임

- 따라서 ⓔ식에서 y_{t-1}대신에 X_{t-1}을 사용하여 OLS로 추정

- 이때 직교조건은 해결되어 불편추정량은 얻을 수 있으나 X_t와 X_{t-1}사이에 다중공선성 문제가 발생할 가능성이 있으며, ν_t가 자기상관이 되어 있으므로 효율적인 추정량을 얻지 못함

② Almon의 추정방법

- Almon의 추정방법은 외생시차변수(exogenous lagged variable)의 모수를 추정한 후 시차모형의 모수를 계산하는 방법으로 다항식 시차(polynomial distributed lag : PDL) 모형이라고 함

- 이 방법은 시차계수의 크기가 시차에 따라 2차 함수 또는 3차 함수 등 다항식의 구조를 가진다고 가정

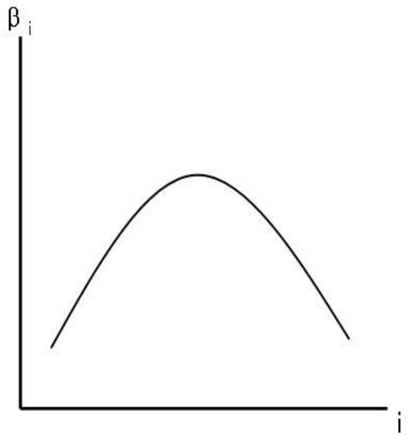
- 즉, 시차계수(β_i)가 다음 식과 같이 시차의 길이인 i의 적절한 차수의 다항식의 근사치로 계산할 수 있다고 가정

$$\beta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \cdots + \alpha_k i^r$$

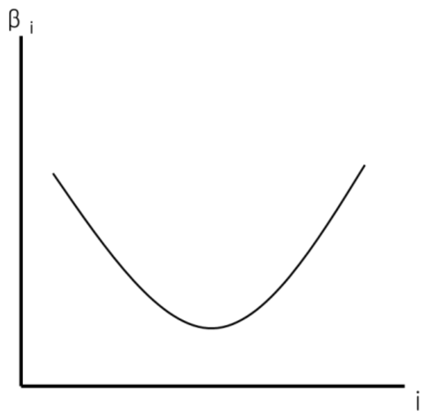
- 예를 들어, r=2인 경우 다음 식과 같이 2차 다항식이 됨

$$\beta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2$$

- 이는 시차계수(β_i)는 시차에 따라 다음 그림과 같은 구조를 갖는다고 가정한 것과 같음



시차계수 구조



시차계수 구조

- 2차 다항식을 가정하는 경우 β는 α들과 다음의 관계에 있으므로 $\alpha_0, \alpha_1, \alpha_2$을 알면 β의 값들을 알 수 있음

$$\beta_0 = \alpha_0$$
$$\beta_1 = \alpha_0 + \alpha_1 + \alpha_2$$
$$\beta_2 = \alpha_0 + 2\alpha_1 + 4\alpha_2$$
$$\vdots$$
$$\beta_k = \alpha_0 + k\alpha_1 + k^2\alpha_2$$

- 이를 시차분포모형에 대입하면 다음의 ⓕ식을 얻게 됨

$$\begin{aligned} y_t &= \alpha + \alpha_0 X_t + (\alpha_0 + \alpha_1 + \alpha_2) X_{t-1} + \cdots + (\alpha_0 + k\alpha_1 + k^2\alpha_2) X_{t-k} + u_t \\ &= \alpha + \alpha_0 (X_t + X_{t-1} + \cdots + X_{t-k}) + \alpha_1 (X_{t-1} + 2X_{t-2} + \cdots + kX_{t-k}) \\ &\quad + \alpha_2 (X_{t-1} + 2^2 X_{t-2} + \cdots + k^2 X_{t-k}) + u_t \\ &= \alpha + \alpha_0 \left(\sum_{k=0}^{K} X_{t-k}\right) + \alpha_1 \left(\sum_{k=0}^{K} kX_{t-k}\right) + \alpha_2 \left(\sum_{k=0}^{K} k^2 X_{t-k}\right) + u_t \\ &= \alpha + \alpha_0 Z_{1t} + \alpha_1 Z_{2t} + \alpha_2 Z_{3t} + u_t \quad ⓕ \end{aligned}$$

단, $Z_{1t} \equiv \sum_{k=0}^{K} X_{t-k}$, $Z_{2t} \equiv \sum_{k=0}^{K} kX_{t-k}$, $Z_{3t} \equiv \sum_{k=0}^{K} k^2 X_{t-k}$

- ⓕ식을 OLS로 추정하여 $\hat{\alpha}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2$을 구하고 β와 α의 관계를 이용하여 $\hat{\beta}_0, \hat{\beta}_1, \ldots\ldots, \hat{\beta}_k$를 계산

- Almon의 추정방법에서 다항식의 형태(r의 수) 및 시차 수(k)의 결정은 여러 가지 조합을 시행하여 회귀계수의 t값과 모형의 R^2를 높이는 방향으로 함