12주차 2차시 : 이분산(개념 및 유형, 검정, 해결책)

1. 개념 및 유형
2. 검정
3. 해결책

1.개념 및 유형

- X의 모든 관측치에 대해 교란항의 확률분포가 일정하지 않음

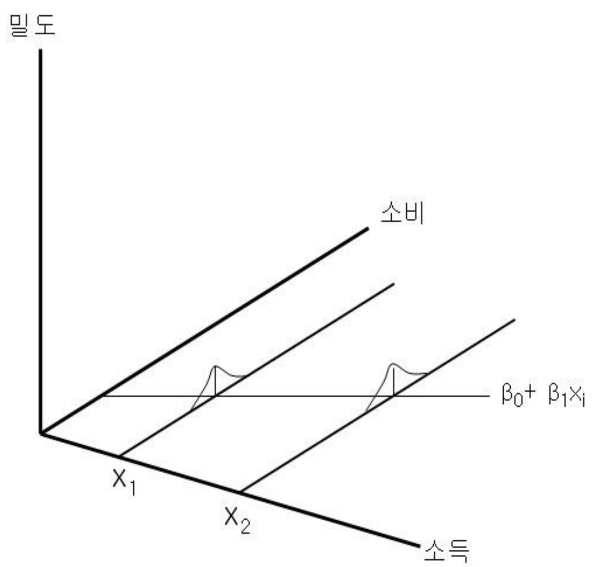
- X가 증가함에 따라 교란항의 분산이 커질 수도 있고, 작아질 수도 있음

- 일반적으로 거시자료보다 미시자료에, 시계열자료보다 횡단면자료에 이분산이 존재할 가능성이 큼

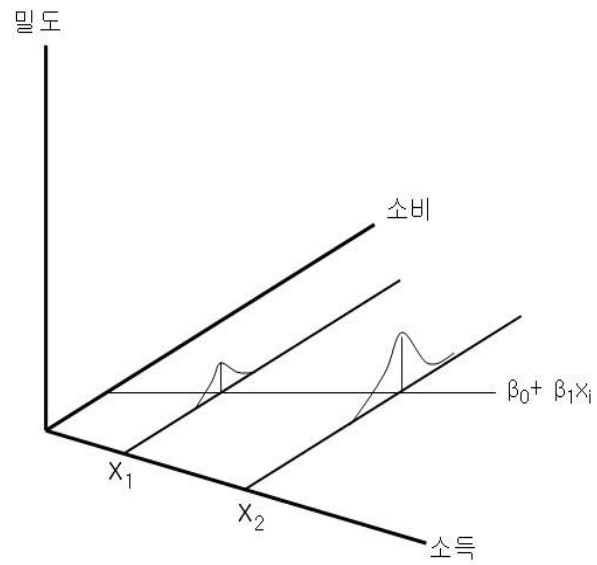
- 그 이유는 횡단면자료는 다양한 규모의 가계나 기업 등에 관한 자료가 포함되어 있기 때문임

- 그러나 시계열자료라고 해서 항상 동분산만 있는 것은 아니고, 한 경제단위의 장기에 걸친 시계열자료의 경우 이분산이 나타날 가능성이 있음

- (예 : 소득수준에 따른 소비수준의 변화)
  · 소득수준이 낮으면 소비지출에 기복이 작고 단순하여 교란항의 분포가 좁아져 분산이 작음
  · 소득수준이 높으면 소비지출에 다양해져 교란항의 분포가 퍼져서 분산이 크게 됨

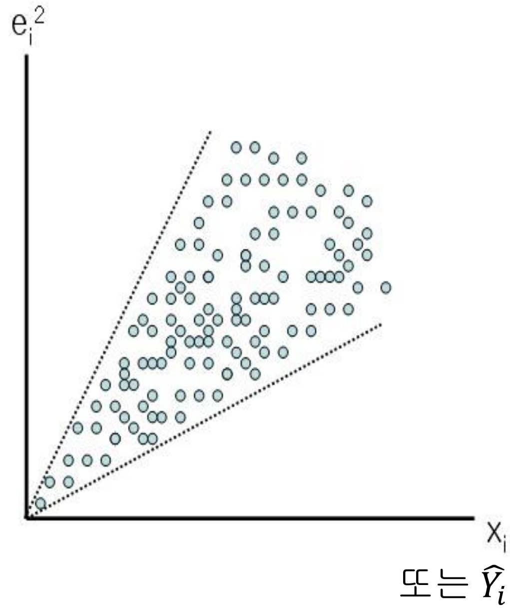


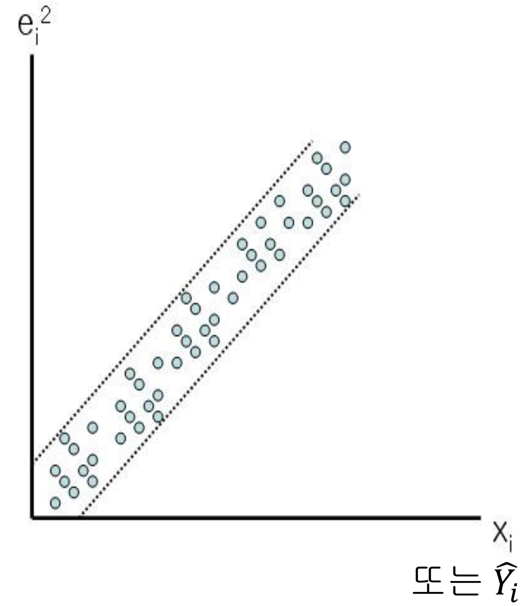
교란항의 분포(동분산)



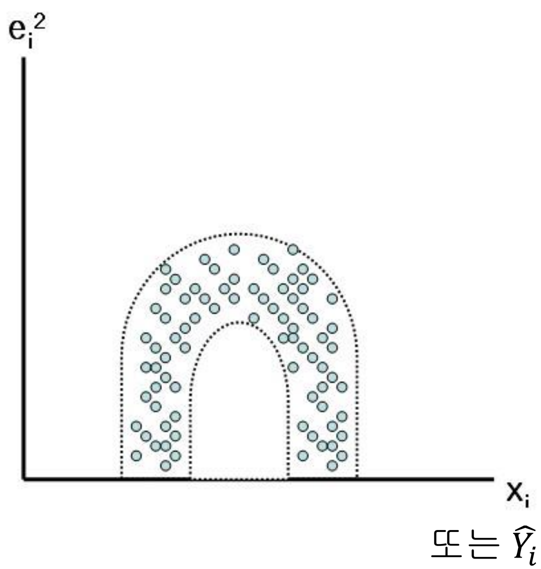
교란항의 분포(이분산)

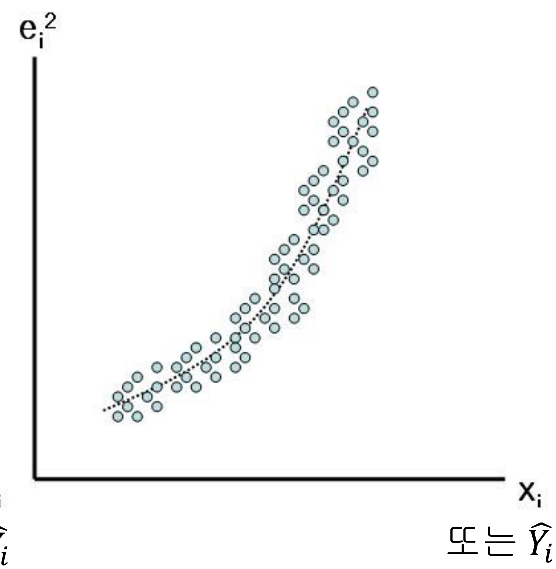
선형관계


e_i^2
X_i
또는 $\widehat{Y}_i$


e_i^2
X_i
또는 $\widehat{Y}_i$

비선형관계


e_i^2
X_i
또는 $\widehat{Y}_i$


e_i^2
X_i
또는 $\widehat{Y}_i$

2.검정

- 이분산 여부를 탐지하는 방법으로는
· 그래프 분석(residual plotting): 회귀식으로부터 도출된 잔차를 그려 보아 자기상관 여부를 판단하는 방법
· 통계적 검정: White 검정, Goldfeld-Quant 검정, Glejser 검정 (이분산이 있을 경우 일반적으로 교란항 분산의 크기는 설명변수 또는 종속변수와 관련이 있어 통계적 검정은 이러한 측면을 고려함)

①그래프분석
-잔차의 제곱(잔차의 평균이 0이므로 잔차의 제곱은 교란항의 분산이 됨)과 종속변수의 추정치 또는 독립변수를 그려 보고 이들 사이에 일정한 관계 (이분산의 유형)가 발견되면 이분산이 있는 것으로 판단

②White 검정

- 교란항의 분산과 모형에 설명변수 뿐만 아니라 이 변수들의 2차항 사이에 상관관계가 나타나는 지를 검정
- 다음 식에 대한 White 검정은 다음과 같음

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad ⓐ$$

- ⓐ식을 추정한 후 다음과 같이 잔차를 구함

$$\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}$$

- 잔차의 제곱을 이용하여 다음 식의 회귀모형을 추정하는데 이를 이를 보조회귀식(auxiliary regression)이라고 함

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + \nu_i$$

- 보조회귀식에서 검정하고자 하는 귀무가설은 $H_0 : \alpha_2 = \cdots = \alpha_6 = 0$ 즉, 이분산이 없다는 것임
- 귀무가설 하에서 다음의 LM통계량과 그 분포를 구할 수 있음

$$LM = n\,R^2 \sim \chi_q^2$$

- 단, n은 관측치의 수, R^2는 보조회귀식의 결정계수이며 χ^2−분포의 자유도는 q 즉, 제약식의 수

3.해결책

- 이분산 문제를 해결하는 방법은 이분산의 구조에 따라 달라짐
- 예를 들어, 다음의 회귀식과 교란항의 분산을 가정하면

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \ , \ E\left(u_i^2\right) = \sigma_i^2 I_n$$

① σ_i^2의 값을 알고 있는 경우

$$-\left(\frac{Y_i}{\sigma_i}\right) = \beta_1\left(\frac{1}{\sigma_i}\right) + \beta_2\left(\frac{X_{2i}}{\sigma_i}\right) + \beta_3\left(\frac{X_{3i}}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right)$$

또는 $Y_i^* = \beta_1\left(\frac{1}{\sigma_i}\right) + \beta_2 X_{2i}^* + \beta_3 X_{3i}^* + u_i^*$

- 위의 변환된 회귀모형(transformed regression model)에서 Y_i^*를 $\left(\frac{1}{\sigma_i}\right)$와 X_{2i}^*, X_{3i}^*에 대해 OLS로 추정하면 BLUE를 얻을 수 있음

- 이러한 추정방법은 일반화최소자승법(GLS)의 하나인데 이를 특별히 가중최소자승법(Weighted Least Squares : WLS)라고 함

② σ_i^2의 값을 모를 경우

- σ_i^2의 추정치를 이를 가중치로 사용
- ⓐ식 및 보조회귀식을 추정한 후 추정된 회귀계수를 이용하여 다음 식과 같이 잔차제곱에 대한 추정치(fitted value)를 구함

$$\hat{\sigma}_i^2 = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \hat{\alpha}_3 X_{3i} + \hat{\alpha}_4 X_{2i}^2 + \hat{\alpha}_5 X_{3i}^2 + \hat{\alpha}_6 X_{2i} X_{3i}$$

- $\hat{\sigma}_i^2$을 σ_i^2의 값을 알고 있는 경우 처럼 σ_i 대신에 사용하는데 이러한 추정방법을 실행 가능한 일반최소자승법(Feasible Generalized Least Squares : FGLS)이라고 함

– 만약에 회귀식과 교란항의 분산이 다음과 같다고 하면,

$$Y_i = \beta_1 + \beta_2 X_i + u_i \ , \ E(u_i^2) = \sigma^2 X_i$$

– 원래 회귀식을 다음과 같이 변환하여 이분산이 없는 모형으로 도출

$$\frac{Y_i}{\sqrt{X_i}} = \beta_1 \frac{1}{\sqrt{X_i}} + \beta_2 \sqrt{X_i} + \frac{u_i}{\sqrt{X_i}} \ , \ \mathrm{Var}\left(\frac{u_i}{\sqrt{X_i}}\right) = \frac{1}{X_i}\sigma^2 X_i = \sigma^2$$

또는 $Y_i^* = \beta_1 \left(\frac{1}{\sqrt{X_i}}\right) + \beta_2 X_{2i}^* + u_i^*$