Economics 113 Professor Spearot
Introduction to Econometrics
Fall 2011 – Final
Name Key ID

Final – 120 Points

You must answer all questions. Please write your name on every page. The exam is closed book and closed notes. You may use calculators, but they must not be graphing calculators. No cell phones. Do not use your own scratch paper.

You must show your work to receive full credit

I have neither given nor received unauthorized aid on this examination, nor have I concealed any similar misconduct by others.

Signature Key

Problem 1 (40 Points)

Consider the following simple specification that tests for regional differences in hours worked:

$$hours = \beta_0 + \beta_1 urban + u$$

hours is average hours worked per week, and *urban* is a dummy variable that takes on a value of 1 if the respondent lives in a metropolitan area, and 0 otherwise. The results from estimating this equation are below:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  43.7386     0.4448  XXXXXXXXXXXXXXX
urban         0.2658     0.5251  XXXXXXXXXXXXXXX
---
Multiple R-squared: 0.0002747,     Adjusted R-squared: -0.0007969
F-statistic: 0.2563 on 1 and 933 DF, SSR=48731.95
```

a.) Please construct and interpret a 95% confidence interval for the intercept. **(10 Points)**

$t_{crit} = 1.96$ +2

$43.7 - 0.44 \cdot 1.96 < \beta_0 < 43.7 + 0.44 \cdot 1.96$

$42.84 < \beta_0 < 44.56$ +4

With 95% confidence, a person that lives in a rural (non-urban) location ~~works~~ between 42.84 and 44.56 hours per week. +4

b.) I claim that urban residents work a number of hours that is significantly different than rural residents. What is the probability that I'm wrong? **(10 Points)**

$$Pr\left(|z| > \frac{0.2658}{0.5251}\right) = 2\left(1 - Pr\left(z < \frac{0.2658}{0.5251}\right)\right)$$

$$= 2(1 - Pr(z < 0.506))$$

$$= 2(1 - 0.6950)$$

$$= \boxed{0.61}$$

+4

+6

reasonable work is ok

c.) Suppose that instead of the regression in 'a', I run the following regression:

$$hours = \beta_0 + \beta_1 urban + \beta_2 educ + u$$

where *educ* is the years of education of the respondent. The results from estimating this equation are below:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  39.8146     1.4889  XXXXXXXXXXXXXXX
urban         0.1613     0.5246  XXXXXXXXXXXXXXX
educ          0.2969     0.1076  XXXXXXXXXXXXXXX
---
Multiple R-squared: 0.008383, Adjusted R-squared: 0.006255
F-statistic:  3.94 on 2 and 932 DF, SSR=48336.7
```

In comparing the regression in 'a' and the regression in 'c', what is the correlation between *educ* and *urban*? Why? **(10 Points)**

+5

The correlation is positive. In 'a' educ is an omitted variable. Given the positive relationship between educ and hours, and given Burban goes down, there must be a positive correlation between educ and urban.

+5

d.) Does the model in 'c' tell us anything about hours worked? If a hypothesis test is warranted, test this hypothesis at the 95% level, stating your null and alternative hypotheses. If not, provide other evidence for your answer. **(10 Points)**

$H_0: \beta_1 = \beta_2 = 0$

H_A: H_0 not true } +1

$F_{stat} = 3.911$

$q=2 \Rightarrow F_{crit} = 3.00$ } +1

$F_{stat} > F_{crit} \Rightarrow$ Reject null in favor of alternative.

+2

e.) Suppose that I modify the regression in 'c' to include *age* and *age2*, which are the age and age squared of the respondent.

$$hours = \beta_0 + \beta_1 urban + \beta_2 educ + \beta_3 age + \beta_4 age^2 + u$$

The results from this regression are below:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 42.078061  31.524587  XXXXXXXXXXXXXXX
urban        0.166803   0.525423  XXXXXXXXXXXXXXX
educ         0.299204   0.108045  XXXXXXXXXXXXXXX
age         -0.200219   1.914023  XXXXXXXXXXXXXXX
I(age^2)     0.003918   0.028749  XXXXXXXXXXXXXXX
---
Multiple R-squared: 0.009078, Adjusted R-squared: 0.004816
F-statistic:  2.13 on 4 and 930 DF,  SSR=48302.81
```

At what age is average hours worked minimized? Show your work!! **(10 Points)**

$$\frac{\partial \widehat{hours}}{\partial age} = \hat{\beta}_3 + 2\hat{\beta}_4 age = 0 \quad +5$$

$$age = -\frac{\hat{\beta}_3}{2\hat{\beta}_4}$$

$$= \frac{0.200}{2 \cdot 0.0039}$$

+5

$$age = 25.64$$

f.) Is the model in 'e' preferred to the model in 'c'? If a hypothesis test is warranted, test this hypothesis at the 95% level, stating your null and alternative hypotheses. If not, provide other evidence for your answer. **(10 Points)**

e is nested within c $\Rightarrow$ F Test

$SSR_{UR} = 48{,}302.8$ $q = 2$ $H_0: \beta_3 = 0, \beta_4 = 0$

$SSR_R = 48{,}336.7$ $df = 930$ H_A: H_0 not true

+1

$$F_{stat} = \frac{\frac{48{,}336.7 - 48{,}302.8}{2}}{\frac{48302.8}{930}} = 0.326$$

+1

$F_{crit} = 3.00$ $F_{stat} < F_{crit}$

Fail to reject H_0.

+2

Problem 2 (40 Points)

a.) For this problem, we wish to study the impact of health insurance on the smoking behavior of pregnant mothers. While difficult to assess, we will leverage a family's eligibility for prenatal care via Medicaid to determine the effects of health insurance on behavior. To do so, we run the following regression:

$$smoke = \beta_0 + \beta_1 faminc + \beta_2 medicaid + \beta_3 faminc \cdot medicaid + u$$

Here, *smoke* takes on a value of 1 if a mother smoked during pregnancy, and zero otherwise. Further, *faminc* is yearly family income (in thousands) and *medicaid* is a dummy variable taking a value of 1 if *faminc* is below 22 (which is \$22,000) and zero otherwise. What kind of regression technique is this? **(10 Points)**

Regression Discontinuity all or nothing

b.) The results from estimating the regression in 'a' are below:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.2053321	0.0329005	XXXXXXXXXXXXXXX	
faminc	-0.0024584	0.0007648	XXXXXXXXXXXXXXX	
medicaid	0.0810480	0.0601873	XXXXXXXXXXXXXXX	
I(medicaid * faminc)	-0.0037862	0.0036022	XXXXXXXXXXXXXXX	

Multiple R-squared: 0.02766, Adjusted R-squared: 0.0252
F-statistic: 11.26 on 3 and 1187 DF, SSR= 135.3847

Please use a t-test to test whether Medicaid eligibility (at the eligibility threshold) affects smoking behavior. Please state your null and alternative hypotheses, and test the null against the alternative at the 99% level. **(10 Points)**

$H_0: B_2 = 0$
$H_A: B_2 \neq 0$ } +4

$t_{stat} = \frac{0.081 - 0}{0.0601} = 1.35$

$t_{crit} = 2.575$ } +4

$|t_{stat}| < t_{crit} \Rightarrow$ Fail to reject H_0

Eligibility does not have an effect on smoking behavior at the threshold. } +2

c.) Does the relationship between family income and maternal smoking behavior depend on whether the family is eligible for Medicaid? Test this hypothesis at the 98% level using a two-sided test. State your null and alternative, and show your work! **(10 Points)**

$H_0: \beta_3 = 0$ $\Rightarrow$ $t_{stat} = \frac{-0.00379}{0.0036} = -1.058$

$H_1: \beta_3 \neq 0$

$t_{crit} = 2.325$

$|t_{stat}| < t_{crit}$ $\Rightarrow$ Fail to reject null

There is no significant relationship between income and smoking that is conditional on medicaid

d.) Suppose that instead of the above model, we estimate the following model:

$$smoke = \beta_0 + \beta_1 motheduc + \beta_2 medicaid + u$$

where *motheduc* is the mother's education level in years. The results are below:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.520542   0.059349  XXXXXXXXXXXXXXX
motheduc    -0.030276   0.004257  XXXXXXXXXXXXXXX
medicaid     0.040296   0.022489  XXXXXXXXXXXXXXX
---
Multiple R-squared: 0.05677,  Adjusted R-squared: 0.05519
F-statistic: 35.75 on 2 and 1188 DF,  SSR=131.3309
```

Please interpret the coefficient on *medicaid*, and test whether this coefficient is significantly different from zero. Please state your null and alternative hypotheses, and test the null against the alternative at the 90% level. **(10 Points)**

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$ + 2

$t_{stat} = \frac{0.040296 - 0}{0.022489} = 1.79$

$t_{crit} = 1.645$ +2

+1 $|t_{stat}| > t_{crit}$ $\Rightarrow$ Reject null in favor of alternative

Mothers eligible for medicaid have a 0.0402 larger probability of smoking. + 6

e.) Which regression is preferred, the regression in '2b' or the regression in '2d'? If a hypothesis test is warranted, test this hypothesis at the 95% level, stating your null and alternative hypotheses. If not, provide other evidence for your answer. **(10 Points)**

Non-Nested $\Rightarrow$ Adjusted R^2

$Adj\ R^2_b = 0.0252$ +4

$Adj\ R^2_d = 0.0552$ +4

$\Rightarrow$ Model in "d" is preferred +2

+3 for F-test

f.) Using the previous regression equation in 'd', we wish to predict the probability of smoking for a mother with 20 years of education that is eligible for Medicaid. Please derive a regression equation that allows us to generate this prediction with standard error, and write the R commands that would estimate this particular equation. Show your work!! **(10 Points)**

$$\Theta = \beta_0 + \beta_1 \cdot 20 + \beta_2 \cdot 1 \quad \Big\} + 3$$

$$\Rightarrow \beta_0 = \Theta - \beta_1 \cdot 20 - \beta_2 \cdot 1$$

$$+4$$

$$\Rightarrow \text{smoke} = (\Theta - \beta_1 \cdot 20 - \beta_2 \cdot 1) + \beta_1 \text{motheduc} + \beta_2 \text{medicaid} + u$$

$$\text{smoke} = \Theta + \beta_1(\text{motheduc} - 20) + \beta_2(\text{medicaid} - 1) + u$$

#) $X\$M20 = X\$\text{motheduc} - 20$

$X\$Med1 = X\$\text{medicaid} - 1$

$\text{lm}(\text{smoke} \sim M20 + Med1, x)$ $\Big\} + 3$

or

Or

$$\text{lm}(\text{smoke} \sim I(\text{motheduc} - 20) + I(\text{medicaid} - 1), x) \quad +3$$

Have a nice holiday!!!