계량경제학 10.1

제 10 강

이분산성

Heteroskedasticity

이분산성의 본질

계량경제학 10.2

- 이분산성(Heteroskedasticity) 은 오차항의 분산이 일정하지 않게 되는 오차항에 있어서의 체계적인 패턴임
- 통상적 최소제곱추정(Ordinary least squares, OLS)는 모든 관측치들이 동등한 정도로 신뢰할 만하다고 가정함
- 유효한(efficient) 추정을 위해서는 관측치에 대해 가중치를 줌으로써 동일한 오차항 분산을 갖도록 해주어야 함

이분산성의 본질

계량경제학 10.3

단순선형모형

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

0의 기대값(zero mean): $E(\varepsilon_t) = 0$

동분산성(homoskedasticity): $var(\varepsilon_t) = \sigma^2$

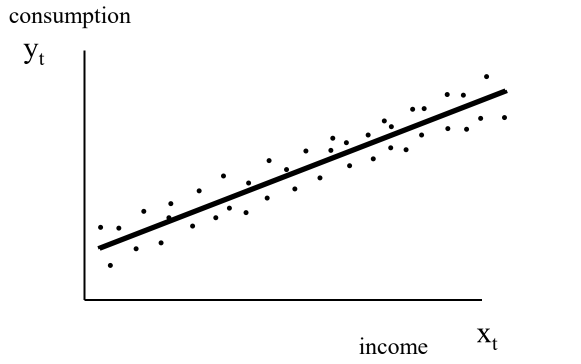
비자기상관(nonautocorrelation): $cov(\varepsilon_t, \varepsilon_s) = 0 \quad t \neq s$

heteroskedasticity: $var(\varepsilon_t) = \sigma_t^{\,2}$

이분산성의 본질

계량경제학 10.4

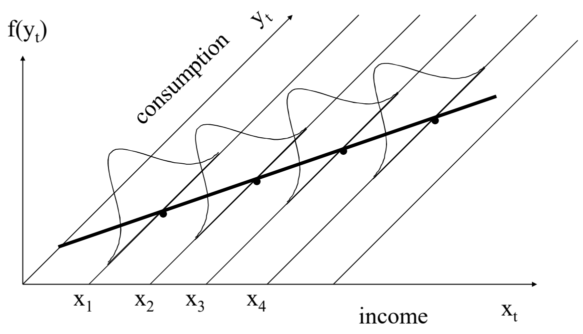
동분산성



이분산성의 본질

계량경제학 10.5

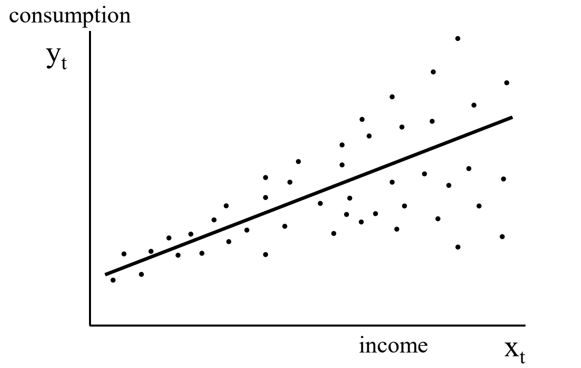
동분산성



이분산성의 본질

계량경제학 10.6

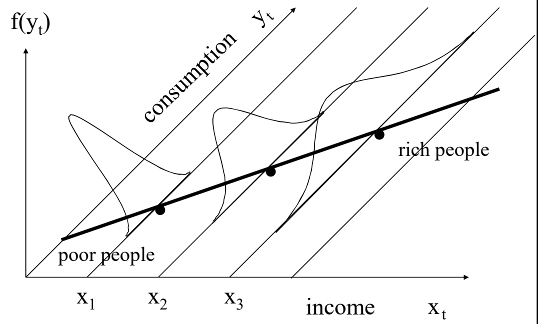
이분산성



이분산성의 본질

계량경제학 10.7

이분산성



이분산성하의 최소제곱추정량

계량경제학 10.8

- 최소제곱추정은 여전히 선형이고 불편 추정
- 최소제곱 추정은 유효 추정이 아님, 즉 BLUE가 아님
- 통상적인 공식은 최소제곱 추정량에 대한 잘못된 표준 오차를 제공함
- 이러한 잘못된 표준오차에 근거한 신뢰구간이나 가설검정은 모두 잘못된 것임

이분산성하의 최소제곱추정량

계량경제학 10.9

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

heteroskedasticity: $\quad var(\varepsilon_t) = \sigma_t^{\,2}$

최소제곱추정량의 분산에 대한 잘못된 공식

$$var(b_2) = \frac{\sigma^2}{\Sigma\,(x_t - \bar{x})^2}$$

최소제곱추정량의 분산에 대한 옳은 공식

$$var(b_2) = \frac{\Sigma\,\sigma_t^{\,2}(x_t - \bar{x})^2}{[\Sigma\,(x_t - \bar{x})^2]^2}$$

Halbert White의 표준오차

계량경제학 10.10

이분산성

화이트의 최소제곱 추정량의 분산에 대한 추정량

$$\widehat{var}(b_2) = \frac{\Sigma\,\hat{\varepsilon}_t^{\,2}(x_t - \bar{x})^2}{[\Sigma\,(x_t - \bar{x})^2]^2}$$

$\boxed{\text{대표본에서, 화이트의 표준오차는 적절함}}$

두 가지 유형의 이분산성

계량경제학 10.11

1. 비례적(Proportional) 이분산성 (continuous function of x_t, for example)
2. 분할된(Partitioned) 이분산성 (discrete categories/groups)

두 가지 유형의 이분산성

계량경제학 10.12

비례적 이분산성

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

$E(\varepsilon_t) = 0$	$var(\varepsilon_t) = \sigma_t^{\,2}$	$cov(\varepsilon_t, \varepsilon_s) = 0 \quad t \neq s$

where $\quad \sigma_t^{\,2} = \sigma^2 x_t$

The variance is assumed to be proportional to the value of x_t

두 가지 유형의 이분산성
계량경제학 10.13

비례적 이분산성

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

variance: $\mathrm{var}(e_t) = \sigma_t^2 \longrightarrow \sigma_t^2 = \sigma^2 x_t$

standard deviation: $\sigma_t = \sigma\sqrt{x_t}$

To correct for heteroskedasticity divide the model by $\sqrt{x_t}$

$$\frac{y_t}{\sqrt{x_t}} = \beta_1 \frac{1}{\sqrt{x_t}} + \beta_2 \frac{x_t}{\sqrt{x_t}} + \frac{\varepsilon_t}{\sqrt{x_t}}$$

두 가지 유형의 이분산성
계량경제학 10.14

비례적 이분산성

$$\frac{y_t}{\sqrt{x_t}} = \beta_1 \frac{1}{\sqrt{x_t}} + \beta_2 \frac{x_t}{\sqrt{x_t}} + \frac{\varepsilon_t}{\sqrt{x_t}}$$

$$y_t^* = \beta_1 x_{1t}^* + \beta_2 x_{2t}^* + \varepsilon_t^*$$

$$\mathrm{var}(\varepsilon_t^*) = \mathrm{var}\left(\frac{\varepsilon_t}{\sqrt{x_t}}\right) = \frac{1}{x_t}\mathrm{var}(\varepsilon_t) = \frac{1}{x_t}\sigma^2 x_t$$

$$\mathrm{var}(\varepsilon_t^*) = \sigma^2$$

ε_t is **heteroskedastic**, but ε_t^* is **homoskedastic**.

두 가지 유형의 이분산성
계량경제학 10.15

비례적 이분산성 - 추정방법

다음 3단계로 이루어지는 가중최소제곱 (weighted least squares) 추정 :

1. 어떤 변수가 이분산성에 비례적인가를 결정함 (x_t in previous example).
2. 원래의 모형의 모든 항들을 그 변수의 제곱근으로 나누어 줌 (divide by $\sqrt{x_t}$).
3. 새로운 종속변수와 설명변수(y_t^*, x_{t1}^* and x_{t2}^* but no intercept) 들을 가지는 변환된 모형에 대해 최소제곱 추정을 적용

두 가지 유형의 이분산성
계량경제학 10.16

분할된 이분산성

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

y_t = 1에이커 당 옥수수 수확량 $t = 1,\ldots,100$

x_t = 1에이커 당 공급된 물 (rain or other)

구 옥수수 종자의 오차항 분산: $\mathrm{var}(\varepsilon_t) = \sigma_1^2$ $t = 1,\ldots,80$

신 옥수수 종자의 오차항 분산: $\mathrm{var}(\varepsilon_t) = \sigma_2^2$ $t = 81,\ldots,100$

두 가지 유형의 이분산성
계량경제학 10.17

분할된 이분산성 - 추정방법

구 종자:	$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$	$\mathrm{var}(\varepsilon_t) = \sigma_1^2$

$$\frac{y_t}{\sigma_1} = \beta_1 \frac{1}{\sigma_1} + \beta_2 \frac{x_t}{\sigma_1} + \frac{\varepsilon_t}{\sigma_1} \qquad t = 1,\ldots,80$$

신 종자:	$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$	$\mathrm{var}(\varepsilon_t) = \sigma_2^2$

$$\frac{y_t}{\sigma_2} = \beta_1 \frac{1}{\sigma_2} + \beta_2 \frac{x_t}{\sigma_2} + \frac{\varepsilon_t}{\sigma_2} \qquad t = 81,\ldots,100$$

두 가지 유형의 이분산성
계량경제학 10.18

분할된 이분산성 - 추정방법

$$\frac{y_t}{\sigma_1} = \beta_1 \frac{1}{\sigma_1} + \beta_2 \frac{x_t}{\sigma_1} + \frac{\varepsilon_t}{\sigma_1} \qquad t = 1,\ldots,80$$

$$\frac{y_t}{\sigma_2} = \beta_1 \frac{1}{\sigma_2} + \beta_2 \frac{x_t}{\sigma_2} + \frac{\varepsilon_t}{\sigma_2} \qquad t = 81,\ldots,100$$

$$y_t^* = \beta_1 x_{1t}^* + \beta_2 x_{2t}^* + \varepsilon_t^* \qquad t = 1,\ldots,100$$

$\mathrm{var}(\varepsilon_t^*) = 1$: Homoskedasticity

However, σ_1 and σ_2 are unknown. Need to estimate them

두 가지 유형의 이분산성

그룹별 이분산성 - 추정방법

각 그룹의 자료에 대해 최소제곱추정을 적용.

> $\hat{\sigma}_1^{\,2}$ provides estimator of $\sigma_1^{\,2}$ using the 80 observations.

> $\hat{\sigma}_2^{\,2}$ provides estimator of $\sigma_2^{\,2}$ using the 20 observations.

> 일반화된 최소제곱추정(Generalized Least Square): 변수들을 표준적 가정에 부합되게 변환하고, 변화된 변수들에 대해 최소제곱 추정을 적용하여 모수들에 대한 유효한 추정을 얻음

이분산성의 탐지

1. 잔차 도표 ⇒ Park Test
   오차항의 분산의 이분산성에 있어서 체계적이고 심각한 영향이 의심되는 설명변수를 확인하는데 유용함

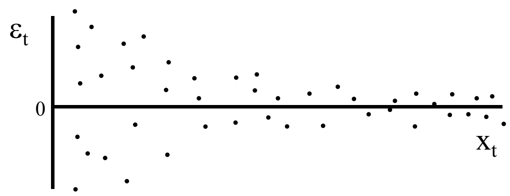
2. Goldfeld-Quandt Test
   두그룹의 관측치들 간에 그 신뢰성에 체계적이고 심각한 차이를 확임

3. White's heteroscedasticity test
   전체적인 설명변수들의 오차항의 분산에 대한 체계적이고 심각한 영향을 확인함

이분산성의 탐지

잔차 도표

의심되는 설명변수 하나를 크기에 따라 Sorting하고 그에 대해 잔차를 그려봄으로써 자료에 이분산적 패턴이 있는가 확인



이분산성의 탐지

Park 검정

H_0 : No heteroscedasticity exists (homoscedasticity) i.e., $Var(\varepsilon_t) = \sigma^2$

H_1 : Yes, heteroscedasticity exists i.e., $Var(\varepsilon_t) = \sigma_t^2$

<u>Park test procedures</u>:

1. Run OLS on regression: $Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$, **obtain** $\hat{\varepsilon}_t$
2. Take square and take log : $\ln(\hat{\varepsilon}_t^2)$
3. Run OLS on regression: $\ln(\hat{\varepsilon}_t^2) = \alpha_1 + \alpha_2 X_t + \upsilon_t$

   Suspected variable that causes heteroscedasticity

4. Use **t-test** to test $H_0 : \alpha_2 = 0$ (Homoscedasticity)

   $H_1 : \alpha_2 \neq 0$

이분산성의 탐지

Park 검정- 일반화

이분산성의 구조는 훨씬 복잡할 수 있음:

$$\sigma_t^{\,2} = \sigma^2 \exp\{\alpha_1 z_{t1} + \alpha_2 z_{t2}\}$$

> z_{t1} 와 z_{t2} 는 분산이 의존하는 것으로 의심되는 임의의 관측가능한 변수들임

> Note: The function exp{·} ensures that σ_t^2 is positive.

이분산성의 탐지

Park 검정- 일반화

$$\sigma_t^{\,2} = \sigma^2 \exp\{\alpha_1 z_{t1} + \alpha_2 z_{t2}\}$$

$$\ln(\sigma_t^2) = \ln(\sigma^2) + \alpha_1 z_{t1} + \alpha_2 z_{t2}$$

$$\ln(\sigma_t^2) = \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2}$$

where $\alpha_0 = \ln(\sigma^2)$

H_o: $\alpha_1 = 0, \alpha_2 = 0$ H_1: H_0 not true	Least squares residuals, $\hat{\varepsilon}_t$ $\ln(\hat{\varepsilon}_t^2) = \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \nu_t$ **the usual F test**

이분산성의 탐지

GQ 검정

골드펠드-퀀트 검정(The Goldfeld-Quandt test)은 비례적 이분산성 혹은 분할된 이분산성의 경우 모두 이분산성을 확인하는데 이용될 수 있음

- 비례적 이분산성의 경우, 먼저 어떤 변수가 오차항 분산에 비례적인가를 확인하는 것이 필요함
- 그리고 자료를 그 변수에 대해 큰 값에서 작은 값으로 sorting함

이분산성의 탐지

GQ 검정

- 통상적으로는 가운데 r개의 관측치는 제외함 (for example, $r \approx T/6$),
- 그리고 처음 T_1 관측치들 그리고 마지막 T_2 관측치들에 대해 최소제곱 추정을 적용.
- 어느 그룹의 관측치들이 잠재적으로 큰 분산을 가지고 있는가를 알기 때문에 다음과 같은 단측 가설 검정을 하게 됨

$$H_0:\quad \sigma_1^2 = \sigma_2^2 \qquad H_1:\quad \sigma_1^2 > \sigma_2^2$$

이분산성의 탐지

GQ 검정

Goldfeld-Quandt Test Statistic

$$GQ = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F_{[T_1-K,\ T_2-K]}$$

Small values of **GQ** support H_0 while large values support H_1.

두 관측치 그룹간에 분산이 다르다는 것이 의심되지만 어느쪽 분산이 잠재적으로 큰지 모른다면 양측 가설검정이 적절할 것임

이분산성의 탐지

White 이분산성 검정(교차항 없음)- 참고

H_0 : homoscedasticity Var (ε_t) = σ^2
H_1 : heteroscedasticity Var (ε_t) = σ_t^2

Test procedures:

(1) Run OLS on regression: $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + ... + \beta_K X_{Kt} + \varepsilon_t$,
obtain the residuals, $\hat{\varepsilon}_t$

(2) Run the auxiliary regression (보조적 회귀):

$$\hat{\varepsilon}_t^2 = \delta_1 + \delta_2 X_{2t} + ... + \delta_K X_{Kt} + \delta_{K+1} X_{2t}^2 + ... + \delta_{2K-1} X_{Kt}^2 + v_t$$

H_0: $\delta_2 = \delta_3 = \ldots = \delta_{2K-1} = 0$

(3) Compute **W** = **T·R²** $\underset{asy}{\sim} \chi^2_{df}$ (F검정도 이용가능)

(4) Compare the **W** and χ^2_{df}
(where the **df** is # of slope coefs in (2) = 2K-2)
if **W** > χ^2_{df} ==> reject the Ho

이분산성의 탐지

White 이분산성 검정(교차항 있음)-참고

H_0 : homoscedasticity Var (ε_t) = σ^2
H_1 : heteroscedasticity Var (ε_t) = σ_t^2

Test procedures:

(1) Run OLS on regression: $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t$,
obtain the residuals, $\hat{\varepsilon}_t$

(2) Run the auxiliary regression:
$\hat{\varepsilon}_t^2 = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X^2_{2t} + \alpha_5 X^2_{3t} + \alpha_6 X_{2t} X_{3t} + v_t$

(3) Compute **W (or LM)** = **T·R²** $\underset{asy}{\sim} \chi^2_{df}$ **(F검정도 이용가능)**

(4) Compare the **W** and χ^2_{df}
(where the **df** is # of slope coefs in (2))
if **W** > χ^2_{df} ==> reject the Ho