# L8. Modeling Problems - Other Diagnostics

## Modeling Assumption Violations:

(1) Heteroscedasticity

(2) Non-Normality Residuals

(3) False assumption of linearity between response variable $Y$ and any predictor.

## 1. Heteroscedasticity.

Recall Model Assumption: $\varepsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2)$, $i = 1, \cdots, n$

or $\vec{\varepsilon} \sim MVN(\vec{0}, \sigma^2 I_n)$

The OLSE method requires all $\varepsilon_i$s have a constant variance $\sigma^2$, s.t. the parameter estimate

$\vec{b} = (\vec{X}^T\vec{X})^{-1}\vec{X}^T\vec{y}$ is the "best" estimator in "BLUE".

- Heteroscedasticity occurs when this assumption is violated, which means the variance of ~~residuals~~ errors are non-constant, more commonly, it refers to the speed of residuals changes systematically with predictors.

(a) Problem.

(i) The OLSE estimate $\vec{b}$ is still linear and unbiased, but not the "best" anymore: we can't say $\vec{b}$ has the smallest variance among all unbiased linear estimators. There is another estimator with a smallest variance.

(ii) $se(\hat{\beta}_k)$ for all OLS output are incorrect estimate of the standard deviation of $\hat{\beta}_k$. It will result misleading t test and C.I.

(iii) Predictions of y are still unbiased, but the prediction intervals are incorrect.

Recall: $Var(\vec{b}) = E(\vec{b}-\vec{\beta})(\vec{b}-\vec{\beta})^T$

$= E((\vec{X}^T\vec{X})^{-1}\vec{X}^T\vec{\varepsilon}\,\vec{\varepsilon}^T\vec{X}(\vec{X}^T\vec{X})^{-1}]$ lecture 4

$= (\vec{X}^T\vec{X})^{-1}\vec{X}^T \cdot E(\vec{\varepsilon}\,\vec{\varepsilon})^T \cdot \vec{X}(\vec{X}^T\vec{X})^{-1}$

When applying OLS method:

assume $E(\vec{\varepsilon}\,\vec{\varepsilon})^T = \sigma^2 I_n = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}$

so $Var(\vec{b})_{ols} = \sigma^2(\vec{X}^T\vec{X})^{-1}$

and is estimated by $MSE(\vec{X}^T\vec{X})^{-1}$

which is an unbiased estimator of $Var(\vec{b})$ when the assumption of constant variance holds.

When Heteroscedasticity exist:

the real $Var(\vec{b}) = (\vec{X}^T\vec{X})^{-1}\vec{X}^T \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix} \vec{X}(\vec{X}^T\vec{X})^{-1}$

so $MSE(\vec{X}^T\vec{X})^{-1}$ is now biased estimate for $Var(\vec{b})$

In this case, if we still apply OLS method:

(i) $Var(\vec{b})_{OLS} \neq Var(\vec{b})$

Not "best" anymore

(ii) $\widehat{Var}(\vec{b})_{OLS} = MSE \cdot (\vec{X}^T \vec{X})^{-1}$ is

not unbiased anymore, so

$se(\hat{\beta}_k)_{OLS}$ is incorrect.

(iii) Let's give some result about the

prediction first

Textbook section 6.7

For a new observation with $\vec{X}_h = \begin{pmatrix} 1 \\ X_{h1} \\ \vdots \\ X_{h,p-1} \end{pmatrix}$

The fitted value $\hat{Y}_h = \vec{X}_h^T \vec{b}$

- unbiased: $E(\hat{Y}_h) = \vec{X}_h^T \vec{\beta} = E(Y_h)$

- $Var(\hat{Y}_h) = Var(\vec{X}_h^T (\vec{X}^T \vec{X})^{-1} \vec{X}^T \vec{Y})$

When assumption holds

$= \sigma^2 \vec{X}_h^T (\vec{X}^T \vec{X})^{-1} \vec{X}^T$

- Confidence Interval for the Mean Response. $\mu_h$

$$\hat{Y}_h \pm t_{\alpha/2, df=n-p} \; se(\hat{Y}_h)$$

Where $se(\hat{Y}_h) = \sqrt{MSE \; \vec{X}_h^T (\vec{X}^T \vec{X})^{-1} \vec{X}^T}$

- Prediction Interval for the new observation $Y_h$

$$\hat{Y}_h \pm t_{\alpha/2, df=n-p} \sqrt{se(\hat{Y}_h)^2 + MSE}$$

When Heteroscedasticity exists.

$$Var(\hat{Y}_h) = Var(\vec{X}_h^T (\vec{X}^T \vec{X})^{-1} \vec{X}^T \vec{y})$$

$$= \vec{X}_h^T (\vec{X}^T \vec{X})^{-1} \vec{X}^T \underbrace{Var(\vec{y})}_{\neq \sigma^2 I_n}$$

So $MSE \; \vec{X}_h^T (\vec{X}^T \vec{X})^{-1} \vec{X}^T$ is not unbiased estimate anymore.

Therefore prediction/confidence interval calculated this way would be incorrect.

(b) Detection.

(i) Residual v.s. Fitted Value Plot:

You would observe obvious change of bandwidth in the plot

(ii) **Brensch-Pagan Test**

Basic Idea is the variance of error should not change given difference prediction values.

If the assumption of constant variance is violated, then the variance would be changed with predictor values.

1) Fit your model $y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{ip-1} + \varepsilon_i$

2) Obtain the residuals $e_1, e_2, \cdots, e_n$

3) Build an auxiliary regression model:

$$e_i^2 = \gamma_0 + \gamma_1 X_{i1} + \cdots + \gamma_{p-1} X_{ip-1} + \xi_i \quad (E)$$

4) $H_0$: $\gamma_1 = \gamma_2 = \cdots = \gamma_{p-1} = 0$ v.s $H_a$: at least one $\gamma_i \neq 0$

5) Test stat : $\chi_s^2 = nR^2$

where $R^2$ is the coef. of determination of model (E)

where Ho is true: $\chi_s^2 \sim \chi^2(p)$

6) Test Result:

When Fail to reject Ho: $\chi_s^2 < \chi^2_{\alpha, df=p}$ or p-value $> \alpha$ ~~0.05~~

we conclude there's not significant heteroscedasticity.

when reject Ho: $\chi_s^2 \geq \chi^2_{\alpha, df=p}$ or p-value $\leq \alpha$.

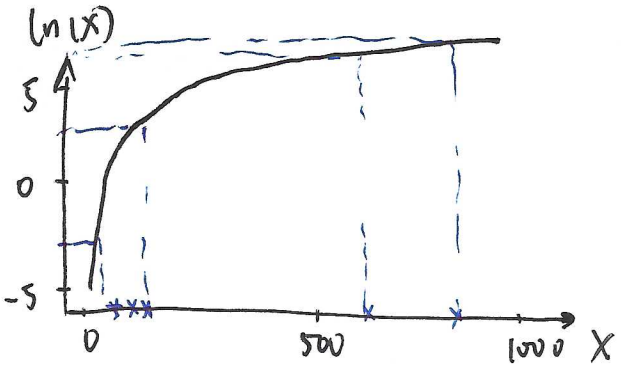
we conclude there's significant heteroscedasticity problem.

(iii) White Test * is another common one.

(c) Solution

(i) log-transformation on y

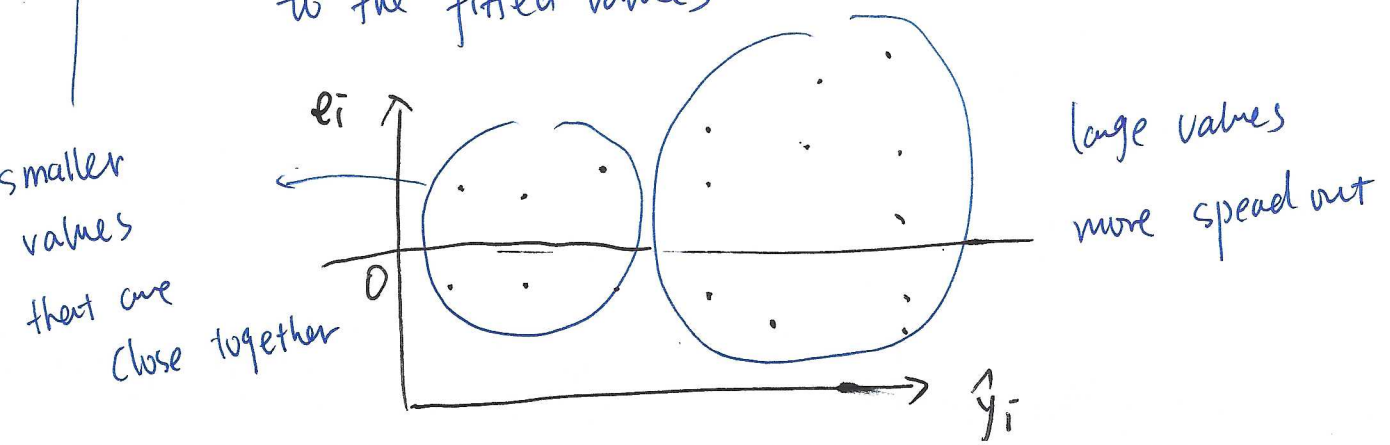
- Natural logarithm, ln, transformation is the most common transformation in MLR




From the plot, the effects of taking natural log transformation are:

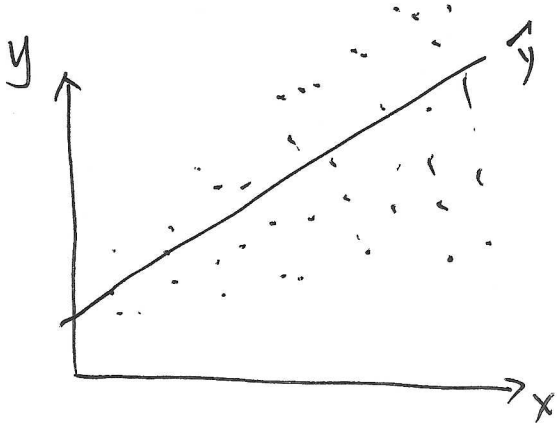
- small values that are close together are spread further out
- large values that are spead out are brought close together

Therefore, it works when the residuals changes proportionally to the fitted values



since $e_i = y_i - \hat{y}_i$

the heteroscedasticity problem lands on "y" in the data.




distribution of y around the fitted line also changes from clusters of smaller values to the spread of large values.

- In this case, a natural-log transformation on y could help "even-out" the points:

$y_{new} = \ln(y)$

smf.ols("$y_{new} \sim X_1 + \cdots + X_{p-1}$")

$\hat{y}_{new} = \ln(\hat{y}) \rightarrow \hat{y} = e^{\hat{y}_{new}}$

Since Iny only works for the case of $e_i$ is proportional with $\hat{y}_i$.

or "Funnel" shape, other solutions can be

(ii) Weighted least square regression.

Recall in OLSE: assumes $Var(\vec{\varepsilon}) = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}$

find $\vec{b}_{OLS}$ = $\arg\min SS(\vec{\beta})_{OLS} = (\vec{X}^T\vec{X})^{-1}\vec{X}^T\vec{y}$

where $SS(\vec{\beta})_{OLS} = \sum e_i^2$

and $Var(\vec{b}_{OLS}) = (\vec{X}^T\vec{X})^{-1}\vec{X}^T \cdot E(\vec{\varepsilon}\vec{\varepsilon}^T)\vec{X}(\vec{X}^T\vec{X})^{-1}$

$= Var(\vec{\varepsilon}) = \Sigma$

$= \sigma^2(\vec{X}^T\vec{X})^{-1}$

and is estimated by $MSE(\vec{X}^T\vec{X})^{-1}$

so $se(\hat{\beta}_{k})_{OLS} = [MSE(\vec{X}^T\vec{X})^{-1}]_{k+1,\,k+1}$

When heteroscedasticity exists.

Since lny only works for the case of $e_i$ is proportional with $\hat{y}_i$ or "Funnel" shape, other solutions can be

(ii) Weighted least-square regression. — count different variances in the estimation.

Recall OLSE assumes $Var(\vec{\varepsilon}) = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}$

while reality in heteroscedasticity is

$$Var(\vec{\varepsilon}) = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots \\ & & & \sigma_n^2 \end{pmatrix} = \Sigma$$

$$Var(\vec{b}) = (\vec{X}^T\vec{X})^{-1}\vec{X}^T \cdot \underbrace{E(\vec{\varepsilon}\vec{\varepsilon})^T}_{\text{AKA } Var(\vec{\varepsilon}) = \Sigma} \cdot \vec{X}(\vec{X}^T\vec{X})^{-1}$$

we need to estimate $\Sigma$ by $\hat{\Sigma} = \begin{pmatrix} e_1^2 & & 0 \\ & e_2^2 & \\ 0 & & \ddots \\ & & & e_n^2 \end{pmatrix}$

then estimate $Var(\vec{b})$ by $(\vec{X}^T\vec{X})^{-1}\vec{X}^T\hat{\Sigma}\vec{X}(\vec{X}^T\vec{X})^{-1}$

$$Var(\vec{\varepsilon}) = \begin{pmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \ddots \sigma_n^2 \end{pmatrix} = \Sigma$$

- Weighted-least-square estimates take account of the different variances in the way that

  - An observation with <u>small</u> error variation should weigh <u>more</u> in the model.

  - An observation with <u>bigger</u> error variation should weigh <u>less</u> in the model

  - It includes the $\Sigma$ in the estimation in the way that

Define the Weight Matrix

$$W = \begin{pmatrix} w_1 & & \\ & w_2 & 0 \\ & 0 & \ddots \; w_n \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_1^2} & & \\ & \frac{1}{\sigma_2^2} & 0 \\ & 0 & \ddots \; \frac{1}{\sigma_n^2} \end{pmatrix}$$

- Defined the weighted sum of squares:

$$SS_{WLS}(\vec{\beta}) = \sum w_i e_i^2$$

and $\vec{b}_{WLS} = \text{argmin } SS_{WLS} = (X^T \hat{W} X)^{-1} X^T \hat{W} Y$

where $\hat{W} =$ ?

Practically how to decide the weights can be tricky:

If we assume $\varepsilon_1 \sim N(0, \sigma_1^2), \cdots, \varepsilon_n \sim N(0, \sigma_n^2)$

we only have one observed sample $e_i$ from each $\varepsilon_i$.

Here we just give an simple example. It's NOT required to run WLS yourself:

When $e_i$ v.s. $X_K$ has a funnel shape,

you can run model = "$e_i^2 \sim X_K$"

then use fitted values of the model as the estimated variance of $Var(\varepsilon_i)$

then $w_i = 1/\hat{e}_i^2$

(iii) Generalized least squares: again deal with the estimate of $\Sigma$ in a broader choices of cases.

logistic model is a case of GLM.

(iv) Use more Robust Regression: which is not sensitive to assumptions.

(v) Sometimes, after correcting Non-Normality, Non-linearity and model selection, heteroscedasticity can be greatly improved as well.

# Summary:

- Heteroscedasticity: The assumption of constant variance of $\varepsilon_i$ is violated.

## Problem

- In OLS result

  $se(\hat{\beta}_{1c})$ is incorrect, result in misleading t test and anova test results.

- $Var(\hat{Y}_h)$ is incorrect. result in misleading C.I. and prediction intervals.

*: The estimation of coefficients and fitted values are still unbiased, therefore reliable.

## Detection

- Residual v.s. Fitted plot

  the bandwidth changes

- Breusch-Pagan Test or White Test:

  p-value $\leq 0.05$ indicated serious heteroscedasticity problem.

## Solution:

(i) If $e_i$ v.s. $\hat{y}_i$ is Funnel shape, perform natural log transformation on y.

(ii) Perform Weighted-~~square~~ least-square regression.

* Practically hard to decide a good weight

(iii) Use Robust Regression methods in ML:

Huber, RANSAC, Theil Sen