제 9 장 이분산(Heteroskedasticity)

1. 이분산의 성격과 문제점

1) 이분산(Heteroskedasticity)이란:

→ 오차항(ε_i)의 분산이 회귀모형에 포함되는 변수에 따라 일정하지 않게 나타나는 오차항에있어서의 체계적 패턴현상: $Var(\varepsilon_i) = \sigma_i^2$

⇔ 통상적 회귀모형 추정에서 기본가정들중 하나는 동분산(homoskedasticity) 가정이다: 오차항(ε_i)의 분산의 크기는 σ^2 로 모든관측치에 대해 일정하다: $Var(\varepsilon_i) = \sigma^2$

→ 이분산현상은 독립변수값이 변화할 때 이에 대응하여 변화하는 종속변수값들의 분산이 상이하게 될 때 나타나게 된다

→ 이분산 현상은 잔차항의 분산이 설명변수와 함수관계에 있을 때 발생 할 수 있다

2) 이분산성의 유형

- 비례적 이분산성: 분산의 크기가 설명변수(X_i)값에 대해 비례적으로 변화하는 경우 → $\sigma_i^2 = \sigma^2 X_i$
- 분할된 이분산성: 설명변수의 값이 여러 개의 범주(그룹)으로 분할되는 경우,

분산의 크기가 분할된 그룹에 따라 달라지는 경우

예) 1980~2000 년 물가상승자료

→ 1980~1990 년 물가상승율의 분산: σ_1^2

→ 1991~2000 년 물가상승율의 분산: σ_2^2

3) 추정시 이분산 현상의 문제점

- 회귀모형에서 기본가정들이 모두 충족한 경우 최소승자 추정치는 BLUE 의 특성, 불편성(unbiasedness), 선형성(linearity), 일치성(consistency) 그리고 효율성(efficiency)을 모두 갖게 된다
- 기본가정중 동분산 가정이 충족되지 못하고 이분산 현상이 나타날 경우 회귀모형, $Y_i = \alpha + \beta X_{1i} + \varepsilon_i$ 을 최소자승법에 의해 회귀계수(α, β)추정시 나타나는 문제점은,

a) 이분산 현상이 발생한 경우 최소승차 추정치는 불편성과 일치성은 유지하나 추정량의 분산이 커져 최소분산을 갖는 효율성을 갖지 못해 BLUE 가 되지 못한다

b) 최소 추정치 자체는 불편성을 유지하나, 그 추정치의 분산추정량 (s_α^2, s_β^2)은 하향편의(downward bias)를 갖게된다,

→ 유의성 검정통계량인 t-통계량 ($t = \beta^\wedge / s_\beta^2$)이 하향편의를 가진 분산추정량($s_\beta^2$)

으로 인해 부당하게 커짐에 따라 귀무가설 H_0: β = 0 를 부당하게 기각시킬 가능성이 커진다

→ 회귀계수의 분산에 대해 **White 의 이분산 일치 (heteroskedasticity-consistent) 추정량**을 이용하면 이분산이 나타나는 경우에도 일반적인 해결하기위해 t-통계량을 이용하여 검정이 가능하다: $Var(\hat{\beta}) = [\sum e^2(X_i - \mathbf{X})^2]/[\sum(X_i - \mathbf{X})^2]^2$

←→ 동분산의 경우: $Var(\hat{\beta}) = \sum e^2]/\sum(X_i - \mathbf{X})]^2$

2. 이분산 여부의 검정

- 오차항에 이분산 현상이 있으면, 회귀모형의 추정과 검정에 중요한 문제가 발생하므로 모형 사용전에 이분산 존재의 여부를 점검해야 한다

1) 잔차항의 그래프를 이용하는 방법

- 모집단 회귀식의 오차항(ε_i)을 간측할 수 없으므로 이에 대한 추정치인 표본회귀 오차항의 자승값(e_i^2)을 이용하여 이분산 여부를 파악할 수 있다

→ 표본회귀 오차항의 자승값(e_i^2)을 추정값($\hat{Y}$)이나 어느 특정한 설명변수(X)와 대응시켜 그래프로 나타내면 이분산이 존재할 경우 그래프가 일정한 유형을 갖게 된다

→ 그래프

2) 함수유형을 이용하는 방법

- 오차항의 크기를 구체적인 함수형태로 표현하여 이분산 형태의 유형을 파악하는 방법

- White 의 이분산 검정법

→ 오차항의 분산과 모형에 포함된 설명변수 그리고 이변수들의 2 차항 사이에 상관관계가 나타나는지를 점검 하는 방법

→ 검정 방법: 회귀모형, $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

i) 최소자승법을 이용하여 회귀모형을 추정, 잔차항을 도출한다.

$e_i = Y_i - \alpha - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}$

ii) 잔차항의 제곱(e_i^2)을 상수항과 설명변수 그리고 설명변수들의 2 차항에 대해 회귀분석을 한다.

$e_i^2 = \gamma_1 + \gamma_2 X_{1i} + \gamma_3 X_{2i} + \gamma_4 X_{1i}^2 + \gamma_5 X_{2i}^2 + \gamma_6 X_{1i} X_{2i}^2 + \nu_i$

→ 검정 통계치, $nR^2 \sim \chi^2(5)$ (n: 관측치의 수, R^2: 추정결과 도출된 결정계수)

→ 검정통계치를 이용 귀무가설, $H_0: \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 0$ 을 점검하여 오차항에 대한 추정치인 잔차항이 설명변수드과 체계적인 상관관계가 있는지를 판단 한다

→ 귀무가설이 채택되면 이분산이 존재하지 않으며 최소자승법의 적용이 가능하다

3. 이분산의 해결방안 및 대응책

- 비례적 이분산성의 경우 : 가중최소자승법(Weighted Least Square)

→ 회귀함수를 오차항(ε_i)을 각각의 분산(σ_i)으로 나누어 준뒤($1/\sigma_i$)의 가중치를 부여) 도출된 오차자승합을 최소화 시킴으로써 도출한다

- 분활된 이분산성의 경우: 일반화된 최소자승법(GLS: generalized least squares)

→ 이분상현상이 더 이상 없는 것으로 볼 수 있도록 원래의 관측치를 함수전환시킨 다음 최소자승법을 적용시키는 것이다.