6주차 2차시 : 다중회귀분석(OLS 추정량의 특성)

1. OLS 추정량의 특성

(1) 선형추정량

(2) 불편추정량

(3) 최소분산추정량

1.OLS 추정량의 특성

(1)선형추정량

-회귀계수의 선형성(linearity in Y)에 대해 살펴보면 $\hat{B} = (X'X)^{-1}X'Y$에서 X는 비확률변수(즉, 확정변수)이므로 $\hat{B}$은 Y에 대해서 선형이다
단, $(X'X)^{-1}X'$는 단순회귀분석에서 있는 $\frac{x_i}{\sum x_i^2}$의 행렬식 표현이다

(참고)단순회귀모형

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i (Y_i - \bar{Y})}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2} - \frac{\bar{Y} \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} = \sum_{i=1}^{n} w_i Y_i$$

단, $w_i = \dfrac{x_i}{\sum_{i=1}^{n} x_i^2}$

(2)불편추정량

$$\hat{B} = (X'X)^{-1}X'Y$$

$$= (X'X)^{-1}X'(XB + U)$$

$$= B + (X'X)^{-1}X'U$$

$$E(\hat{B}) = B + (X'X)^{-1}X'E(U)$$

$$= B$$

(참고)단순회귀모형

$$\hat{\beta}_1 = \sum_{i=1}^{n} w_i Y_i$$
$$= \sum_{i=1}^{n} w_i\,(\beta_0 + \beta_1 X_i + u_i)$$
$$= \beta_0 \sum_{i=1}^{n} w_i \ + \beta_1 \sum_{i=1}^{n} w_i X_i + \sum_{i=1}^{n} w_i u_i$$
$$= \beta_1 + \sum_{i=1}^{n} w_i u_i \ (\because \sum_{i=1}^{n} w_i \ = 0, \sum_{i=1}^{n} w_i X_i \ = 1)$$

$$E(\hat{\beta}_1) = \beta_1 + E(\sum_{i=1}^{n} w_i u_i)$$
$$= \beta_1 + \sum_{i=1}^{n} w_i E(u_i)$$
$$= \beta_1$$

$$\sum_{i=1}^{n} w_i X_i = \sum_{i=1}^{n} \frac{x_i(x_i - \bar{X})}{\sum_{i=1}^{n} x_i^2}$$

$$= \sum_{i=1}^{n} \frac{x_i^2}{\sum_{i=1}^{n} x_i^2} - \sum_{i=1}^{n} \frac{x_i \bar{X}}{\sum_{i=1}^{n} x_i^2}$$

$$= \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2}$$

$$= 1$$

$$\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} \frac{x_i}{\sum_{i=1}^{n} x_i^2}$$

$$= \frac{\sum_{i=1}^{n} (X_i - \bar{X})}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} X_i - n\bar{X}}{\sum_{i=1}^{n} x_i^2}$$

$$= 0$$

(3)최소분산추정량

-다른 선형추정량을 다음과 같다고 하자

$$\tilde{B} = [(X'X)^{-1}X' + C]Y$$

$$= [(X'X)^{-1}X' + C](XB + U)$$

$$= (X'X)^{-1}X'XB + (X'X)^{-1}X'U + CXB + CU$$

$$= B + CXB + [(X'X)^{-1}X' + C]U$$

-$\tilde{B}$가 불편추정량이 되기 위해서는 $CX = 0$이 되어야 한다

$$E(\tilde{B}) = B + CXB + [(X'X)^{-1}X' + C]E(U)$$

$$= B + CXB$$

$$= B \ if \ CX = 0$$

-한편, $\tilde{B} - B = [(X'X)^{-1}X' + C]U$이므로 $\tilde{B}$의 분산을 계산해 보자

$$Var(\tilde{B}) = E[(\tilde{B} - B)(\tilde{B} - B)']$$

$$= E[((X'X)^{-1}X'U + CU))((U'X(X'X)^{-1} + U'C')]$$

$$= E[((X'X)^{-1}X'UU'X(X'X)^{-1}) + ((X'X)^{-1}X'UU'C') + (CUU'X(X'X)^{-1} + CUU'C']$$

$$= (X'X)^{-1}\ X'E(UU')X(X'X)^{-1} + ((X'X)^{-1}X'E(UU')C') + CE(UU')X(X'X)^{-1} + CE(UU')C'$$

$$= \sigma_u^2[((X'X)^{-1} + CC'] \qquad \because E(UU') = \sigma_u^2 I_n, X'C' = 0, CX = 0$$

−다음에 논의하겠지만 $Var(\hat{B}) = \sigma_u^2(X'X)^{-1}$이므로 $Var(\tilde{B})$는 $Var(\hat{B})$보다 CC'(non-negative definite)만큼 더 크다

−따라서 $\hat{B}$은 다른 어떤 선형불편추정량보다 분산이 작게 되는데 이러한 특성을 가진 OLS추정량을 최량선형불편추정량(Best Linear Unbiased Estimator: BLUE)이라고 한다

(참고)단순회귀모형

다른 선형추정량 $\tilde{\beta}_1$을 다음의 (3.31)식과 같다고 하자.

$$\tilde{\beta}_1 = \sum c_i Y_i \qquad (3.31)$$

여기서 c_i는 w_i와 같을 수도 있고 같지 않을 수도 있다. 즉, $c_i = w_i$이면 $\hat{\beta}_1 = \tilde{\beta}_1$이고, $c_i \neq w_i$이면 $\hat{\beta}_1 \neq \tilde{\beta}_1$이다.

$$\begin{aligned} E(\tilde{\beta}_1) &= \sum c_i E(Y_i) \\ &= \sum c_i E(\beta_0 + \beta_1 X_i + u_i) \\ &= \beta_0 \sum c_i + \beta_1 \sum c_i X_i \end{aligned}$$

따라서 $\tilde{\beta}_1$가 불편추정량이 되기 위해서는 $\sum c_i = 0, \sum c_i X_i = 1$이 되어야 한다.

한편, $\tilde{\beta}_1$의 분산을 계산해 보면 (3.32)식과 같게 된다.

$$
\begin{aligned}
Var(\tilde{\beta}_1) &= Var(\sum c_i Y_i) \\
&= \sum c_i^2 Var(Y_i) \\
&= \sigma_u^2 \sum c_i^2 \\
&= \sigma_u^2 \sum (c_i - \frac{x_i}{\sum x_i^2} + \frac{x_i}{\sum x_i^2})^2 \\
&= \sigma_u^2 \sum (c_i - \frac{x_i}{\sum x_i^2})^2 + \sigma_u^2 \frac{\sum x_i^2}{(\sum x_i^2)^2} + 2\sigma_u^2 \sum (c_i - \frac{x_i}{\sum x_i^2})(\frac{x_i}{\sum x_i^2}) \\
&= \sigma_u^2 \sum (c_i - \frac{x_i}{\sum x_i^2})^2 + \sigma_u^2 \frac{\sum x_i^2}{(\sum x_i^2)^2} + 2\sigma_u^2 \frac{\sum c_i x_i}{\sum x_i^2} - 2\sigma_u^2 \frac{\sum x_i^2}{(\sum x_i^2)^2} \\
&= \sigma_u^2 \sum (c_i - \frac{x_i}{\sum x_i^2})^2 + 2\sigma_u^2 \frac{1}{\sum x_i^2} - \sigma_u^2 \frac{1}{\sum x_i^2} \\
&= \sigma_u^2 \sum (c_i - \frac{x_i}{\sum x_i^2})^2 + \sigma_u^2 \frac{1}{\sum x_i^2} \\
&= \sigma_u^2 \sum (c_i - w_i)^2 + \sigma_u^2 \frac{1}{\sum x_i^2} \qquad (3.32)
\end{aligned}
$$

(3.32)식에서 $c_i = w$이면 $var(\tilde{\beta}_1) = var(\hat{\beta}_1)$이고, $c_i \neq w_i$이면 $var(\tilde{\beta}_1) \geq var(\hat{\beta}_1)$인데 그 이유는 $\sigma_u^2 \sum (c_i - \frac{x_i}{\sum x_i^2})^2 \geq 0$이기 때문이다. 따라서 $\hat{\beta}_1$은 다른 어떤 선형불편추정량보다 분산이 작게 되는데 이러한 특성을 가진 OLS 추정량을 최량선형불편추정량(Best Linear Unbiased Estimator: BLUE)이라고 하며 이를 요약한 것이 Gauss-Markov 정리이다.