Eco671, Spring 2007, Prof. Bill Even
First Midterm Exam

Directions: Put all answers on the paper provided. Use only one side of a sheet of paper and put your answers in order. Be careful to read the questions carefully and provide explanations when requested. Ambiguity in answers will result in unfavorable grades.

1. (25 points) Using data from 1999 Current Population Survey, I estimated regression models of the **log(wage)** [log=natural log] as the dependent variable. The explanatory variables in the regression included the person's age and dummy variables listed below that should be self-explantory. The first specification of the model of does not include education controls. The second specification adds the education dummies.

Dependent variable: log(wage)	Specifcation (1)		Specication (2)	
	Coefficient	t-statistic	Coefficient	t-statistic
Intercept	2.2	72.16	1.82	61.7
age	0.008	15.27	0.005	11.12
female	-0.24	-22.19	-0.24	-24.64
Smoking status (never smoked omitted)				
quit smoking	0.03	1.85	0.05	3.97
currently smoke	-0.08	-6.26	0.02	1.56
Marital status (divorced/separted omitted)				
married	0.23	15.83	0.2	15.09
single	0.11	5.75	0.13	7.31
Race (other race omitted)				
white	-0.03	-1.11	0.02	0.7
black	-0.09	-3.07	0.00	-0.28
Education (less than high school omitted)				
High school graduate			0.29	18.02
Some college			0.42	25.13
Bachelors degree			0.74	41.43
Master's degree			0.88	36.45
Professional degree			0.99	21.06
Doctorate			1.08	21.76
Sample size	10976		10976	
Sum of squared errors	3525.95		2817.81	

a. **Based on specification (2),** other things being the same, how do the **wages** of smokers and quitters compare? (provide a numerical comparison)

b. If specification (1) was reestimated with a dummy for "never smoked" included and the dummy for "quit smoking" excluded,

i.. what would be the value of the coefficient on the intercept? (no explanation required.)

ii. what would be the value of the coefficient on the "never smoked" dummy? (no explanation required.)

c. Comparing specification (1) and (2), it's clear that inclusion of the education controls has a substantial effect on the coefficient for the "currently smoke" dummy. What does this tell you about the relationship between education and smoking? Justify your answer with reference to your knowledge about the source of bias in OLS estimators.

d. If age was measured with error, how would this likely affect the estimated effect of smoking on earnings? Explain any assumptions you must make in justifying your answer.

e. Using specification (2) as the "base model", explain how you could test the null hypothesis that the effect of smoking status on wages is identical for men and women. Be precise. Describe the regressions you would estimate, how the test statistic would be calculated, the distribution of the test statistic (including degrees of freedom), and under what conditions you would reject the null hypothesis.

2. (30 points) In a 1998 article,[1] the effect of children on a range of different measures of women's labor supply was examined. An important concern addressed in the article is that children are endogenous to women's labor supply decision (alternatively, a woman simultaneously chooses the number of children and labor supply.) To correct for the effect of this endogeneity , they used two stage least squares using the sex composition of children as the identifying variables for the labor supply equation. **A summary of the estimated effect of number of children on different labor supply measures is provided in the table below:**

TABLE 8—OLS AND 2SLS ESTIMATES OF LABOR-SUPPLY MODELS USING 1990 CENSUS DATA

	All women			Married women			Husbands of married women		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation method	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Instrument for *More than 2 children*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*
Dependent variable:									
Worked for pay	−0.155	−0.092	−0.092	−0.147	−0.104	−0.104	−0.102	0.017	0.017
	(0.002)	(0.024)	(0.024)	(0.002)	(0.024)	(0.024)	(0.001)	(0.009)	(0.009)
			[0.743]			[0.576]			[0.989]
Weeks worked	−8.71	−5.66	−5.64	−8.25	−5.76	−5.76	−1.03	1.01	1.01
	(0.08)	(1.16)	(1.16)	(0.09)	(1.15)	(1.15)	(0.05)	(0.63)	(0.63)
			[0.391]			[0.670]			[0.708]
Hours/week	−6.80	−4.08	−4.10	−6.39	−3.94	−3.95	−0.06	0.85	0.83
	(0.07)	(0.98)	(0.98)	(0.07)	(0.96)	(0.96)	(0.05)	(0.69)	(0.69)
			[0.489]			[0.665]			[0.180]

Focus on the models for women (specifications 1-6). OLS estimates of the effect of children on labor supply are larger (in absolute terms) than the 2SLS estimates for all three measures of labor supply.

a. Comparing the 2SLS and OLS estimates of the effect of children for women (specifications 1-6), what can you conclude about the nature of the endogeneity problem? Explain.

b. Explain the 2 steps of 2SLS for this particular example. Be precise about the 2 steps using the specifics of this problem (i.e. describe which variables are used in which regressions).

c. Explain why the OLS estimates were biased (what statistical assumption is violated) and how 2SLS eliminates the bias found in the OLS estimates.

d. In this study, specifications 3, 6, and 9 include two instruments: a dummy variable indicating whether the mother's first two children were boys, and another dummy indicating whether the first two children were girls. (All women in the sample had at least two kids). What conditions must these two dummy variables satisfy to be "good" instruments? If it's possible to test whether these instruments satisfy these conditions, describe the relevant test.

---

[1] "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." Joshua D. Angrist, William N. Evans. *The American Economic Review*, Vol. 88, No. 3. (Jun., 1998), pp. 450-477.

Eco671, Spring 2007, Prof. Bill Even
First Midterm Exam

3. (25 points) The model below is a probit model explaining whether a person smokes cigarettes as a function of age, marital status, sex, and years of education. Refer to the estimates to answer the questions that follow.

```
Probit estimates                                  Number of obs   =       7023
                                                  LR chi2(4)      =     252.79
                                                  Prob > chi2     =     0.0000
Log likelihood = -4172.2257                       Pseudo R2       =     0.0294

------------------------------------------------------------------------------
     smoknow |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0247233   .0033506    -7.38   0.000    -.0312904   -.0181562
     married |  -.3855384   .0475146    -8.11   0.000    -.4786654   -.2924114
        male |   .2318823   .0535697     4.33   0.000     .1268876     .336877
     educyrs |   -.049293   .0045926   -10.73   0.000    -.0582943   -.0402917
       _cons |   1.521375   .2031785     7.49   0.000     1.123153    1.919598
------------------------------------------------------------------------------
```

a. What is the probability that a 60 year old married male with 12 years of education smokes? Provide a brief description of how you derived your answer.

b. What is the marginal effect of an additional year of education on the probability of smoking for the person described in (a)? Provide a brief description of how you derived your answer.

c. Suppose that you want to test whether marriage has a different effect on the probability of smoking for men and women. Explain how you could test this hypothesis. What regression(s) would you estimate, how would you calculate a relevant test statistic, what's its distribution, and under what conditions would you reject the null hypothesis that the effect of marriage is identical for men and women?

Eco671, Spring 2007, Prof. Bill Even
First Midterm Exam

4. (20 points) A multinomial logit model of a worker's pension plan type was estimated. The four possible outcomes are labeled: 1: defined benefit (DB) plan; 2=defined contribution (DC) plan; 3=both a DB and DC plan; 4=no pension plan. As can be seen below, outcome 4 was chosen as the "reference" outcome and coefficients on outcome 4 are normalized to zero. The explanatory variables are defined as follows: **fs100=1** if there are more than 100 employees at the firm and fs100=0 otherwise; **annual** is the person's annual income in 10,000s of dollars; the age variables are a series of age dummies representing the person's age (e.g. age2630=1 implies the person is between 26 and 30 in age). **Female** is a dummy indicating that the worker is female. The sample includes workers aged 21 to 54.

```
Multinomial regression                          Number of obs   =      18337
Log likelihood = -20154.321                     Pseudo R2       =     0.0968
------------------------------------------------------------------------------
    plantype |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1  (DB only) |
       fs100 |   1.632133   .0550525    29.65   0.000     1.524233    1.740034
      annual |   .0030319   .0016961     1.79   0.074    -.0002923    .0063561
     age2630 |   .6274585   .1347746     4.66   0.000     .3633051    .8916119
     age3135 |   .9304695   .1284837     7.24   0.000     .6786462    1.182293
     age3640 |   1.312306   .1247968    10.52   0.000     1.067709    1.556903
     age4145 |   1.594885   .1244391    12.82   0.000     1.350989    1.838781
     age4650 |   1.608275   .1256892    12.80   0.000     1.361928    1.854621
     age5154 |   1.991174   .1307112    15.23   0.000     1.734985    2.247364
      female |  -.4799718   .0495038    -9.70   0.000    -.5769975   -.3829461
       _cons |   -3.33853   .1216725   -27.44   0.000    -3.577003   -3.100056
-------------+----------------------------------------------------------------
2  (DC only)           |
       fs100 |   1.477032   .0400898    36.84   0.000     1.398458    1.555607
      annual |   .0063577   .0013934     4.56   0.000     .0036267    .0090888
     age2630 |   .7719201   .0880759     8.76   0.000     .5992945    .9445456
     age3135 |   .9774938   .0853748    11.45   0.000     .8101622    1.144825
     age3640 |   1.064954   .0851808    12.50   0.000     .8980022    1.231905
     age4145 |   1.223434   .0859084    14.24   0.000     1.055057    1.391812
     age4650 |    1.15493   .0877542    13.16   0.000     .9829353    1.326925
     age5154 |   1.199215   .0968637    12.38   0.000     1.009366    1.389064
      female |  -.5988376   .0384591   -15.57   0.000     -.674216   -.5234592
       _cons |   -2.16179    .080363   -26.90   0.000    -2.319299   -2.004282
-------------+----------------------------------------------------------------
3  (both DB and DC)
       fs100 |   2.538268   .0762499    33.29   0.000     2.388821    2.687715
      annual |   .0044632   .0015726     2.84   0.005     .0013809    .0075455
     age2630 |    1.12633   .1920817     5.86   0.000     .7498565    1.502803
     age3135 |   1.881973   .1808749    10.40   0.000     1.527465    2.236482
     age3640 |   2.044699   .1797938    11.37   0.000      1.69231    2.397088
     age4145 |     2.4131   .1792045    13.47   0.000     2.061865    2.764334
     age4650 |   2.350956   .1806604    13.01   0.000     1.996868    2.705044
     age5154 |    2.76251   .1843167    14.99   0.000     2.401256    3.123764
      female |  -.7683511   .0542164   -14.17   0.000    -.8746134   -.6620889
       _cons |  -4.879732   .1838243   -26.55   0.000    -5.240021   -4.519443
------------------------------------------------------------------------------
Outcome plantype==4 (No plan) is the comparison group
```

a. Compute the probability that a 40 year old male worker at a firm with less than 100 employees has
   i. no pension
   ii. only a DC plan
   Provide a general idea of how you derived your answers.
b. If the reference outcome were switched from plantype=4 to plantype=2, what would the coefficient on female be for plantype=4? for plantype=3? Give a **brief** description of how you derived your answer.
c. Calculate the marginal effect of an additional $10,000 of income on the probability of having only a DC plan.
d. Is the IIA assumption likely to be satisfied for this model? Why or why not?