EE325 Introductory Econometrics **Section 2** Weerawat Phattarasukkumjorn

Assignment 1

Assigned on Apr 9th, 2022. Due date Saturday, May 7th, 2022 before midnight.

6304640094

AUS ATSAVAKOVITH

Estimate the model (1.1) reports in the Table 1.1

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

Table 1.1

Source	SS	df	MS			
				Number of obs	=	2,725
				F(5, 2719)	=	24.29
Model	85.9532425	5	17.1906485	Prob > F	=	0.0000
Residual	1924.39391	2,719	.707757967	R-squared	=	0.0428
				Adj R-squared	=	0.0410
Total	2010.34716	2,724	.738012906	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			

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)

avgsen

- $narr86_i = \beta_1 + \beta_2 pcnv_i + \beta_3 avgsen_i + \beta_4 tottime_i + \beta_5 ptime_i86 + \beta_6 qemp_i86 + u_i$
  $= 0.7061 - 0.1512 pcnv_i - 0.0070 avgsen_i + 0.0121\ tottime_i - 0.3926\ ptime_i86 - 0.1031\ qemp_i86$

- An increase of 1 month of $avgsen_i$ = A decrease of 0.0070 $narr86_i$
- $H_0 : \beta_3 = 0$
  $H_a : \beta_3 \neq 0$
- t_{cal} for avgsen
  $t_{cal}(\beta_3) = \dfrac{\hat{\beta}_3 - \beta_3}{se(\hat{\beta}_3)} = \dfrac{0.007 - 0}{0.0124} = -0.5645$
- $n-k = 2725 - 6 = 2719$
- critical value
  $\alpha = 0.05$
  $t_{\frac{\alpha}{2}} = t_{0.025} = \pm 1.96$, $[-1.96, 1.96]$
- Since t_{cal} doesn't exceed critical value, we cannot reject null hypothesis

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)

- H_0 : Ethnic background and legal income have no marginal distribution in model
  H_a : otherwise
- F-test → measure the additional variables only
  $F_{cal} = \dfrac{\frac{R^2_{1.2} - R^2_{1.1}}{k'}}{\frac{1-R^2}{n-k}} = \dfrac{\frac{0.0723 - 0.0428}{3}}{\frac{1-0.0723}{2725-9}} = 28.79$
- Critical Value $\alpha = 0.05$
  $F_{3,2716} = 2.6$
- Since F_{cal} lies exceed the critical value, we can reject null hypothesis and ethnic background and legal income have no marginal distribution in model

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.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)

F-test + using R^2

- $H_0 : \beta_i = 0$
  H_a : otherwise, for both test
- Model 1
  $F_{cal} = \dfrac{\frac{R^2}{k}}{\frac{1-R^2}{n-k}} = \dfrac{\frac{0.0428}{6}}{\frac{1-0.0428}{2725-6}} = 20.26$
- Model 2
  $F_{cal} = \dfrac{\frac{R^2}{k}}{\frac{1-R^2}{n-k}} = \dfrac{\frac{0.0723}{9}}{\frac{1-0.0723}{2725-9}} = 23.52$
- Critical Value $\alpha = 0.01$
  M1 : $F_{6,2719} = 2.64$
  M2 : $F_{9,2716} = 2.41$
- Since F_{cal} lies exceed the critical value, we can reject null hypothesis for both model

EE325 Introductory Econometrics **Section 2** Weerawat Phattarasukkumjorn
Assignment 1
Assigned on Apr 9th, 2022. Due date Saturday, May 7th, 2022 before midnight.

6304640094
AUS ATSAVAKOVITH

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.

- $H_0 : \beta_k = 0$
  $H_a : \beta_k \neq 0$ when $k = 1,2,3,4$
- T-test

$$t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{9.1748 - 0}{0.0035} = 2621.37$$

$$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{0.587 - 0}{0.0072} = 81.53$$

$$t_{cal} = \frac{\hat{\beta}_3 - \beta_3}{se(\hat{\beta}_3)} = \frac{-0.0336 - 0}{0.005} = -6.72$$

$$t_{cal} = \frac{\hat{\beta}_4 - \beta_4}{se(\hat{\beta}_4)} = \frac{0.0444 - 0}{0.0102} = 4.35$$

- d.f. = n-k = 97,878 - 4 = 97,874
- critical value $\alpha = 0.05$
  $t_{\frac{\alpha}{2}} = \pm 1.96$
- Since all t_{cal} exceed the critical value, we can reject H_0 for all.

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

Consider β_2 which is the coefficient of civil servant, it can be interpret as the different between civil servant and state employee.

Since it is a positive, means on average civil servant earn more than state employee by $100(e^{\hat{\beta}_2} - 1) = 79.86\%$

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

Consider β_3 which represent as a coefficient of year since we would find average between year 2019-2020.

Since it is a negative, means on average of pandemic effect will drop the overall wage by $100(e^{\beta_3} - 1) = 3.3\%$.

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.

We need to calculate separetely

- $\ln \widehat{wage} = 9.1748 + 0.587(1) - 0.0336(1) + 0.0444(1)\cdot(1)$
- $\ln \widehat{wage} = 9.1748 + 0.587(0) - 0.0336(1) + 0.0444(0)\cdot(1)$

even though the wage went down by 0.0336 from year coefficient
it increase again by 0.0444 from civil groups

$\therefore$ the civil group still better-off during the pandemic by $100(e^{\hat{\beta}_3 + \hat{\beta}_4} - 1) = 1.09\%$ increase for overall wage
the state employee worst-off $\longrightarrow$ by $100(e^{\hat{\beta}_3} - 1) = 3.3\%$ decrease for overall wage

6304640094
AUS ATSAVAKOVITH

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			
Model	5020.64468	3	1673.54823	Number of obs	=	30
Residual	2642.24237	26	101.624706	F(3, 26)	=	16.47
				Prob > F	=	0.0000
				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.

- $H_0 : \beta_k = 0$
  $H_a : \beta_k \neq 0$ when $k = 1,2,3,4$
- T-test

$$t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{-34.1349 - 0}{15.6763} = 2.18$$

$$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{1.5386 - 0}{3.0005} = 0.51$$

$$t_{cal} = \frac{\hat{\beta}_3 - \beta_3}{se(\hat{\beta}_3)} = \frac{3.9152 - 0}{1.1531} = 3.4$$

$$t_{cal} = \frac{\hat{\beta}_4 - \beta_4}{se(\hat{\beta}_4)} = \frac{18.8023 - 0}{8.3426} = 2.25$$

- d.f. = n-k = 30-4 = 26
- critical value $\alpha = 0.05$
  $t_{\frac{\alpha}{2}} = \pm 2.056$
- Since t_{cal} except β_2 exceed the critical value, we can reject H_0 except β_2
- $R^2 = 0.6552$
- Since we ecknowledge that none of perameters ere differed from 0 we mey conclude thet multicollinearity is not present here.

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 = Best Linear Unbiased Estimator
 └ the least variance measured
 └ leads to true parameters

BLUE is not affected by multicollinearity but hypothesis testing does.

6304640094
AUS ATSAVAKOVITH

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

$$Inf_t = \beta_1 + \beta_2 unem_t + u_t \qquad \textbf{(4.1)}$$

where Inf_t = inflation rates (%)

$unem_t$ = unemployment rates (%)

Table 4.1

Source	SS	df	MS			
Model	32.3284496	1	32.3284496	Number of obs	=	59
Residual	478.096987	57	8.38766644	F(1, 57)	=	3.85
				Prob > F	=	0.0545
				R-squared	=	0.0633
				Adj R-squared	=	0.0469
Total	510.425437	58	8.80043856	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)

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

chi2(1) = 1.12

4.a) Interpret the intercept and slope coefficients.

$\hat{\beta}_1$ = intercept at 1.01
L when unemployment rate is 0, inflation rate is 1.01

$\hat{\beta}_2$ = slope is 0.5
L when unemployment rate increase by 1%, inflation rate increase by 0.5%

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)

- LM_{cal} for χ^2_{k-1} = 1.0266
- Critical value, α = 0.05, χ^2_1 = 3.8415
- LM_{cal} is not exceed χ^2_1, thus we cannot reject the null hypothesis

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

Since we cannot reject H_0, BLUE is not violated.