APPLIED STATISTICS

Model Diagnostics for Linear Regression II

Dr Tao Zou

Research School of Finance, Actuarial Studies & Statistics
The Australian National University

Last Updated: Wed Aug 23 09:03:19 2017

Overview

- R-Squared and Adjusted R-Squared

- Graphical Tools for Model Diagnostics

  4. Leverage plot.

  5. Standardized (Studentized) residuals versus fitted values plot.

  6. Cook's distance plot.

- Weighted Regression

References

1. **F.L. Ramsey and D.W. Schafer** (2012)
   Chapter 8, Chapter 10 and Chapter 11 of *The Statistical Sleuth*
2. ANU STAT2008 Lecture Notes
3. **W.H. Green** (2012) *Econometric Analysis*
4. The slides are made by **R Markdown**.
   http://rmarkdown.rstudio.com

Review: Sum of Squared Errors (SSE)

The sum of squared errors (SSE) for a MLR

$$\mu\{Y|X\} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k, \text{ where } X = (X_1, \cdots, X_k),$$

is defined by

$$\text{SSE} = \sum_{i=1}^{n} \text{res}_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2.$$

In Tutorial 2, we have shown for SLR, the sample variance of the residuals is

$$s_{\text{res}}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\text{res}_i - \overline{\text{res}})^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \boxed{\frac{1}{n-1}}\text{SSE},$$

which measures **the variation in the residuals**, where

$$\overline{\text{res}} = \frac{1}{n} \sum_{i=1}^{n} \text{res}_i = 0.$$

This result is also true for MLR. ~~Thus,~~ SSE also measures **the variation in the residuals.**

Total Sum of Squares (SST)

In any dataset, there will be **variation in the values of the response variable**. This variation can be measured by the sample variance of the response values

$$\boxed{s_Y^2} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2, \text{ where } \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

We call the total sum of squares (SST) of the response

$$\boxed{\mathrm{SST} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2,}$$

which measures **the variation in the values of the response variable** too.

Sum of Squares due to Regression (SSR)

In Tutorial 2, we have shown for SLR, the sample variance of the fitted values is

$$s_{\hat{Y}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 .$$

which measures **the variation in the fitted values**. This result is also true for MLR. We call the sum of squares due to regression (SSR)

$$\mathrm{SSR} = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 ,$$

which measures **the variation in the fitted values** too.

Partitioning Variability

Intuitively,

$$\begin{aligned}\sum_{i=1}^{n}(\underset{\overset{\uparrow}{\text{SST}}}{Y_i} - \bar{Y})^2 &= \sum_{i=1}^{n}\{(Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})\}^2 \\ &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right), \text{ namely}\end{aligned}$$

$$\text{SST} = \text{SSE} + \text{SSR}.$$

One important interpretation of a regression model is that it explains **variation in the values of the response variable** (SST).

The **variation in the values of the response variable** (SST) can be split into the following two parts: **the variation in the residuals** (SSE) + **the variation in the fitted values** (SSR).

The variation in the fitted values (SSR) is explained by the regression model.

The variation in the residuals (SSE) is the unexplained variation.

R-Squared

R-squared for MLR

$$\mu\{Y|X\} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k, \text{ where } X = (X_1, \cdots, X_k),$$

is the % of the total response variation explained by the regression model

$$R^2 = \frac{\boxed{\text{SSR}}}{\boxed{\text{SST}}} = \underline{1 - \frac{\boxed{\text{SSE}}}{\text{SST}}}.$$

$$0\% \leq R^2 \leq 100\%.$$

Hence, if R^2 is close to 100%, the regression model can explain the variation in response a lot. The regression model is "good".

Change in R-Squared

For MLR,

$$\mu\{Y|X\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3, \text{ where } X = (X_1, X_2, X_3),$$

recall that in F-test, we define

$$\text{the } \textbf{reduced model} \text{ is } \mu(Y|X_2) = \beta_0 + \beta_2 X_2, \text{ and}$$

$$\text{the } \textbf{full model} \text{ is } \mu(Y|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3.$$

One can verify that

$$\mathrm{SSE}_{\mathrm{full}} \leq \mathrm{SSE}_{\mathrm{reduced}}.$$

Conclusion: If we add more explanatory variables in the model, SSE will decrease.

However,

$$\mathrm{SST} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \text{ where } \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

will not change at all since it does not depend on the regression models above.

Change in R-Squared (Con'd)

Due to the definition of R-squared

$$R^2\uparrow = 1 - \frac{\mathrm{SSE}\downarrow}{\mathrm{SST}},$$

we have the **conclusion**: if we add more explanatory variables in the model, R^2 will increase.

Hence in MLR, we can add as many as explanatory variables possible such that R^2 is close to 100%. In this case, the regression model is "good".

Recall that SSE (deviance) also measures the goodness of fit for MLR.

That R^2 close to 100% means the regression model is "good", is consistent with the statement that the smaller the SSE (deviance) is, the better fitting of a model.

But this does not necessarily mean that the model can be used for prediction very well based on the new data! Sometimes, it is called over-fitting.

In order to solve this problem, we propose the adjusted R-squared.

/L14

Adjusted R-Squared

Adjusted R-squared for MLR

$$\mu\{Y|X\} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k, \text{ where } X = (X_1, \cdots, X_k), \text{ is}$$

$$\text{Adjusted } R^2 = 1 - \frac{\mathrm{SSE}/(n-k-1)}{\mathrm{SST}/(n-1)}.$$

To investigate where we get the above idea to define the adjusted R-squared, recall that we estimate σ by

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} \mathrm{res}_i^2}{n-k-1}} = \sqrt{\frac{\mathrm{SSE}}{n-k-1}},$$

where $n-k-1$ is the number of degrees of freedom for SSE.

Furthermore, the sample variance of the response values

$$s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \frac{\mathrm{SST}}{n-1},$$

where $n-1$ is the number of degrees of freedom for SST.

Adjusted R-Squared (Con'd)

This adjustment allows for the number of regression coefficients $k+1$ in the formula.

Conclusion: If we add more explanatory variables in the model, adjusted R^2 will not necessarily increase, or in other words, may decrease.

If an additional explanatory variable is added to the model, which results in the decrease in the adjusted R-squared, then this explanatory variable may not have prediction power, and hence should not necessarily be added in the model.

Adjusted R-squared will be used as a measure in variable selection in the later lectures.

Model Checking and Refinement

Model-building efforts can be wasted if

1. violation of model assumptions;

True $\mu\{Y|X_1 \cdots X_k\} = \beta_0+\beta_1 X_1 + \cdots + \beta_k X_k$

2. omitted variables;

fitting $\mu\{Y|X_1\} = \underline{\beta_0+\beta_1 X_1}$

3. problematic observations;

are not identified.

We have talked about diagnostic tools and plots to inspect the data under study, in order to deal with 1. violation of model assumptions. See Lecture Notes 3.

Model Checking and Refinement (Con'd)

For 2. omitted variables, we can detect it by residuals versus fitted values plot (see Lecture Notes 3) and we can try the following ideas first:

(1). The explanatory variables that answer the goals of the study should be included in the model. The control variables should also be included. See Lecture Notes 4.

(2). Potential confounding variables should be included. See Lecture Notes 2.

(3). The model should potentially include squared terms, interaction terms and etc. See Lecture Notes 4 and Lecture Notes 5, respectively.

This lecture will focus on graphical tools to identify 3. problematic observations and how to deal with this problem.

4. Leverage Plot (Explanatory Variables)

Consider to use the observations

$$(X_1, Y_1),$$

$$\vdots$$

$$(X_n, Y_n),$$

to construct a SLR model. The leverage of the *i*-th observation is defined by

$$h_i = \frac{1}{n-1}\left\{\frac{X_i - \bar{X}}{s_X}\right\}^2 + \frac{1}{n} \Rightarrow h_i \geq 0 \text{ and } \frac{1}{n}\sum_{i=1}^{n} h_i = \frac{2}{n},$$

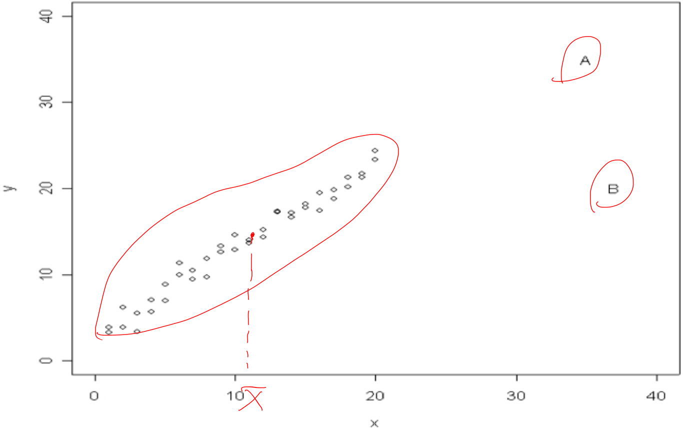
where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ and $s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

The leverage of an observation is a measure of the distance between its explanatory variable values and the average of the explanatory variable values in the entire data set.

$$\Rightarrow$$

Detect the observations with distant explanatory variable values.

4. Leverage Plot (Con'd)



Observations "A" and "B" both have high leverages.

4. Leverage Plot (Con'd)

Consider to use the observations

$$(X_{1,1}, \cdots X_{k,1}, Y_1),$$
$$\vdots$$
$$(X_{1,n}, \cdots X_{k,n}, Y_n),$$

to construct a MLR model. The leverage of the i-th observation is defined by

$$h_i = e_{n,i}^\top \mathbb{X} \left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \mathbb{X}^\top e_{n,i} \Rightarrow h_i \geq 0 \text{ and } \frac{1}{n}\sum_{i=1}^{n} h_i = \frac{k+1}{n},$$

where the $n \times (k+1)$ design matrix

$$\mathbb{X} = \begin{pmatrix} 1 & X_{1,1} & \cdots & X_{k,1} \\ 1 & X_{1,2} & \cdots & X_{k,2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1,n} & \cdots & X_{k,n} \end{pmatrix} \text{ and } e_{n,i} = \begin{array}{c} \\ \\ \\ \text{Row } i \\ \\ \\ \\ \end{array} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \text{ is an } n \times 1 \text{ vector.}$$

4. Leverage Plot (Con'd)

A "rule of thumb" cut-off for leverage is $2(k+1)/n$, twice the average of all the leverages.

If $h_i > 2(k+1)/n$, then observation i is **an observation with distant explanatory variable values**.

The leverage plot is the plot with the x-axis being the indices of the observations (i) and the y-axis being the corresponding leverages (h_i).

$i = 1, \ldots n$

5. Studentized (Standardized) Residuals versus Fitted Values Plot

Recall residual: $\mathrm{res}_i = \hat{\mathcal{E}}_i = Y_i - \hat{Y}_i$ for observation i.

One can obtain

$$\mathrm{SD}(\mathrm{res}_i) = \sigma\sqrt{1-h_i}, \text{ and}$$

$$\mathrm{SE}(\mathrm{res}_i) = \hat{\sigma}\sqrt{1-h_i}, \text{ where } \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} \mathrm{res}_i^2}{n-k-1}}.$$

A studentized residual is a residual divided by its standard error, namely

$$\mathrm{studres}_i = \frac{\mathrm{res}_i}{\mathrm{SE}(\mathrm{res}_i)}$$

Using studentized residuals allows the residuals to be viewed on the same scale.

These studentized residuals are roughly standard normally distributed ($N(0,1)$), if the observation is from the MLR model with the all the assumptions satisfied.

5. Studentized (Standardized) Residuals versus Fitted Values Plot (Con'd)

Due to the nature of the $N(0,1)$ distribution, most values of the $N(0,1)$ distribution concentrate in the middle around 0.

Hence, if $\mathrm{studres}_i$ falls into the two tails of the $N(0,1)$ distribution, namely $|\mathrm{studres}_i|$ is too large, then it is unlikely that observation i is from the MLR model with the all the assumptions satisfied.

$\mathrm{studres}_i > 1.96$ (97.5 % quantile of $N(0,1)$), or
$\mathrm{studres}_i <$- 1.96 (2.5 % quantile of $N(0,1)$).
$\Rightarrow$
$\mathrm{studres}_i$ falls into the two tails of the $N(0,1)$ distribution.
$\Rightarrow$
$|\mathrm{studres}_i|$ is too large.
$\Rightarrow$
Observation i is **an outlier**.

Sometimes we use 2 instead of 1.96 for simplicity.

6. Cook's Distance Plot

For observation i, Cook's distance is represented as

drop obs i　　　all obs

$$D_i = \sum_{j=1}^{n} \frac{\left(\hat{Y}_{j(-i)} - \hat{Y}_j\right)^2}{(k+1)\hat{\sigma}^2}, \text{ where}$$

- $\hat{Y}_j$ is the j-th fitted value in a MLR fit using all the observations $1, \cdots, n$;

→ measures the fitting after we remove obs i

- $\hat{Y}_{j(-i)}$ is the j-th fitted value in a MLR fit using observations $1, \cdots, i-1, i+1, \cdots n$;

(n-1) obs ⇒ LS est ⇒ fitted value for j

$\hat\beta_{0(-i)} \cdots \hat\beta_{k(-i)}$

$\hat{Y}_{j(-i)} = \hat\beta_{0(-i)} + \hat\beta_{1(-i)} X_{1,j} + \cdots + \hat\beta_{k(-i)} X_{2,j}$

- $k+1$ is the number of regression coefficients;

- $\hat{\sigma}^2$ is the estimated variance from the MLR fit using all the observations $1, \cdots, n$.

6. Cook's Distance Plot (Con'd)

The Cook's distance measures how much removing observations i alters the fitted model.

If observation i's Cook's distance is large, then we call it an influential observation.

Removing an influential observation may result in regression coefficients changing signs and statistical tests changing from significant to not significant.

Least squares regression analysis is not resistant to influential observations. It is possible for one or two influential observations to "drive" a least squares analysis.

One approach to dealing the problem is:

whether or not will explained later.

Examine data for influential points and potentially exclude these observations. Often these observations can provide important information.

6. Cook's Distance Plot (Con'd)

An equivalent expression for Cook's distance is:

$$D_i = \frac{1}{k+1}\,(\mathrm{studres}_i)^2\,\frac{h_i}{1-h_i}.$$

Large $|\mathrm{studres}_i|$ (outlier) or large h_i (distant explanatory variable values) may cause large Cook's distance (influential observation).

But an outlier or an observation with distant explanatory variable values is not necessarily an influential observation.

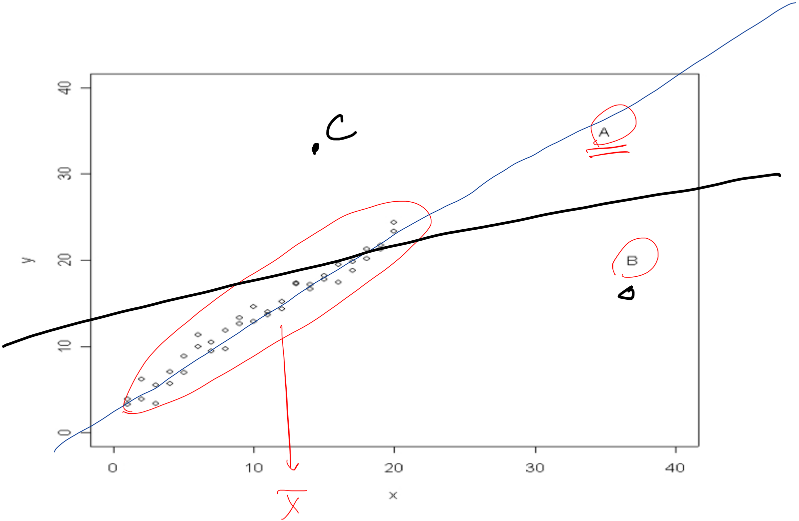
The Cook's distance plot is the plot with the x-axis being the indices of the observations (i) and the y-axis being the corresponding Cook's distances (D_i).

A rough "rule of thumb" cut-off for Cook's distance is 1. Hence, if $D_i > 1$, then the observation i is **an influential observation**.

Another option is to identify **an influential observation** that has a relatively large value of Cook's distance.

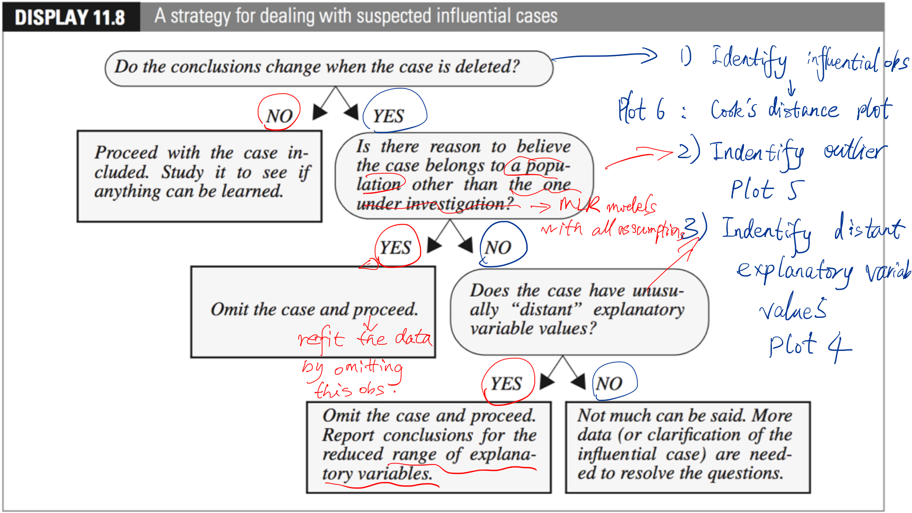
/ 415

6. Cook's Distance Plot (Con'd)



Observations "A" and "B" both have distant explanatory variable values, but only "B" is an influential observation.

Conclusion for Plots 4 – 6



DISPLAY 11.8 A strategy for dealing with suspected influential cases

Do the conclusions change when the case is deleted?

NO → *Proceed with the case included. Study it to see if anything can be learned.*

YES → *Is there reason to believe the case belongs to a population other than the one under investigation?*

YES → *Omit the case and proceed.*

NO → *Does the case have unusually "distant" explanatory variable values?*

YES → *Omit the case and proceed. Report conclusions for the reduced range of explanatory variables.*

NO → *Not much can be said. More data (or clarification of the influential case) are needed to resolve the questions.*

1) Identify influential obs

Plot 6 : Cook's distance plot

2) Indentify outlier

Plot 5

→ MLR models with all assumptions

3) Indentify distant explanatory variable values

Plot 4

refit the data by omitting this obs?

Display taken from class text: "The Statistical Sleuth".

Example: Brain Weight (Con'd)

```r
library(Sleuth3)
brain<-case0902
Y=log(brain$Brain)
X1=log(brain$Gestation)
X2=log(brain$Body)
X3=log(brain$Litter)
brain.reg = lm(Y ~ X1 + X2 + X3) #full model

CD=cooks.distance(brain.reg) #Cook's distance
Rstd<-rstudent(brain.reg) #Studentized residuals
X=cbind(X1,X2,X3)
lev=hat(X) #Leverage

par(mfrow=c(3,1))
plot(CD)
abline(h=1,col='red')
plot(brain.reg$fitted,Rstd)
abline(h=1.96,col='red')
abline(h=-1.96,col='red')
plot(lev)
abline(h=2*(3+1)/length(Y),col='red')
```

→ Plot 6

→ Plot 5

→ Plot 4

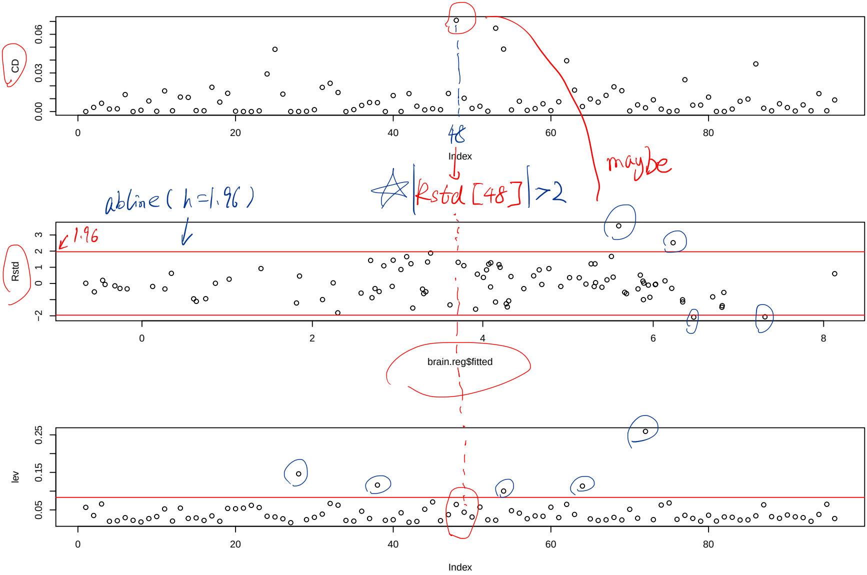
SLR

abline(fit)

abline(a , b)

→ abline(h=1.96,)

Example: Brain Weight (Con’d)



48

maybe

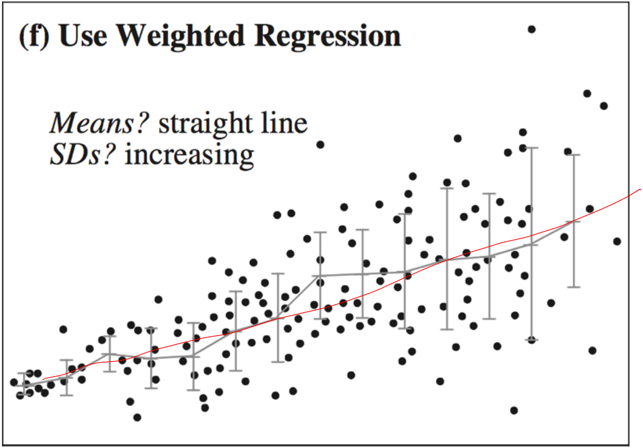
abline (h=1.96)

☆ |Rstd [48]| >2

1.96

Weighted Regression

In many situations, non-constant variances cannot be corrected by transformations or other means.



Picture taken from class text: “The Statistical Sleuth”.

Weighted Regression

Given all the observations

$$(X_{1,1}, \cdots X_{k,1}, Y_1), \qquad \text{1-st obs}$$

$$\vdots$$

$$(X_{1,n}, \cdots X_{k,n}, Y_n), \qquad \text{n-th obs}$$

a non-constant variance MLR model has the form:

$$\mu\{Y_i | X_{1,i}, \cdots, X_{k,i}\} = \beta_0 + \beta_1 X_{1,i} + \cdots + \beta_k X_{k,i}, \text{ and}$$

$$☆ \quad \sigma\{Y_i | X_{1,i}, \cdots, X_{k,i}\} = \sigma_i, \text{ for } i = 1, \cdots, n.$$

$\sigma_i = \sigma$ is a special case

The other MLR assumptions remain the same. The standard deviations σ_i's are unkown.

Generalized Least Squares (GLS)

Now if we pretend that we know those standard deviations, we can consider the following weighted least squares (also called generalized least squares, GLS) estimates of $\beta_0, \cdots, \beta_k$. Those estimates are given by the values that minimise

$$Q(b_0, \cdots, b_k) = \underbrace{\sum_{i=1}^{n} \boxed{w_i} \{Y_i - (b_0 + b_1 X_{1,i} + \cdots + b_k X_{k,i})\}^2}_{}, $$

where $\boxed{w_i = 1/\sigma_i^2.}$ This method gives more weight to the observation with less variation.

The solution of the estimates in matrix notation is

$$\begin{pmatrix} \hat{\beta}_{0,\mathrm{GLS}} \\ \hat{\beta}_{1,\mathrm{GLS}} \\ \vdots \\ \hat{\beta}_{k,\mathrm{GLS}} \end{pmatrix} = \underbrace{(\mathbb{X}^\top \mathbb{W} \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{W} \mathbb{Y}}_{},$$

minimize the above function

where $\mathbb{X}$ is the $n \times (k+1)$ design matrix, and

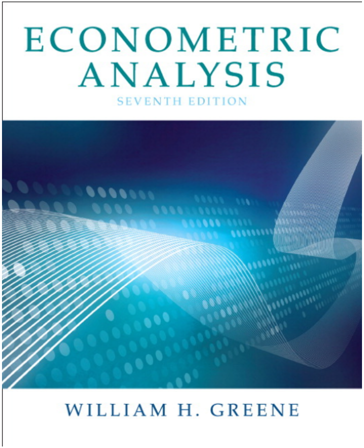
Generalized Least Squares (Con'd)

$$\mathbb{W} = \begin{pmatrix} w_1 & 0 & \cdots & 0 & 0 \\ 0 & w_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & w_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & w_n \end{pmatrix},$$

which is an $n \times n$ matrix with diagonals being w_i and off-diagonals being 0.

But in practice, we do not know where $w_i = 1/\sigma_i^2$, so how can we implement the above idea?

Feasible Generalized Least Squares (FGLS)



Homoscedasticity (Homoskedasticity): Constant Variance.
Heteroscedasticity (Heteroskedasticity): Non-contant Variances.

/16

Feasible Generalized Least Squares Steps

1. Use the (ordinary) least squares (OLS) estimation

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{Y},$$

   to fit the data first, and obtain $\mathrm{res}_i = Y_i - \hat{Y}_i$.

2. Let $w_i = 1/\mathrm{res}_i^2$ and we have the FGLS estimation

plug-in idea

$w_i = 1/\sigma_i^2$

$$\begin{pmatrix} \hat{\beta}_{0,\mathrm{GLS}} \\ \hat{\beta}_{1,\mathrm{GLS}} \\ \vdots \\ \hat{\beta}_{k,\mathrm{GLS}} \end{pmatrix} = (\mathbb{X}^\top \mathbb{W} \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{W} \mathbb{Y}.$$

Feasible Generalized Least Squares (Con'd)

If the constant variance assumption is violated, standard errors for the OLS estimates inaccurately measure uncertainty.

In this case, standard errors for the FGLS estimates are accurate.

Example: Simulation

```
rm(list=ls())
n=100
set.seed(1)
x1=rf(n,2,46)
x2=rnorm(n)
beta0=1
beta1=2
beta2=0.1
epsilon=rep(0,n)
for ( i in 1:n){
  epsilon[i]=rnorm(1,0,x1[i])
}
y=beta0+beta1*x1+beta2*x2+epsilon
plot(x1,y)
```

the model in page 29 with non-constant variances

k=2

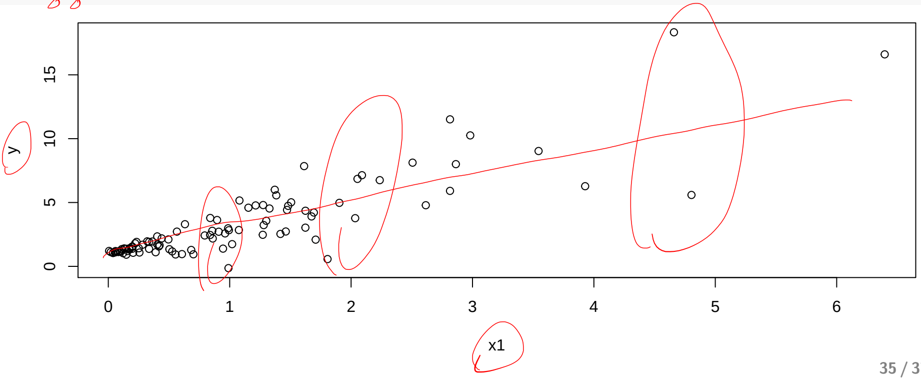
non-constant variance

simulated data → does not satisfy the constant variance

based

↓ using old method

what will happen ?



Example: Simulation (Con'd)

```
#OLS
fitOLS=lm(y~x1+x2)
summary(fitOLS)
```

old method

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9970 -0.4749  0.1001  0.5432  6.5622
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.7425     0.2130   3.486 0.000739 ***
## x1            2.3438     0.1364  17.185  < 2e-16 ***
## x2            0.2155     0.1583   1.361 0.176618
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.533 on 97 degrees of freedom
## Multiple R-squared:  0.7532, Adjusted R-squared:  0.7481
## F-statistic:   148 on 2 and 97 DF,  p-value: < 2.2e-16
```

Example: Simulation (Con'd)

```r
#FGLS
fitFGLS=lm(y~x1+x2,weights=1/fitOLS$res^2)
summary(fitFGLS)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, weights = 1/fitOLS$res^2)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0220 -0.9962  0.9975  1.0082  1.2523
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.737474   0.010684   69.02   <2e-16 ***
## x1          2.346421   0.016331  143.68   <2e-16 ***
## x2          0.218472   0.003894   56.10   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.003 on 97 degrees of freedom
## Multiple R-squared:  0.9957, Adjusted R-squared:  0.9956
## F-statistic: 1.116e+04 on 2 and 97 DF,  p-value: < 2.2e-16
```

insignificance reasons so far

1. no enough samples
2. constant variance assumption violated

✓