Final Exam EE325 KEY (Section 2 Semester 1/2011)

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

1.a) (4 points) 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 wheather the average sentence served from prior convictions has an impact on the number of arrests in the current year (1986). Show your work. (Use a = 0.05)

- $\widehat{narr86}_i = 0.71 - 0.15\, pcnv_i - 0.007\, avgsen_i + 0.01\, tottime_i - 0.04\, ptime86_i - 0.1\, qemp86_i$
- A month of prior average sentenced for prior conviction increase is corresponding to the decrease of number of arrests in 1986 by 0.07 case.
- Calculate t cal for avgsen:

  $t_{cal}(\beta_3) = \dfrac{\hat{\beta}_3 - \beta_3}{se(\hat{\beta}_3)} = -\dfrac{0.007 - 0}{0.012} = -0.58$
- The critical value for the degrees of freedom $(n-K)$ $2{,}725 - 6 = 2{,}716$ is ± 1.96 when $a = 0.05$
- Since the calculated statistic does not exceed the critical value (in the negative zone), we cannot reject the null hypothesis. Therefore, we cannot make sure that the parameter is different from zero.

1.b) (4 points) What is the overall significance of the regression from model (1.1) and Model (1.2)? What test do you use? (Use a = 0.01)

- Overall significance can be tested by an F-test, using R^2
- H_0: All β_K are 0 simultaneously. H_a: otherwise, for both tests.
- Model 1.1: $F_{cal} = \dfrac{R^2/K-1}{1-R^2/(n-K)} = \dfrac{0.0428/5}{(1-0.0428)/2{,}725-6} = 26.48$
- Model 1.2: $F_{cal} = \dfrac{R^2/K-1}{1-R^2/(n-K)} = \dfrac{0.0723/8}{(1-0.0723)/(2{,}725-9)} = 26.48$
- The critical values are 2.64 and 2.41 for model 1.1 and model 1.2 respectively ($F_{6,2719}$ and $F_{9,2716}$) when $a = 0.01$
- We can reject null hypothesis for both models. On the other hand, we can just simply consider the p-value (prob > F) to answer this question

1.c) (4 points) 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)

- For this question, we can apply the 'marginal contribution' test, comparing between Model 1.1 and 1.2 since the model 1.2 has added those variables.
- H_0: ethic background and legal income have no marginal contribution to the model.

  H_a: otherwise.
- Calculate the F-test: $F_{cal} = \dfrac{R^2_{1.2} - R^2_{1.1} / (K \text{ of new regressors})}{1 - R^2_{1.2}/(n - K_{1.2})} = \dfrac{0.0723 - 0.0428/3}{(1-0.0723)/(2{,}725-9)} = 28.79$
- The critical value of $F_{3,2716}$ is 2.6 when $a = 0.05$
- Fcal exeeds the critical value, which means that we can reject the null hypothesis and we can make sure that ethic background and legal income have marginal contribution to the model.

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			
				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			
avgsen	-.0070487	.0124122			
tottime	.0120953	.0095768			

Omitted for the purpose of this exam

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			

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) (3 points) Test all the parameters individually if each of them is significantly different from zero of not.

- setting up hypothesis : $H_0 : \beta_K = 0$ and $H_0 : \beta_K \neq 0$ when $K = 1, 2, 3, 4$
- Calculate t cal for all parameters :

$$t_{cal}(\beta_1) = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{9.1748 - 0}{0.0035} = 2{,}621.37$$

$$t_{cal}(\beta_3) = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{0.587 - 0}{0.0072} = 81.53$$

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

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

- The critical value for the degrees of freedom (n−K) 97,878−4 = 97,874 is $\pm$ 1.96 when $\alpha = 0.05$
- All the t_{cal} exceeds the critical value. Therefore, we can reject all the hypotheses of β_K and conclude that all the parameters are significantly different from zero (not simultaneously but individually).

2.b) (3 points) How much on average does a civil servant and state employee earns more or less than the others disregarding the year?

- If we only consider separated effect of civil servant, we can consider β_2 which is a coefficient representing difference between civil servants and other groups.
- This is positive coefficient, therefore, it means that on average civil servants earn more than the other groups by $100 \times (e^{\hat{\beta}_2} - 1) = 79.86$ percent.

2.c) (3 points) How much on average does the pandemic affect wage overall?

- Consider the year parameter since the data are from the second quarter of 2019 and 2020, the pandemic effect is realized in the second quarter of 2020. Hence, β_3 is the representation of that effect.
- This is negative coefficient. Therefore, it means that in 2020, overall wage drops by $100 \times (e^{\hat{\beta}_3} - 1) = 3.30$ percent for all groups.

2.d) (3 points) 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.

- When we consider the interaction term, we can see that once the civil servant variable is interacted with year (2020), there is a positive effect and significantly different from zero.
- Consider only in 2020, the civil sevant group overall wage can be displayed as follows.
  $\ln \widehat{wage}_i = 9.1748 + 0.587(1) - 0.0336(1) + 0.00444(1)\cdot(1)$ while the other groups are
  $\ln \widehat{wage}_i = 9.1748 + 0.587(1) - 0.0336(1) + 0.00444(0)\cdot(1)$
- The interaction term does not add to the other groups which means that even the civil servant wage went down by the year coefficient (-0.0336) in 2020, the interaction term shows a bounce for this group as well $(+0.0444)$.
- In summary, the civil servant group (control group) is still better-off doing the pandemic by $100 \times (e^{\hat{\beta}_3 + \hat{\beta}_4} - 1) = 1.09$ percent increase of overall wage for this group. In the meantime, the other groups are worse-off by $100 \times (e^{\hat{\beta}_3} - 1) = 3.30$ decrease of overall wage.
- This makes economic sense since civil servants' wage did not drop during the pandemic while the rest may find themselves worked less due to the lowered hours or limited work placements

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) (5 points) 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.

- VIF and coefficient of correlation do not reveal any sign of multicollinearity if we rely on the rules of thumb at 0.8 and 10.

- Another kind of test we can pose critical value is considering each t-test and R^2. If multicollinearity is present, they must be conflicting, namely high R^2 but rare significant parameter. Therefore, we should test all the parameters.
- setting up hypotheses: $H_0: \beta_K = 0$ and $H_0: \beta_K \neq 0$ when $K = 1,2,3,4$
- Calculate t cal for all parameters:

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

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

$$t_{cal}(\beta_3) = \frac{\hat{\beta}_3 - \beta_3}{se(\hat{\beta}_3)} = \frac{18.8023 - 0}{8.3426} = 2.25$$

- The critical value for the degrees of freedom $(n-K)$ $30-4 = 26$ is ± 2.056 when $a = 0.05$. Therefore, we can reject the null hypothesis for all the parameters but β_2.
- Now we can see clearly that the R^2 is 0.6552 and almost all the parameters are significantly different from zero, we may conclude that conflicting test is not found in this model and may be able to conclude that multicollinearity is not present here.

3.b) (3 points) 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 refers to Best Linear Unbiased Estimator, meaning that the estimators have the lowest variance possible (best). They are unbiased or when there are more samples. probability limit of the estimators tend towards true parameters.

- BLUE is not affected by multicollinearity. However, multicollinearity causes misleading conclusion of hypothesis testing (High R^2 but rare significant parameter)

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

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

where Inf_t = inflation rates (%)

$unem_t$ = unemployment rates (%)

Table 4.1

Source	SS	df	MS			
				Number of obs	=	59
				F(1, 57)	=	3.85
Model	32.3284496	1	32.3284496	Prob > F	=	0.0545
Residual	478.096987	57	8.38766644	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

Answer the following questions.

4.a) (2 points) Interpret the intercept and slope coefficients.

- For $\hat{\beta}_1$, the intercept of 1.01, it means that when unemployment rate is 0, inflation rate is 1.01 on average.
- For $\hat{\beta}_2$, the slope of 0.5, it means that when unemployment rate increases by 1 percent, inflation rate increases by 0.5 percent on average.

4.b) (3 points) 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)

- Using the White's test, we have the LM_{cal} for χ^2_{K-1} already, which is 1.0266.
- The critical value when χ^2_1 and $\alpha = 0.05$ is 3.84146.
- $LM_{cal} < \chi^2_1$, so we cannot reject the null hypothesis of homoscedasticity.

4.c) (3 points) 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 the null hypothesis in 4.b), therefore, we can make sure that BLUE property is not violated.