Economics 113 Professor Spearot
Introduction to Econometrics
Spring 2008 – Final Exam
Name John P. Answer Key

Final Exam – 150 Points

You must answer all the questions. The exam is closed book and closed notes. You may use calculators, but they must not be graphing calculators. Do not use your own scratch paper.

<u>You must show your work to receive full credit</u>

You have plenty of time to finish. Take your time and relax. And, have a safe and wonderful Summer!

Problem 1 **(30 Points)**

You roll two dice. Both dice are fair. The first one has THREE sides and the second one has SIX sides.

a.) Draw and label the Venn diagram describing all possible sample points. **(5 Points)**

	1	2	3	4	5	6
1						
2						
3						

+2

+2

Also possible to use a tree diagram.

+1 for work

b.) What is the probability that you will get a total of four or more points between the two dice? **(5 Points)**

	1	2	3	4	5	6
1			X	X	X	X
2		X	X	X	X	X
3	X	X	X	X	X	X

+3

May refer to "a"

$$Pr(T \geq 4) = \frac{15}{18} = \boxed{\frac{5}{6}}$$

+2

+2 for work

c.) Given that you roll a two with one of the two dice what is the chance that the two dice together will total 4? **(10 Points)**

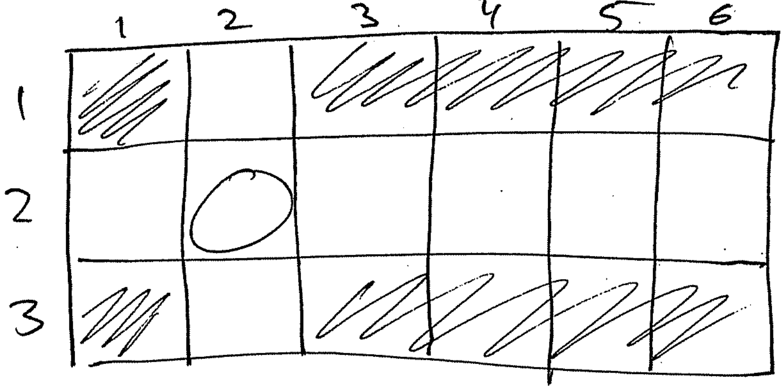
$$Pr(T=4 \mid 2) = \frac{Pr(T=4 \cap 2)}{Pr(2)} \quad +2$$

$$Pr(2) = \frac{8}{18} \quad +2$$

$$Pr(T=4 \cap 2) = \frac{1}{18} \quad +2$$

$$Pr(T=4 \mid 2) = \frac{\frac{1}{18}}{\frac{8}{18}} = \frac{1}{8}$$

+2



+2 for shading

d.) Given that you roll a three with one of the two dice what is the chance that the two dice together will total a value greater than 4? **(10 Points)**

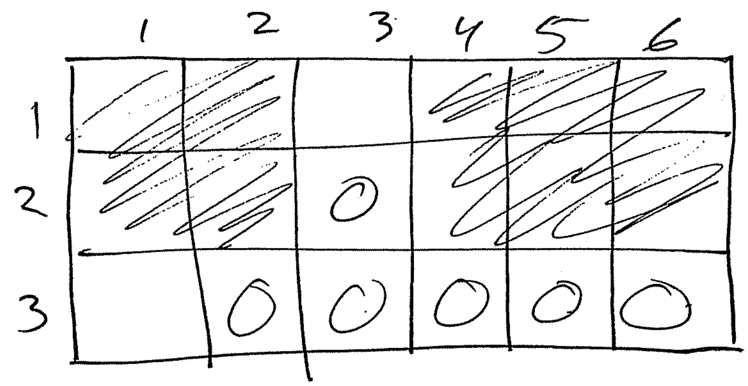
$$Pr(T>4 \mid 3) = \frac{Pr(T>4 \cap 3)}{Pr(3)} \quad +2$$

$$Pr(3) = \frac{8}{18} \quad +2$$

$$Pr(T>4 \cap 3) = \frac{6}{18} \quad +2$$

$$Pr(T>4 \mid 3) = \frac{\frac{6}{18}}{\frac{8}{18}} = \boxed{\frac{6}{8}}$$

+2



+2

Problem 2 (90 points)

Suppose that I run the following regression predicting the effects of classroom performance on students' final exam grades:

$$final = \beta_0 + \beta_1 section + \beta_2 mt1 + \beta_3 hwtotal + u$$

Here, *final*, *mt1*, *hwtotal*, *section* are the percent scores on the final, midterm, homework, and section participation, respectively. The results from running this regression are below.

```
. regress final section mt1 hwtotal

      Source |       SS       df       MS              Number of obs =     142
-------------+------------------------------           F(  3,   138) =   26.49
       Model |  12155.6037     3  4051.86791           Prob > F      =  0.0000
    Residual |   21109.933   138  152.970529           R-squared     =  0.3654
-------------+------------------------------           Adj R-squared =  0.3516
       Total |  33265.5367   141  235.925792           Root MSE      =  12.368

------------------------------------------------------------------------------
       final |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     section |   .0795122    .058387    xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
         mt1 |   .4671107   .0669202    xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
     hwtotal |    .235302   .0720338    xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       _cons |   14.46427   7.723413    xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

a.) Please interpret the constant. **(5 points)**

+1

On average, a student that earns 0% in section, the first midterm, and hw total will earn 14.46% on the final exam

+1 +1 +2

b.) I claim that getting a higher grade on homework increases your predicted grade on the final. Conduct a one-sided hypothesis test at the 5% level for the coefficient on *hwtotal*, β_3. Please state your null and alternative hypotheses, and briefly interpret the result. **(10 Points)**

+1 $H_0: B_3 = 0$

+1 $H_A: B_3 > 0$

$t_{stat} = \dfrac{0.235 - 0}{0.0721} = 3.26$ +2

$t_{crit} = 1.64$ +1

$t_{stat} > t_{crit}$ +1

Reject null!! +1

Homework performance has a positive effect on grades which has a statistically significant difference from zero. +3

c.) Construct a 99% confidence interval for the coefficient on *section*, β_1. **(10 Points)**

+5 $\hat{\beta}_1 - se(\hat{\beta}_1) \cdot t_{crit} < \beta_1 < \hat{\beta}_1 + se(\hat{\beta}_1) \cdot t_{crit}$

$0.0795 - 0.0584 \cdot 2.57 < \beta_1 < 0.0795 + 0.0584 \cdot 2.57$

+5 $-0.0706 < \beta_1 < 0.23$

d.) I have reason to suspect that the variability of final exam scores changes with previous performance (homework, midterms, section). What is this called? What can be done about it? What Stata commands are necessary? **(5 Points)**

Heteroskedasticity +2

Robust standard errors +2

Use 'robust' command. +1

e.) I want to test the suspicion in 'd' rigorously. I run the following regression:

$$\hat{u} = \delta_0 + \delta_1 section + \delta_2 mt1 + \delta_3 hwtotal + \varepsilon$$

Here, $\hat{u}$ is the residual from the regression in 'a'. The estimates are as follows:

Source	SS	df	MS
Model	7.2760e-12	3	2.4253e-12
Residual	21040.4593	138	152.467096
Total	21040.4593	141	149.223115

Number of obs	=	142	
F(3,	138)	=	0.00
Prob > F	=	1.0000	
R-squared	=	0.0000	
Adj R-squared	=	-0.0217	
Root MSE	=	12.348	

uhat	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
section	-3.84e-09	.0582909	-0.00	1.000	-.1152588	.1152588
mt1	-1.13e-08	.06681	-0.00	1.000	-.1321036	.1321036
hwtotal	1.12e-08	.0719152	0.00	1.000	-.1421981	.1421982
_cons	1.88e-07	7.710694	0.00	1.000	-15.24638	15.24638

The f-statistic for the full exclusionary test is very low (zero), which implies that the variables of the model tell us very little about the dependent variable. Does this address the assertion in 'd'? If not, suggest an alternative. What assumption is at play here? **(10 Points)**

No, Does not. +2

Since $E(u|x)=0$, the residuals will be independent of the explanatory variables. +4

Instead, you should regress: +4

$$\hat{u}^2 = \delta_0 + \delta_1 \text{ section} + \delta_2 \text{ mt1} + \delta_3 \text{ hwtotal} + \varepsilon$$

And run a full-exclusion test.

f.) I suspect that the return to homework scores is dependent on whether or not you attend sections. To examine this possibility, I run the following regression:

$$final = \beta_0 + \beta_1 section + \beta_2 mt1 + \beta_3 hwtotal + \beta_4 hwtotal * section + u$$

The results from estimating this equation are below:

```
. regress final section mt1 hwtotal hwtotal*section

      Source |       SS       df       MS              Number of obs =     142
-------------+------------------------------           F(  4,   137) =   19.90
       Model |  12225.0771     4  3056.26926           Prob > F      =  0.0000
    Residual |  21040.4597   137  153.579998           R-squared     =  0.3675
-------------+------------------------------           Adj R-squared =  0.3490
       Total |  33265.5367   141  235.925792           Root MSE      =  12.393

------------------------------------------------------------------------------
          final |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        section |   .0842367   .0589235     1.43   0.155    -.0322803    .2007538
            mt1 |   .7592581   .4395156     1.73   0.086    -.1098538     1.62837
        hwtotal |   .4647203   .3486564     1.33   0.185    -.2247237    1.154164
hwtotal*section |  -.0034208   .0050861   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
          _cons |  -5.333285   30.43569    -0.18   0.861    -65.51776    54.85119
------------------------------------------------------------------------------
```

Derive the return to section attendance. Plug in the estimated coefficients where necessary. Please interpret briefly. **(10 Points)**

$$\frac{\partial final}{\partial section} = B_1 + B_1 \, hwtotal \qquad + \cancel{8}$$

$$= 0.0842 - 0.0034 \, hwtotal \qquad + 3$$

The returns to ~~those~~ section are smaller as you get better grades on homework $+3$

g.) What is the homework score which yields a negative return to section attendance? Given that homework scores are between 0 and 100, is the return to section attendance always positive? **(10 Points)**

$$\frac{\partial final}{\partial section} = 0 \quad \text{if} \quad \hat{\beta}_1 + \hat{\beta}_4 \, hwtotal = 0$$

$$hwtotal = -\frac{\beta_1}{\beta_4}$$

$$= -\frac{(0.0842)}{(-0.00342)}$$

$$= 24.62$$

For homework scores above 24.62, the returns to section are negative.

h.) Is there a significant interaction between homework and section attendance? Conduct a two-sided test at the 1% level, stating your null and alternative hypotheses, also briefly interpreting the result. **(10 Points)**

$H_0: \beta_4 = 0$ +1

$H_A: \beta_4 \neq 0$ +1

$$t_{stat} = \frac{-0.00342 - 0}{0.00509} \quad +2$$

$$= -0.672$$

$$t_{crit} = 2.57 \quad +1$$

+1

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

+1

~~Return~~ ~~Return to section attendance~~

+3

There exists no significant interaction between the returns to section and homework, and vice versa.

i.) Rather than using interactions as in 'f', I have added in squared terms of homework, *hwtotalsqr*, and section attendance, *sectionsqr*. Their coefficients are β_5 and β_6, respectively.

```
. regress final section mt1 hwtotal hwtotalsqr sectionsqr

      Source |       SS       df       MS              Number of obs =     142
-------------+------------------------------           F(  5,   136) =   15.84
       Model |  12244.9991     5  2448.99983           Prob > F      =  0.0000
    Residual |  21020.5376   136  154.562776           R-squared     =  0.3681
-------------+------------------------------           Adj R-squared =  0.3449
       Total |  33265.5367   141  235.925792           Root MSE      =  12.432

------------------------------------------------------------------------------
       final |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     section |  -.0044826   .2068736    -0.02   0.983    -.4135877    .4046226
         mt1 |   .4710029   .0674799     6.98   0.000     .3375573    .6044485
     hwtotal |   .4722948   .3601857     1.31   0.192    -.2399944    1.184584
  hwtotalsqr |  -.0017249   .0025807    -0.67   0.505    -.0068284    .0033786
  sectionsqr |   .0006319   .0015609     0.40   0.686    -.0024549    .0037186
       _cons |   9.323664    13.5762     0.69   0.493    -17.52409    36.17142
------------------------------------------------------------------------------
```

Which model is preferred, the one in 'i', the one in 'a'? Please justify your answer. If a test is required, state your null and alternative hypotheses, test it at the 5% level, an~~d briefly interpret the result~~. **(10 Points)**

$H_0: \beta_5 = \beta_6 = 0$ +1

$H_A:$ H_0 not true +1

$SSR_{UR} = 21020$ +1

$MSE_{UR} = 154.6$ +1

$SSR_R = 21109$ +1

$$F_{stat} = \frac{\frac{21109 - 21020}{2}}{154.6} = 0.287 \quad +2$$

$F_{crit} = 3.07$ +1

Fail to reject H_0. +2

j.) Suppose that natural ability is an unobserved variable, which does not change over time. I am worried that not including it may be causing omitted variable bias. What technique is appropriate for this problem, and why? **(10 Points)**

Differencing +5

+5 Because natural ability is time invariant, taking the difference between two years yields a specification where ability has no effect.

Problem 3 (30 Points)

Professor Spearot is getting older. He is worried about a receding hair line. To analyze male hair patterns as a function of demographics, he estimates the following linear probability model using a sample of men:

$$Bald = \beta_0 + \beta_1 \text{Age} + \beta_2 Dad + u$$

Bald takes on the value of 1 if the respondent is bald, and 0 otherwise. *Age* is the Age of the respondent, and *Dad* is an indicator variable taking the value of 1 if the respondent's father is bald and 0 otherwise.

a.) Suppose that β_2 is positive. How do I interpret the estimate of the coefficient on *Dad*, β_2? **(5 Points)**

+2

Having a Dad that is bald increases the +2

probability that I am bald by B_2.

+

b.) Suppose that I estimate the model, and I generate predictions for each respondent. Some predictions are negative. Is this sensible? What alternative estimation procedure could remedy this problem? Why? **(10 Points)**

No, since the predictions are probabilities

and negative probabilities don't make +6

sense.

Use Probit or Logit. +4

Probability $\in [0, 1]$ with probit and logit.

c.) Suppose that Stress, and unobserved variable, increases with age. Stress also lead to a higher likelihood of baldness. What is this called? In what direction is the bias? **(5 Points)**

in B_1

Omitted variable bias. +3

Positive Bias. B_1 is over-estimated. +2

d.) Professor Spearot's father is Bald (sorry Dad!). Professor Spearot is 29 years old. Please **derive** the estimating equation required to generate a prediction for somebody with Professor Spearot's characteristics. Please also write the precise STATA commands required to run this regression. **(10 points)**

$$\theta = \beta_0 + \beta_1 \cdot 29 + \beta_2 \cdot 1 \qquad +2$$

$$Bald = \beta_0 + \beta_1 Age + \beta_2 Dad \qquad +1$$

$$+2$$

$$= \theta - \beta_1 \cdot 29 - \beta_2 \cdot 1 + \beta_1 Age + \beta_2 Dad$$

$$Bald = \theta + \beta_1 (Age - 29) + \beta_2 (Dad - 1) \qquad +2$$

gen Dad_1 = Dad - 1 +1

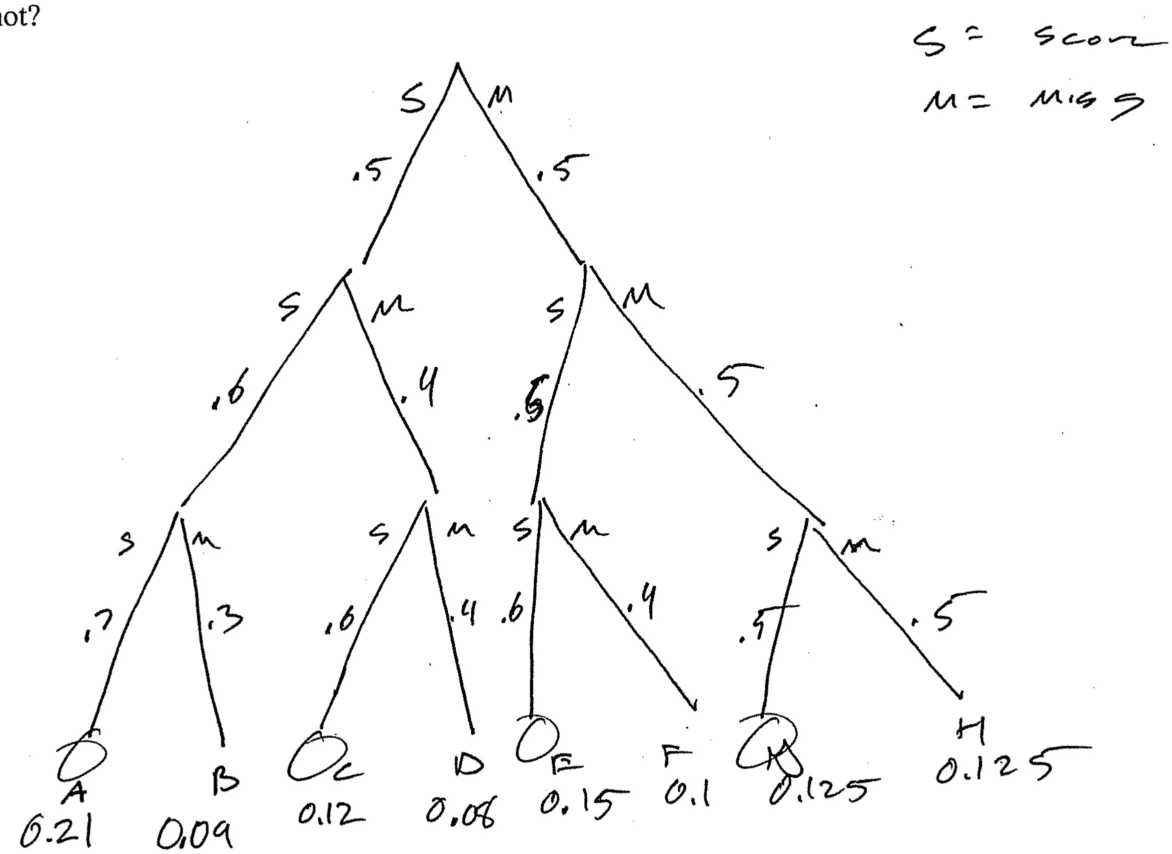
gen Age_29 = Age - 29 +1

regress Bald Age_29 Dad_1 +1

Extra Credit: (10 Points)

Bob Baden was once a college hockey player (no joke here). Skilled and graceful, he was an offensive weapon.

Suppose that Bob takes three shots at the net. The probability of scoring on the first shot is 0.5. Each time he scores, the probability of scoring on the next shot goes up by 0.1. What is the probability of scoring on the 3rd shot?



$$\Pr(\text{Score on third shot}) = \Pr(A \text{ or } C \text{ or } E \text{ or } G)$$

$$= \Pr(A) + \Pr(C) + \Pr(E) + \Pr(G)$$

$$= 0.21 + 0.12 + 0.15 + 0.125$$

$$\boxed{= 0.605}$$