Economics 113 Professor Spearot Introduction to Econometrics
Spring 2014 – Final
Name ___________ANSWER KEY_____ ID_____________________________

Final – 100 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 no graphing calculators. No cell phones. Do not use your own scratch paper.

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

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

*Signature*____

Problem 1 (50 Points)

We wish to predict real wage outcomes using the following regression:

$$\log(rw_{it}) = \beta_0 + \beta_1 college_i + \beta_2 female_i + \beta_3 black_i + \alpha_t + u_{it}$$

Here, rw_{it} is the real wage for respondent i interviewed in year t, $college_i$ takes on a value of 1 if respondent i is a college graduate (0 otherwise), $female_i$ takes a value of 1 if respondent i is female (0 otherwise), and $black_i$ takes on a value of 1 if respondent i is black (0 otherwise). The term α_t represents year fixed effects, which are suppressed in the following results:

```
.reg ln_rw college female black i.year

      Source |       SS       df       MS              Number of obs =  598475
-------------+------------------------------           F(  9,598465) = 9167.91
       Model |  25658.4537     9  2850.9393           Prob > F      =  0.0000
    Residual |  186104.335    65  .310969454           R-squared     =  0.1212
-------------+------------------------------           Adj R-squared =  0.1212
       Total |  211762.789    74  .353837908           Root MSE      =  .55765
------------------------------------------------------------------------------
       ln_rw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     college |   .3899047   .0018827   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
      female |  -.2447752    .001444   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       black |  -.1123664   .0024996   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       _cons |   2.781718   .0020319   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
------------------------------------------------------------------------------
```

a.) Please interpret **<u>precisely</u>** the coefficient on college. **(10 Points)**

First, exponentiate the effect and subtract 1.

exp(0.389)-1 = 0.476 +4

<u>Within years</u>, having a <u>college degree increases wages by 47.6%</u> <u>relative to those without a college degree</u>
+2 +2 +2

b.) We wish to test whether there are any interactions between *female* and *black* and *college* using the following specification:

$$\log(rw_{it}) = \beta_0 + \beta_1 college_i + \beta_2 female_i + \beta_3 black_i + \beta_4 college_i \cdot female_i + \beta_5 college_i \cdot black_i + \alpha_t + u_{it}$$

The results from running this regression (again suppressing year estimates) are below:

```
.reg ln_rw college female college_female black college_black i.year

      Source |       SS       df       MS              Number of obs =  598475
-------------+------------------------------           F( 11,598463) = 7504.87
       Model |  25670.1273    11  2333.64794           Prob > F      =  0.0000
    Residual |  186092.661    63  .310950987           R-squared     =  0.1212
-------------+------------------------------           Adj R-squared =  0.1212
       Total |  211762.788    74  .353837908           Root MSE      =  .55763
------------------------------------------------------------------------------
       ln_rw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     college |   .3968689   .0026744   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
      female |  -.2415216   .0015945   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
college_female |  -.0182506   .0037573   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       black |  -.1162495   .0026832   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
 college_black |   .0293265   .0073784   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
        cons |    2.78049   .0020615   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
------------------------------------------------------------------------------
```

Which regression is preferred, the regression in '1a' or the regression here in '1b'? Please test this hypothesis at the 95% level, stating your null and alternative hypotheses. **(10 points)**

H0: B_3=0, B_5=0 +1
HA: H_0 not true +1
q = 2 +0.5
df_{ur}=598463 +0.5
SSR_{ur}=186092 +0.5
SSR_r =186104.335 +0.5

F_{stat} = ((186104-186092)/2)/(186092/598463) = 19.30 +3
F_{crit}= 3 +1

F_{stat}>F_{crit} =>> Reject the null! +2

The interactions between female, black, and college are a jointly significant determinant of the real wage.

c.) Please write the Stata code required to generate `college_female` and `college_black`, and provide a different command than in '1b' to estimate the specification with year fixed effects. **(10 points)**

*gen college_female = college*female* +3
*gen college_black = college*black* +3

xtreg ln_rw college female college_female black college_black, fe i(year)
+4

d.) Does the black-white wage gap depend on whether the respondent is college educated? Test this hypothesis at the 99% level, stating your null and alternative hypothesis. Show your work! **(10 points)**

$H0: B_5=0$ +1
$HA: B_5!=0$ +1

$T_{stat}=(.0293265/.0073784)=3.97$ +3
$T_{crit}=2.575$ +1

$|T_{stat}|>|T_{crit}|$ =>> *reject the null!!* +1

The black-white wage gap is significantly affected by a college education. +3

e.) What is the **precise** difference in predicted wages between a black college-educated male and a white female without a college degree? **(10 points)**

```
BM_C = 2.78049 + 0.3968689 - 0.1162495 + 0.0293265 +1

WF_NC = 2.78049 - 0.2415216                         +1

BM_C - WF_NC = 0.551                                +2
(Taking the difference properly is worth 4 total points.  I don't care how one gets it)

exp(0.551)-1 = 0.735                                +3
```

A black, college educated male makes 73.5% more than white female without a college education.
+3

Problem 2 (50 Points)

a.) We now use our wage panel dataset from 1980-1987 to examine the determinants of annual hours worked:

$$hours_{it} = \beta_0 + \beta_1 educ_i + \beta_2 manu_{it} + \beta_3 union_{it} + \alpha_t + u_{it}$$

Here, $hours_{it}$ is annual hours worked for individual i in year t, $educ_i$ is the time-invariant education level of individual i, $manu_{it}$ equals 1 if individual i works in a manufacturing job in year t (0 otherwise), and $union_{it}$ equals 1 if individual i works in a union job in year t (0 otherwise). Note that manufacturing and union jobs are not mutually exclusive outcomes. Estimating this equation using Pooled OLS, we get the following.

```
. reg hours educ union manuf i.year

      Source |       SS       df       MS              Number of obs =    1200
-------------+------------------------------           F( 10,  1189) =    7.28
       Model |  25891174.6    10  2589117.46           Prob > F      =  0.0000
    Residual |   422789266  1189  355583.908           R-squared     =  0.0577
-------------+------------------------------           Adj R-squared =  0.0498
       Total |   448680441  1199  374212.211           Root MSE      =  596.31
------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   -23.3217   10.21055   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       union |   -56.6942   42.35081   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       manuf |    60.6815   39.21068   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
        year |
       1981  |    162.335   68.86171   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       1982  |    213.217   69.00531   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       1983  |    291.039   68.90454   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       1984  |    315.447   69.11393   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       1985  |    358.572   68.91055   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       1986  |    381.867   68.89213   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       1987  |    454.073   69.04009   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

       _cons |    2230.84   130.5198   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
------------------------------------------------------------------------------
```

Please construct and interpret a 95% confidence interval for the constant in this regression. **(10 Points)**

*2230.84 - 1.96*130.5198 < B_0 < 2230.84 + 1.96*130.5198* *+2*

1975.021 < B_0 < 2486.659

With 95% confidence, a respondent with zero years of education that works in a non-union, non-manufacturing job,

+1 *+3*

worked between 1975.02 and 2486 hours in 1980.

+2 +2

b.) I claim that being in a union has a significant effect on annual hours worked. Using the results in '2a', what is the probability that I'm wrong? **(10 Points)**

t_{stat} = *-56.6942/42.35081=-1.34* *+3*

Pvalue $= Pr(|T|>|t_{stat}|)$
$= Pr(T > |t_{stat}|) + Pr(T < -|t_{stat}|)$
$= 2(1 - Pr(T < |t_{stat}|))$
$= 2(1 - Pr(T < 1.34)) = 2(1 - 0.9099) =$ ***0.1802*** *+7*

c.) Hours worked cannot be negative, though pooled OLS may yield negative values for predictions. What are the two techniques we can use to remedy this issue? **(5 Points)**

Tobit and Poisson +2.5 each

d.) We now augment the regression equation in '2a' to include individual fixed effects, α_i, but removing the time fixed effects.

$$hours_{it} = \beta_0 + \beta_2 manu_{it} + \beta_3 union_{it} + \alpha_i + u_{it}$$

What happened to education, and why? **(5 Points)**

Education does not vary by time within the individual. So, it is absorbed in the fixed effect.
+5 *+5*

e.) After initializing the panel dimension of the dataset, we estimate the model from '2d':

```
. xtreg hours union manuf, fe

Fixed-effects (within) regression               Number of obs      =      1200
Group variable: nr                              Number of groups   =       150

R-sq:  within  = 0.0009                         Obs per group: min =         8
       between = 0.0079                                        avg =       8.0
       overall = 0.0037                                        max =         8

                                                F(2,1048)          =      0.47
corr(u_i, Xb)  = 0.0402                         Prob > F           =    0.6247

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       union |   5.724433   48.58498   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
        manu |   43.92626   46.45788   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       _cons |   2221.753   21.03889   xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
-------------+----------------------------------------------------------------
     sigma_u |  425.22232
     sigma_e |  470.32479
         rho |  .44976438   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(149, 1048) =     6.50           Prob > F = 0.0000
```

Please interpret the coefficient on *manu*, and test whether it is significantly different from zero at the 95% level. Show your work! **(10 points)**

Within individuals, being in a manufacturing job increases annual hours worked by 43.9 relative to non-manufacturing jobs. +3

H_0: $B_2=0$ +1
H_A: $B_2!=0$ +1

$T_{stat}=(43.92626/46.45788)=0.945$ +2
$T_{crit}=1.96$ +1

$|T_{stat}|<|T_{crit}|$ =>> fail to reject the null!! Within individuals, the effect of being in a manufacturing industry on hours worked is insignificant. +2

f.) Again assuming that the panel dataset is already initialized, please write out the code to estimate the following:

$$\Delta hours_{it} = \beta_2 \Delta manu_{it} + \beta_3 \Delta union_{it} + \Delta u_{it}$$

How does the interpretation for the coefficient on *manu* change for this regression relative to 2d?

gen diff_hours = D.hours +1
gen diff_manu = D.manu +1
gen diff_union = D.union +1

reg diff_hours diff_manu diff_union, noconstant
+2 +2

The interpretation is now "in the short run"

+3

Have a great summer!!!

Normal Distribution from -oo to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

TABLE G.3b

5% Critical Values of the *F* Distribution

	Numerator Degrees of Freedom									
	1	2	3	4	5	6	7	8	9	10
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Example: The 5% critical value for numerator $df = 4$ and large denominator df (∞) is 2.37.
Source: This table was generated using the Stata® function invFtail.