6304641670_Treethep

Instructions

(1) Please read the instruction carefully.
(2) Please read each question carefully and answer the questions straightforwardly. Always provide economic reasons at lease a paragraph for your analysis, or a graph when necessary, even when the question does not indicate so.
(3) Handing and submitting assignments are only available via BE Moodle.

Answering the questions and preparing answer sheets

(1) Answers are to be handwritten, in either digital or analog form, in a blank canvas or any clean paper. Make sure that your handwriting is clearly visible and readable.
(2) There is no need to rewrite the question. Just indicate the question number clearly for each of the answer, such as 1.a).
(3) Default decimal point is 4.
(4) Choose precise wordings, especially when you want to interpret the meaning of a test, confidence interval, or coefficients.
(5) When done, for the digital case, collage all the pages into a single PDF file. For those who write on sheets of paper, take photo of all pages then convert all of them into a single PDF file as well.
(6) Name your PDF file as StudentID_YourNickname, such as 640123456_Bo.

Submitting your answers

(1) Make sure your file does not exceed 10MB. This is the maximum file size for BE Moodle upload.
(2) Login to BE Moodle, head into the course, then the assignment topic.
(3) Choose your file to submit. Done. There will be timestamp for your upload date and time, so please make sure to not submit later than that.

Question 1. (12 points) Economic model of Crime.

1.a) Based on the regression results provided, write out the estimated coefficients in the form of regression equation (1.1). Interpret the estimated coefficients associated with $avgsen$. Based on Model (1.1), test whether the average sentence served from prior convictions has an impact on the number of arrests in the current year (1986). Show your work. (Use α = 0.05)

1.b) What is the overall significance of the regression from Model (1.1) and Model (1.2)? What test do you use? (Use α = 0.01)

1.c) If we are interested in testing whether "ethnic background and legal income" has an impact on the number of arrests in the current year (1986), what kind of null/alternative hypothesis would we be testing? Perform the test and discuss your finding. (Use α = 0.05)

Estimate the model (1.1) reports in the Table 1.1

$$narr86_i = \beta_1 + \beta_2 pcnv_i + \beta_3 avgsen_i + \beta_4 tottime_i + \beta_5 ptime86_i + \beta_6 qemp86_i + u_i \qquad \textbf{(1.1)}$$

Table 1.1

Source	SS	df	MS
Model	85.9532425	5	17.1906485
Residual	1924.39391	2,719	.707757967
Total	2010.34716	2,724	.738012906

Number of obs	=	2,725
F(5, 2719)	=	24.29
Prob > F	=	0.0000
R-squared	=	0.0428
Adj R-squared	=	0.0410
Root MSE	=	.84128

narr86	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
pcnv	-.1512246	.040855	Omitted for the purpose of this exam		
avgsen	-.0070487	.0124122			
tottime	.0120953	.0095768			
ptime86	-.0392585	.0089166			
qemp86	-.1030909	.0103972			
_cons	.7060607	.0331524			

Estimate the model (1.2) reports in the Table 1.2

$$narr86_i = \beta_1 + \beta_2 pcnv_i + \beta_3 avgsen_i + \beta_4 tottime_i + \beta_5 ptime86_i + \beta_6 qemp86_i + \beta_4 inc86_i + \beta_5 black_i + \beta_6 hispan_i + u_i \quad \textbf{(1.2)}$$

where

$narr86_i$	= the number of arrests in the current year (1986)
$pcnv_i$	= the proportion of prior arrests that led to a conviction
$avgsen_i$	= the average sentence served from prior convictions (in months)
$tottime_i$	= months spent in prison since age 18 prior to 1986
$ptime86_i$	= months spent in prison in 1986
$qemp86_i$	= the number of quarters that the man was legally employed in 1986
$inc86_i$	= legal income, 1986, (hundred dollars)
$black_i$	= 1 if black ethnic background
$hispan_i$	= 1 if Hispanic ethnic background

Table 1.2

Source	SS	df	MS			
				Number of obs	=	2,725
				F(8, 2716)	=	26.47
Model	145.390104	8	18.173763	Prob > F	=	0.0000
Residual	1864.95705	2,716	.686655763	R-squared	=	0.0723
				Adj R-squared	=	0.0696
Total	2010.34716	2,724	.738012906	Root MSE	=	.82865

narr86	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
pcnv	-.1332344	.0403502	Omitted for the purpose of this exam			
avgsen	-.0113177	.0122401				
tottime	.0120224	.0094352				
ptime86	-.0408417	.008812				
qemp86	-.0505398	.0144397				
inc86	-.0014887	.0003406				
black	.3265035	.0454156				
hispan	.1939144	.0397113				
_cons	.5686855	.0360461				

1.a) Based on the regression results provided, write out the estimated coefficients in the form of regression equation (1.1). Interpret the estimated coefficients associated with *avgsen*. Based on Model (1.1), test whether the average sentence served from prior convictions has an impact on the number of arrests in the current year (1986). Show your work. (Use α = 0.05)

Equation (1.1)

$$narr86_i = 0.7061 - 0.1512\, pcnv_i - 0.0071\, avgsen_i + 0.0121\, tottime_i - 0.0393\, ptime86_i - 0.1031\, qemp86 + u_i$$

$narr86_i = -0.0071\, avgsen_i$ If the average sentence served from prior convictions (in month) increases by 1 unit ($avgsen_i$ increases 1 sentence), the number of arrests in the current year (1986) will fall by 0.0071 unit ($narr86_i$ falls by 0.0071).

Use t-test to test whether β_3 is significantly different from zero

Step 1 : $H_0: \beta_3 = 0$ ($avgsen_i$ has no impact on $narr86_i$)

$H_1: \beta_3 \neq 0$ ($avgsen_i$ has an impact on $narr86_i$)

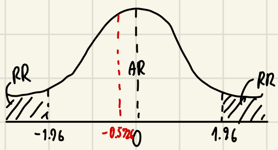
Step 2 : $\alpha = 0.05$

Step 3 : t-cal

$$t_{cal}(\beta_3) = \frac{\hat{\beta}_3 - \beta_3}{se_{\hat{\beta}_3}} = \frac{-0.0071 - 0}{0.0124} = -0.5726$$

$$t_{cri} = t_{0.05,\,2719} = 1.96$$

Step 4 & 5 : Acceptance & Rejection region (AR & RR) and conclude the result



Since our t_{cal} lies on acceptance region, it means we cannot reject the null hypothesis (H_0). As a result, it means β_3 is not significantly difference from 0. Finally, we can conclude that $avgsen_i$ has no impact on $narr86_i$ with 95% confident interval

1.b) What is the overall significance of the regression from Model (1.1) and Model (1.2)? What test do you use? (Use α = 0.01)

By using an F-test, R^2

Step 1: $H_0 = \beta_1 = \beta_2 = \dots = \beta_6 = 0$

$H_1 =$ Otherwise

Step 2 : $\alpha = 0.05$, $F_{5,\,2719} = 2.21$

Step 3 :

$$F_{cal} = \frac{R^2 / d.f._{ESS}}{1 - R^2 / d.f._{RSS}} = \frac{0.0428 / 5}{1 - 0.0428 / 2719} = \frac{8.56 \times 10^{-3}}{3.52 \times 10^{-4}} = 24.3182$$

Step 4 :

Since our F_{cal} lies into the rejection region, we can reject H_0 for model 1.1 with 95% confident interval. In other words, model 1.1 is significantly different from a

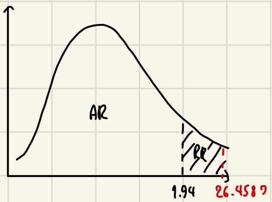
Step 1: $H_0 : \beta_1 = \beta_2 = \dots = \beta_9 = 0$

$H_1 :$ Otherwise

Step 2 : $\alpha = 0.05$, $F_{8,2716} = 1.94$

Step 3 : $F_{cal} = \dfrac{R^2 / d.f._{ESS}}{1-R^2 / d.f._{RSS}} = \dfrac{0.0723/8}{1-0.0723 / 2716} = \dfrac{9.0375 \times 10^{-3}}{3.4157 \times 10^{-4}} = 26.4587$

Step 4 :



Since our F_{cal} lies into rejection region, we can reject the null hypothesis. In other words, with 95% confident interval, we can say that Model 1.2 is significantly different from 0.

1.c) If we are interested in testing whether "ethnic background and legal income" has an impact on the number of arrests in the current year (1986), what kind of null/alternative hypothesis would we be testing? Perform the test and discuss your finding. (Use α = 0.05)

Step 1:

H_0 : Additional variables have no marginal contribution to the model

H_1 : Otherwise

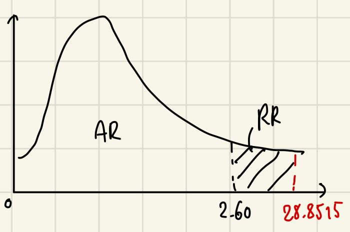
Step 2 : $\alpha = 0.05$

Step 3 :

$$F_{cal} = \frac{\dfrac{ESS_{new} - ESS_{old}}{\text{number of new regressors}}}{\dfrac{RSS_{new}}{n - k_{new}}} = \frac{\dfrac{145.3901 - 85.9532}{3}}{\dfrac{1864.9571}{2725 - 9}} = \frac{19.8123}{0.6867} = 28.8515$$

$F_{cri} = F_{3,2716} = 2.60$

Step 4 : AR and RR



Since our F_{cal} lies into the rejection region, we can reject the null hypothesis. In other words, ethnic background and legal income have a marginal contribution to the model

Question 2. (12 points) Dummy variables and interaction terms.

Using the Thailand labor force survey (LFS) in quarter 2 of 2019 and 2020, employees log of wage is modeled as follows. (Number of observations is 97,878 in total)

$$\ln wage_i = \beta_1 + \beta_2 civil_i + \beta_3 year_i + \beta_4 civil_i \cdot year_i + u_i$$

where $\ln wage_i$ = natural logarithmic scale of monthly wage

$civil_i$ = 1; civil servant and state employee
= 0; otherwise

$year_i$ = 1; year 2020
= 0; otherwise (2019)

This model is also known as Difference-in-Differences (DiD) and its intention is to capture the effect of COVID-19 since March of 2020 on different types of employment. During the pandemic, we assume that civil servant and state employee's wage is not reduced (control group) while others', namely employees in private firms or freelance, etc., is suspected to be reduced (treatment group). The estimation result is shown below with standard errors in parentheses. Answer the following questions.

$$\ln \widehat{wage}_i = \underset{(0.0035)}{9.1748} + \underset{(0.0072)}{0.587}\, civil_i - \underset{(0.005)}{0.0336}\, year_i + \underset{(0.0102)}{0.0444}\, civil_i \cdot year_i + u_i$$

2.a) Test all the parameters individually if each of them is significantly different from zero of not.

2.b) How much on average does a civil servant and state employee earns more or less than the others disregarding the year?

2.c) How much on average does the pandemic affect wage overall?

2.d) Are the control group and the treatment group better-off or worse-off during the pandemic. Discuss each group separately, show your work and explain with economic reasons according to the intention of this model.

2.a) Test all the parameters individually if each of them is significantly different from zero of not.

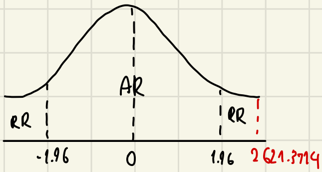
Step 1: $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

Step 2: $\alpha = 0.05$

Step 3: $t_{cal} = \frac{9.1748 - 0}{0.0035} = 2621.3714$

Step 4:



Reject H_0

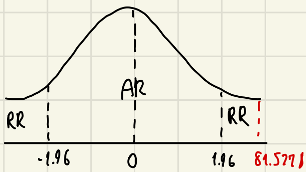
Step 1: $H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

Step 2: $\alpha = 0.05$

Step 3: $t_{cal} = \frac{0.587 - 0}{0.0072} = 81.5278$

Step 4



Reject H_0

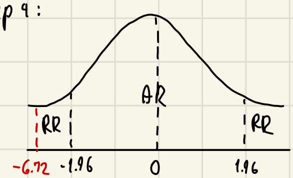
Step 1: $H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

Step 2: $\alpha = 0.05$

Step 3: $t_{cal} = \frac{-0.0336 - 0}{0.005} = -6.72$

Step 4:



Reject H_0

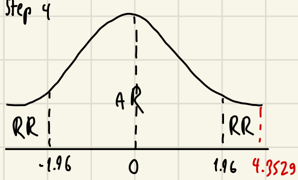
Step 1: $H_0: \beta_4 = 0$

$H_1: \beta_4 \neq 0$

Step 2: $\alpha = 0.05$

Step 3: $t_{cal} = \frac{0.0444 - 0}{0.0102} = 4.3529$

Step 4



Reject H_0

Since we can reject H_0 for all the test, With 95% confidence interval, we can say that the coefficiont is significantly difference from 0.

2.b) How much on average does a civil servant and state employee earns more or less than the others disregarding the year?

civil servant and state employee earns more then the other by $0.587 \times 100 = 58.7\%$ by approximate

civil servant and state employee earns more then the other by $(100 \cdot e^{0.587}) - 100 = 79.8585\%$ by exact

2.c) How much on average does the pandemic affect wage overall?

The pandemic effects the wage overall by $(100 \cdot e^{0.0336}) - 100 = 3.4171\%$ since it is negative coefficient it means the pandemic will affect wage on overall drop by 3.4171%.

2.d) Are the control group and the treatment group better-off or worse-off during the pandemic. Discuss each group separately, show your work and explain with economic reasons according to the intention of this model.

$$\ln \widehat{wage}_i = 9.1748 + 0.587\, civil_i - 0.0336\, year_i + 0.0444\, civil_i \cdot year_i + u_i$$
$$(0.0035) \quad (0.0072) \qquad (0.005) \qquad (0.0102)$$

For the treatment group $E(\ln wage \mid civil = 0, year = 0) = 9.1748$

$E(\ln wage \mid civil = 0, year = 1) = 9.1748 - 0.0336$

On average the treatment group are worse-off by 3.4171 % due to the impact of pandemic

For the control group $E(\ln wage \mid civil = 1, year = 0) = 9.1748 + 0.587 = 9.7618$

$E(\ln wage \mid civil = 1, year = 1) = 9.1748 + 0.587 - 0.0336 + 0.0444 = 9.7726$

If you are the control group, you earn more than the treetment group around 79.8585 %.
The impact of pandemic reduces the overall wage for both group by 3.4171 %.
Moreover, in 2020, the control group earns more than the treatment group by 4.54 %
* 4.54 % is an interaction term shows the additional effect of being control group in year 2020 at the same time
Finally, because there is the interaction term, so being control group still better-off even in the pandemic.

For the economic reason, it is possible that civil servant and state employee wage will not drop during the pandemic compares to the others. This is because usually civil servant and state employee recieves their salary from the government, so it is possible to say that they are guarantee to get the salary compares to the employee from private firm.

Question 3. (8 points) Multicollinearity.

As cheese ages, several chemical processes take place that determine the taste of the final product. The data given pertain to concentrations of various chemicals in a sample of 30 mature cheddar cheeses and subjective measure of taste for each sample.

Estimate the model (3.1) reports in the Table 3.1

$$Taste = \beta_0 + \beta_1 acetic + \beta_2 h2s + \beta_3 lactic + u \qquad \textbf{(3.1)}$$

Where $Taste$ = Measures of taste for each sample

$acetic$ = The natural logarithm of concentration of acetic

$h2s$ = The natural logarithm of concentration of hydrogen sulfide

$lactic$ = Lactic

Table 3.1

Source	SS	df	MS			
				Number of obs	=	30
				F(3, 26)	=	16.47
Model	5020.64468	3	1673.54823	Prob > F	=	0.0000
Residual	2642.24237	26	101.624706	R-squared	=	0.6552
				Adj R-squared	=	0.6154
Total	7662.88705	29	264.237485	Root MSE	=	10.081

taste	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
acetic	1.538645	3.000501	Omitted for the purpose of this exam		
h2s	3.915242	1.153106			
lactic	18.80235	8.342614			
_cons	-34.13491	15.67628			

	acetic	h2s	lactic
acetic	1.0000		
h2s	0.2700	1.0000	
lactic	0.3607	0.6448	1.0000

Variable	VIF	1/VIF
lactic	1.83	0.546648
h2s	1.72	0.582609
acetic	1.15	0.867477
Mean VIF	1.57	

3.a) Is there evidence of multicollinearity in the data? How do you know? Explain your answers in detail and state the critical value for hypothesis testing to receive full points.

3.b) What is the property of BLUE? If there is the multicollinearity problem, is the OLS estimators still retain the property of BLUE? If not, which properties are violated?

Question 4. (8 points) Heteroscedasticity.

The data on U.S. inflation rates (%) and unemployment rates (%), 1948-2006

Estimate the model (4.1) reports in the Table 4.1

$$Inf_t = \beta_1 + \beta_2 unem_t + u_t \tag{4.1}$$

where Inf_t = inflation rates (%)

$unem_t$ = unemployment rates (%)

Table 4.1

Source	SS	df	MS
Model	32.3284496	1	32.3284496
Residual	478.096987	57	8.38766644
Total	510.425437	58	8.80043856

Number of obs	=	59
F(1, 57)	=	3.85
Prob > F	=	0.0545
R-squared	=	0.0633
Adj R-squared	=	0.0469
Root MSE	=	2.8961

inf	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
unem	.5054734	.2574699	Omitted for the purpose of this exam		
_cons	1.010847	1.491583			

White's general test statistic: 1.0266 Chi-sq (2) $\chi^2 = 5.99$

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity $F_{cri} = F_{2,56} = 3.23$

chi2(1) = 1.12

3.a) Is there evidence of multicollinearity in the data? How do you know? Explain your answers in detail and state the critical value for hypothesis testing to receive full points.

There is no evindence that multicollinearity problem arises in the data.
We are looking from 3 variable and using the rule of thumb.
The rule of thumb is the specific level of criterion that is usually and mutually accepted as a threshold.

1st variable, the correlation coefficient, according to the rule of thumb, it shows that correlation in between the variables should not exceed 0.8, which is not present in this data.

2nd variable, the variance inflating factor (VIF), according to the rule of thumb that correlation should not exceed 0.8, VIF should not exceed 10 too because $VIF = \frac{1}{1 - r_{23}^2}$, it is possible to exceed 10 if and only if r_{23}^2 or correlation is more than 0.8.
Also, VIF does not show the evidence of multicollinearity problem

3rd variable, the tolerance (TOL or $\frac{1}{VIF}$), since it is the reciprocal of VIF, it is $TOL = 1 - r_{23}^2$, therefore this variable should be close to 1 not zero because if it closes to 0, it means the correlation is nearly 1 which we wish it is not happen in our data because correlation must not exceed 0.8. As a result, TOL does not show the evidence of a multicollinearity problem

Another way to detect the multicollinearity is the conflict of the test result
In other words, the high R^2 but coefficients are not different from 0.

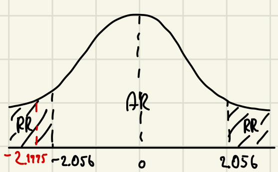
Step 1: $H_0: \beta_0 = 0$

$H_1: \beta_0 \neq 0$

Step 2: $\alpha = 0.05$

Step 3: $t_{cal} = \frac{-34.1349 - 0}{15.6763} = -2.1775$

Step 4: $t_{cri} = t_{0.05, 26} = 2.056$



Reject H_0

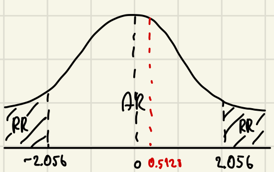
Step 1: $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

Step 2: $\alpha = 0.05$

Step 3: $t_{cal} = \frac{1.5386 - 0}{3.0005} = 0.5128$

Step 4: $t_{cri} = t_{0.05, 26} = 2.056$



Cannot reject H_0

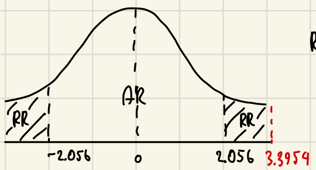
Step 1: $H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

Step 2: $\alpha = 0.05$

Step 3: $t_{cal} = \dfrac{3.9152 - 0}{1.1531} = 3.3954$

Step 4: $t_{cri} = t_{0.05, 26} = 2.056$



Reject H_0

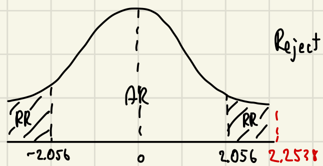
Step 1: $H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

Step 2: $\alpha = 0.05$

Step 3: $t_{cal} = \dfrac{18.8024 - 0}{8.3426} = 2.2538$

Step 4: $t_{cri} = t_{0.05, 26} = 2.056$



Reject H_0

We can clearly see that most of our coefficients are significantly different from 0. Which is not conflict with the high R^2, so there is no multicollinearity problem arises in our data.

3.b) What is the property of BLUE? If there is the multicollinearity problem, is the OLS estimators still retain the property of BLUE? If not, which properties are violated?

BLUE stands for best linear unbiased estimator. The property of BLUE is the variance has the lowest variance as possible in the model.

If there is a multicollinearity problem, OLS estimators still the best linear unbiased estimator because it is the coefficient with the least variance as possible in the model with multicollinearity problem.

Answer the following questions.

4.a) Interpret the intercept and slope coefficients.

4.b) According to the test statistics given after Table 4.1 below, is there any sufficient evidence to conclude that there is heteroscedasticity problem? Show your work on the hypothesis testing. (Use $\alpha = 0.05$)

4.c) Given your test results in a), do the OLS estimators still retain the property of BLUE? If not, which properties are violated?

****

4.a) Interpret the intercept and slope coefficients.

$$Inf_t = \beta_1 + \beta_2 unem_t + u_t$$

where Inf_t = inflation rates (%)

$unem_t$ = unemployment rates (%)

unem	.5054734
_cons	1.010847

The intercept is thet, when unemployment rate is 0%, the infletion is 1.0109 %.
The slope is thet, when unemployment rates increases by 1 %, the inflation rate rises by 0.5055 %.

4.b) According to the test statistics given after Table 4.1 below, is there any sufficient evidence to conclude that there is heteroscedasticity problem? Show your work on the hypothesis testing. (Use α = 0.05)

White's general test statistic: 1.0266 Chi-sq (2)
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
chi2(1) = 1.12

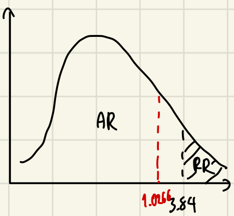
White test

Step 1: H_0: No Heteroscedasticity problems (Homoscedasticity)
H_1: Heteroscedasticity

Step 2: $\alpha = 0.05$

Step 3: $LM_{cal} = 1.0266$

Step 4: $\chi^2_{k-1} = \chi^2_1 = 3.84$



We cannot reject H_0 with 95% confident interval.
This is because our LM_{cal} lies in the acceptance region.
Therefore, we can say that our data has no heteroscedasticity problem

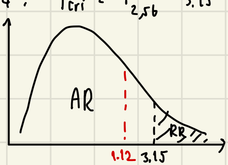
BP-test

Step 1: H_0: Homoscedasticity
H_1: Heteroscedasticity

Step 2: $\alpha = 0.05$

Step 3: $F_{cal} = 1.12$

Step 4: $F_{cri} = F_{2,56} = 3.15$



Since our F_{cal} lies in the acceptance region, it means we cannot reject the null hypothesis. In other words, there is no heteroscedasticity in our data.

4.c) Given your test results in a), do the OLS estimators still retain the property of BLUE? If not, which properties are violated?

The test result in a) still be the BLUE (Best Linear Unbiased estimator) because there is no heteroscedasticity problem arises in our data. (From b))

If not because heteroscedasticity arises, when the heteroscedasticity arises we can use Weight Least Square (WLS) to find the variance for our coefficient, which $var_{WLS} < var_{ordinary\ OLS}$

Therefore, if the heteroscedasticity problem arises, variance and coefficient from ordinary OLS is not efficient anymore. (Variance for WLS is lower)

As a result, if the heteroscedasticity arises

1. the large variance makes it hard to find the significant in our coefficient. (Hard to reject H_0)
2. Our estimators are biased, it can be either positive biased : Overestimate variance
   negative biased : Underestimate variance
3. The conclusion draws might be misleading