9.3 Homoscedasticity

34

35

Assumed model (to be checked)

$M: Y|Z \sim (X\beta, \sigma^2 I_n)$, $\varepsilon = Y - X\beta$

$E(\varepsilon|Z) (= E(\varepsilon)) = 0_n$

$var(\varepsilon|Z) (= var(\varepsilon)) = \sigma^2 I_n$

Elementwise: $Y_i = X_{i\cdot}^T \beta + \varepsilon_i$, $i = 1, \dots, n$

$E(\varepsilon_i|Z_i) = 0$ (= assumption A1)

$var(\varepsilon_i|Z_i) = \sigma^2$ (= assumption A2)

(A1) & (A2) $\Rightarrow$ $var(U|Z) = \sigma^2 M$

$M = I_n - X(X^TX)^{-1}X^T$

additionally (+normality): $var(U_i^{std}|Z) = 1$

used on residual plots

9.3.1 Tests of homoscedasticity

36

→ based on considering the following hypotheses

H_0: $var(\varepsilon_i|Z_i) = const \ (= \sigma^2)$

H_1: $var(\varepsilon_i|Z_i) =$ some (chosen) function of Z_i

- depending on some (chosen) function, different sensitivity towards violation of H_0 is reached

9.3.2 Score tests of homoscedasticity

Model under H_0 (≡ full-rank normal linear model)

$$M: \quad Y|Z \sim \mathcal{N}_n(X\beta, \sigma^2 I_n), \quad \operatorname{rank}(X) = k$$

elementwise: $Y_i = X_i^T \beta + \varepsilon_i$, $\varepsilon_i | Z_i \sim \mathcal{N}(0, \sigma^2)$ (independent)

Model under H_1 (≡ generalization of a general normal linear model)

$$M_{hetero}: \quad Y|Z \sim \mathcal{N}_n(X\beta, \sigma^2 W^{-1}), \quad \mathbb{R}^q \; \mathbb{R}^k \; \mathbb{R}^p$$

$$W = \operatorname{diag}(w_1, \dots, w_n), \quad w_i^{-1} = \tau(\lambda, \beta, Z_i), \quad i = 1, \dots, n$$

τ: a known function such that $\tau(0, \beta, z) = 1$ for all $\beta \in \mathbb{R}^k$, $z \in \mathbb{R}^p$

elementwise: $Y_i = X_i^T \beta + \varepsilon_i$,

$$\varepsilon_i | Z_i \sim \mathcal{N}(0, \sigma^2 \underbrace{\tau(\lambda, \beta, Z_i)}_{=1 \text{ if } \lambda = 0})$$

Test of homoscedasticity

$H_0: \lambda = 0$

$H_1: \lambda \neq 0$

under normality, both M and M_{hetero} are <u>full parametric models</u>

- ML inference is possible

$M \subset M_{hetero}$

$M = M_{hetero}$ with $\lambda = 0$

- Test on λ = test of homoscedasticity in a normal LM

→ classical asymptotic tests based on ML theory (Wald, score, likelihood ratio)

<u>will follow</u>: some special choices of τ function ($var(\varepsilon_i | Z_i) = \sigma^2 \tau(\lambda, \beta, Z_i)$) used in practice

Breusch-Pagan test

≡ score test of $H_0: \omega = 0$ in a model with $\mathrm{var}(\varepsilon_i | Z_i) = \sigma^2 \cdot \underbrace{\exp(\omega X_i^T \beta)}_{\tau(\omega, \beta, Z_i)}$

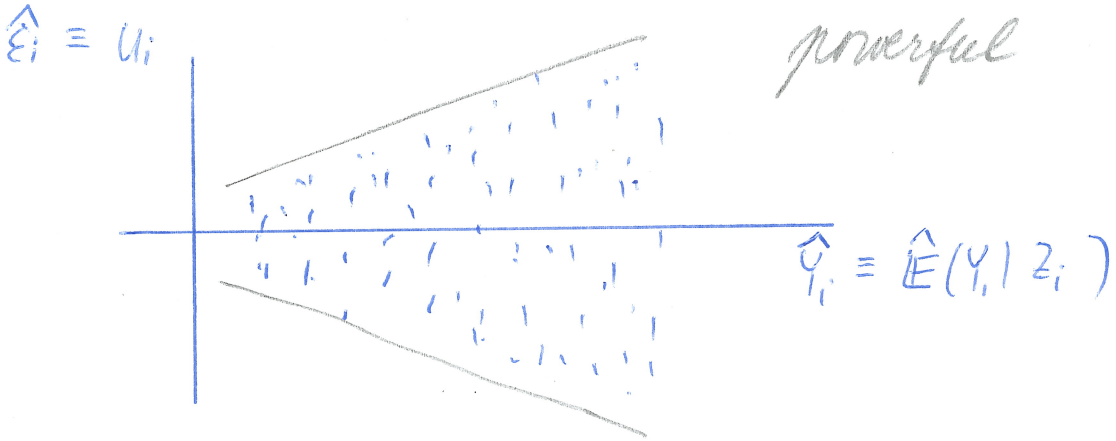
$X_i = t_X(Z_i)$ ≡ regressors used also to model $E(Y_i | Z_i)$ as $X_i^T \beta$

M_{hetero}: $\mathrm{var}(\varepsilon_i | Z_i) = \mathrm{var}(Y_i | Z_i)$

$= \sigma^2 \exp(\omega \cdot E(Y_i | Z_i))$

≡ increase / decrease of residual variability with $E(Y|Z)$

→ situation when BP test powerful



+ additional comments on slide

Linear dependence on the regressors

Hetero: $\text{var}(\varepsilon_i | Z_i) = \text{var}(Y_i | Z_i) =$

$= \sigma^2 \exp(\alpha^T W_i)$

$W_i = t_W(Z_i)$ = some function (parameterization) of covariates

Written differently:

Hetero: $\log(\text{var}(\varepsilon_i | Z_i)) = \underbrace{\log \sigma^2}_{\lambda_0} + \alpha^T W_i$

= linear model for log-residual variance

---

most commonly used:

$X_i = (X_{i0}, X_{i1}, \ldots, X_{ik-1})^T$ regressors used in the model

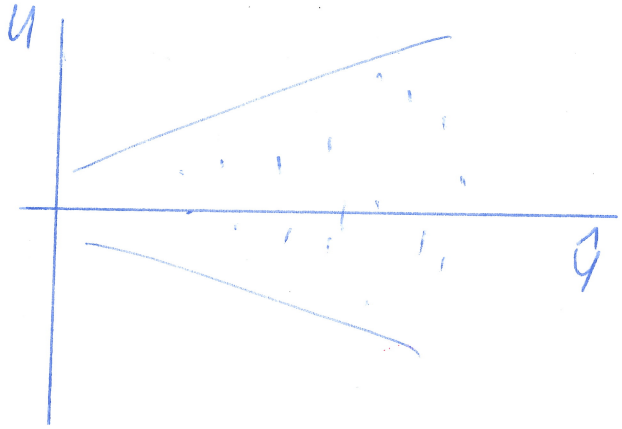
Hetero: $\text{var}(\varepsilon_i | Z_i) = \sigma^2 \exp(\lambda X_{ij})$

$\log(\text{var}(\varepsilon_i | Z_i)) = \underbrace{\log \sigma^2}_{\lambda_0} + \lambda X_{ij}$

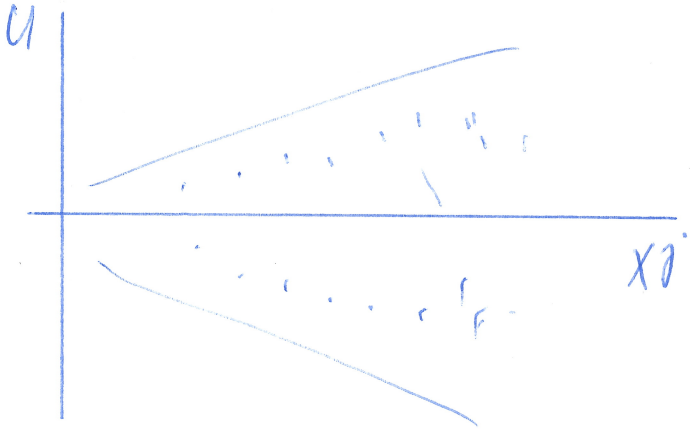
= linear dependence of $\log(\text{var}(\varepsilon_i | Z_i))$ on the j^{th} regressor

- which Metric to consider in practice ?
- which type of heteroscedasticity is hypothesized ?

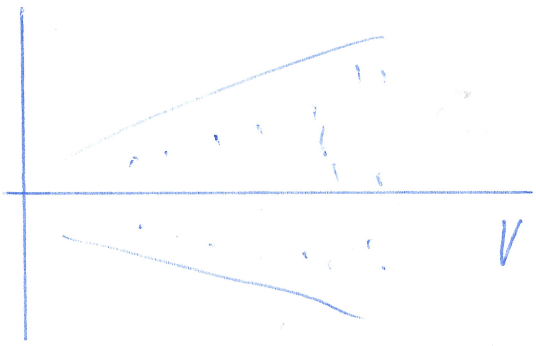
→ residual plots help



$\tau(\omega, \beta, z) = \exp(\omega X^T \beta)$



$\tau(\omega, \beta, z) = \exp(\omega x_j)$



$\tau(\omega, \beta, z) = \exp(\omega V)$

R software : see slide

9.3.3 Some other tests of homoscedasticity

- for info only
- see slides

9.4 Normality

41

42

Assumed model (to be checked)

$M: Y|Z \sim N_n(X\beta, \sigma^2 I_n)$, $\text{rank}(X_{n\times k}) = r \le k$

$\Rightarrow \varepsilon_i = Y_i - X_i^T\beta$ satisfy $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$ $i = 1,\dots,n$

(A4) normality

$\varepsilon_i|Z \overset{indep.}{\sim} N$, $\varepsilon_i \overset{iid}{\sim} N$

(A1) & (A2) & (A3) & (A4) $\Rightarrow U|Z \sim N_n(0, \sigma^2 M)$

$\Rightarrow U_i^{std}|Z \sim (0,1)$ $i = 1,\dots,n$

Problems: Even with normality

given Z: $U_1,\dots,U_n$ are normal

but neither independent

nor homoscedastic

$U_1^{std},\dots,U_n^{std}$ are homoscedastic

but not normal

slide: reminder of notation + expression of important quantities

43

normality of errors $\varepsilon_1, \dots, \varepsilon_n$

$\Rightarrow U/Z \sim N_n(0_n, \sigma^2 M)$

$\Rightarrow U_i^{std}/Z \sim (0,1)$, $i = 1, \dots, n$

---

approximate approaches to test

H_0: distribution of $\varepsilon_1, \dots, \varepsilon_n$ is normal

→ classical tests of normality
(Shapiro-Wilk, Lilliefors, Anderson-Darling, ...)
applied to (i) raw residuals $U_1, \dots, U_n$
(ii) std. residuals $U_1^{std}, \dots, U_n^{std}$

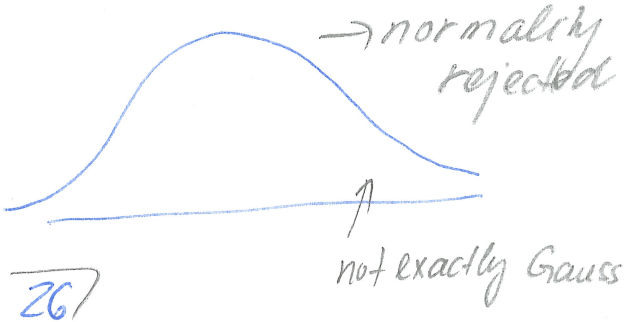
empirical studies show • significance level is maintained
• raw residuals more recommended

---

standard problem on why tests of normality quite useless

(a) small sample size = small power

(b) large sample size = large power to reject normality even if only negligibly violated



BUT, for many things, we have asymptotics and do not need normality!

9.5 Uncorrelated errors

Typical situations when uncorrelated errors cannot be taken for granted

(i) Time series

Y_1 Y_2 Y_3 Y_4 Y_5 . . .

T: t_1 t_2 t_3 t_4 t_5 . . . ← time

z_1 z_2 z_3 z_4 z_5 . . . ← additional covariates

interest to model

$E(Y_i | T_i, z_i)$

model: $Y_i = \boxed{m(T_i, z_i)} + \varepsilon_i$ ← errors

trend which may depend on covariates, possibly linear model

→ see NMSA409 Stochastic Processes 2
NMST537 Time series

(iii) Repeated measurements / longitudinal data

Data: $\underbrace{(Y_{i,1}, \dots, Y_{i,n_i})^T}_{Y_i}$, $i = 1, \dots, N$

Response vector $Y = (Y_1^T, \dots, Y_N^T)^T$

Covariates: $Z_{i,1}, \dots, Z_{i,n_i}$

Primary interest (again) to model $E(Y_{ij} | Z_{ij})$

Example: i = herd, j = cow, Y_{ij} = laboratory measurement of response to some vaccine

Z_{ij} = type of vaccine (A, B), age of cow, ...

herd 1	.....	herd N
Y_{11}		Y_{N1}
⋮		⋮
Y_{1,n_1}		Y_{N,n_N}

Possible model: $Y_{ij} = \underbrace{m(Z_{ij})}_{\text{regression function, perhaps linear model}} + \varepsilon_{ij}$ (error)

$\varepsilon_1 := (\varepsilon_{11}, \dots, \varepsilon_{1,n_1})^T$

⋮

$\varepsilon_N := (\varepsilon_{N1}, \dots, \varepsilon_{N,n_N})^T$

$\perp\!\!\!\perp$ (uncorrelated)

not necessarily uncorrelated (or even independent)

= same environment, ...

→ NMST 432 Advanced regression models

will follow: test for uncorrelated errors

but only for situations where ordering of observations has a practical meaning and may induce dependance between $\varepsilon_1, \dots, \varepsilon_n$

(e.g. time series like data)

9.5.1 Durbin-Watson test

H_0: errors are uncorrelated

H_1: are not

- test based on considering the following models

$$H_0 \equiv M: \quad Y_i = X_i^T \beta + \varepsilon_i \quad , i = 1, \dots, n$$

$$E(\varepsilon_i / X) = 0, \quad var(\varepsilon_i / X) = \sigma^2$$

$$cor(\varepsilon_i, \varepsilon_\ell / X) = 0, \quad i \neq \ell$$

$$\varepsilon_i / X \sim WN(0, \sigma^2)$$

white noise

$$H_1 \equiv M_{AR}: \quad Y_i = X_i^T \beta + \varepsilon_i \quad , i = 1, \dots, n$$

$$\varepsilon_1 = \eta_1, \quad \varepsilon_i = \rho \varepsilon_{i-1} + \eta_i \quad , i = 2, \dots, n$$

$$E(\eta_i / X) = 0, \quad var(\eta_i / X) = \sigma^2$$

$$cor(\eta_i, \eta_\ell / X) = 0, \quad i \neq \ell$$

$$\varepsilon_i / X \sim AR_1(\rho)$$

$-1 < \rho < 1$: additional unknown parameter of the model

Properties of $AR_1(\rho)$: $\operatorname{cor}(\varepsilon_{i+m}, \varepsilon_i \mid X) = \rho^m$ $\quad m \geq 1$

→ Test for uncorrelated errors $H_0: \rho = 0$

$H_1: \rho \neq 0$

(or one-sided $\rho > 0$)

Durbin-Watson test statistic

$U = (U_1, \ldots, U_n)^T$: residuals from the (linear) model M

$$DW = \frac{\sum_{i=2}^{n} (U_i - U_{i-1})^2}{\sum_{i=1}^{n} U_i^2}$$

IDEA: under H_0 (and model M): $E(U \mid X) = 0$,

also under H_1 (and model M_{AR}): $E(U \mid X) = 0$

(only $E(Y \mid X) \in \mathcal{Y}(X)$ is needed to show that $E(U \mid X) = 0$)

$U_1, \ldots, U_n$ ≡ predictions of $\varepsilon_1, \ldots, \varepsilon_n$

→ suitable estimator of $\operatorname{cov}(\varepsilon_i, \varepsilon_{i-1} \mid X)$ is

$$\hat{\sigma}_{12} = \frac{1}{\underbrace{n-1}_{\text{or } n-2}} \sum_{i=2}^{n} U_i U_{i-1}$$ (sample covariance)

→ suitable estimators of $\operatorname{var}(\varepsilon_i \mid X)$:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} U_i^2 \quad \text{or} \quad \frac{1}{n-1} \sum_{i=2}^{n} U_i^2 \quad \text{or} \quad \frac{1}{n} \sum_{i=1}^{n} U_i^2$$

$$DW = \frac{\sum_{i=2}^{n}(u_i - u_{i-1})^2}{\sum_{i=1}^{n} u_i^2} = \frac{\sum_{i=2}^{n} u_i^2 + \sum_{i=1}^{n-1} u_i^2 - 2\sum_{i=2}^{n} u_i u_{i-1}}{\sum_{i=1}^{n} u_i^2}$$

$$\approx \frac{\hat{\sigma}^2 + \hat{\sigma}^2 - 2\hat{\sigma}_{12}}{\hat{\sigma}^2} = 2\left(1 - \frac{\hat{\sigma}_{12}}{\hat{\sigma}^2}\right)$$

$\hat{\rho}$

That is, $\hat{\rho} = 1 - \frac{1}{2} DW$

$$DW = 2(1 - \hat{\rho})$$

under H_0: $DW \approx 2$

---

additional comments → see slide

bootstrap → NMST 434 Modern Statistical Methods