Eco671 Final Exam, Spring 2007
Professor Bill Even

Directions: Record all your answers on the paper provided. Use only one side of each sheet of paper and leave the margins open so that I have room to comment. You have until 3:25 to complete the exam. No extra time will be given.

In the event that you feel a question lacks all the necessary information to provide a correct answer, indicate any assumptions that you make in formulating your answer.

1. (20 points) Several years ago, there was a proposal that would have required that all employers provide a pension plan for their workers (a "universal savings account"). The objective was to improve savings rate in the U.S. Some economists estimated the effect of such a mandate on the level of saving in the U.S. To do this, they estimated a pension savings equation such as:

$S_i = X_i b + e_i$

where S_i is the dollar amount saved by person i, X_i is a vector of characteristics describing person i (e.g. age, education, income), b is the coefficient vector, and e_i is an error term that meets the classical assumptions. For the purposes of this question, **ignore the fact that S_i is truncated at zero.**

The above equation was estimated using a group of people who currently are covered by a pension. Presumably, this is not a random sample of the population because people have some control over where they work and therefore, whether they have pension coverage or not.

a. Suppose that unobservable characteristics that lead a person to be a "saver" make it more likely that he matches with an employer that provides a pension. Also, suppose that higher income workers save more and that income is one of the control variables in the savings equation. If the savings equation described above was estimated with OLS, would the estimated effects of income on pension saving be over- or under-estimated? Justify your answer.

b. Suppose the OLS model of savings (estimated with workers who currently have a pension) is used to estimate how much workers would save if all employers were required to provide a pension. Would the estimates based on the OLS model likely be an over-statement or an under-statement of how much these newly covered workers would save? Explain.

c. Explain how you could improve your prediction of how much workers who currently don't have coverage would save if their employers were forced to offer pension coverage**. Be precise in the details** describing the model you would estimate and how you would use the estimates from that model to predict saving for those who currently do not have pension coverage.

Eco671 Final Exam, Spring 2007
Professor Bill Even

2. (25 points) Taylor et. al (2006)[1] examine the determinants of research productivity among economists. Using data on publications collected in a survey, they construct a measure of "quality adjusted pages of publications per year" (call it PAGES) for each person in the survey. The publications were counted over the period 1998 through 2002.

Because the dependent variable was truncated at zero, a tobit model was used to analyze the data. The explanatory variables and resulting tobit coefficients are listed in the table below.

Table 2. Tobit Coefficients by Highest De

	Overall $n = 714$ (1)	Assumed value for (a)
Personal		
Gender	5.06** (2.25)[a]	1 (implies male)
Average coauthors	24.33** (9.12)	1
Average coauthors squared	−6.63** (6.35)	1
Years to Ph.D.	−0.61* (1.44)	5
Review	1.50* (1.58)	0
Presentations	2.31** (5.25)	0
Experience	−0.30** (2.90)	5
Institutional		
Degree: masters	−2.74 (1.03)	0
Degree: doctorate	6.03* (1.54)	0
Peers who publish	0.14** (3.36)	0
Summer stipend[b]	7.76** (3.07)	0
Teaching		
Teaching hours	−0.51** (2.87)	15
Summer hours	−0.96** (2.35)	0
Service		
Committee	−1.09** (1.80)	0
Committee chair	−2.75** (2.52)	0
Department chair[b]	−7.65** (2.83)	0
Log likelihood	−2036.9	

[1] Taylor, Susan Washburn, Fender, Blakely Fox, and Kimberly Gladden Burke. 2006. "Unraveling the Academic Productivity of Economists: The Opportunity Costs of Teaching and Service." *Southern Economic Journal* 72, no. 4: 846-859.

In the right-most column of the above table, I have provided a vector of characteristics for a hypothetical male (call him Jack). Using the Tobit coefficients above (**and assume that there is an intercept in the tobit equal to -8.0 and that the standard deviation of the regression residual is 2 -- these parameters are not provided in the above table**), estimate each of the following and provide a brief outline of how you performed the calcuation:
a. The probability that Jack has more than zero pages published.
b. The predicted number of pages that Jack will publish.
c. The predicted number of pages that Jack will publish, conditional on publishing something.

This past year, the Miami economics department reduced its standard teaching load by 3 credit hours per year. Calculate the expected effect of this reduced teaching load on
d.. the probability that Jack publishes something.
e. the predicted number of pages that Jack will publish.

f. Suppose that faculty have unobservable characteristics that cause them to differ in terms of their research productivity. Also, suppose that the faculty with unobservable characteristics that cause them to be unusually productive in publishing avoid committee work and are more likely to receive a summer stipend. Explain how this could bias the estimated effects of committee work and summer stipends in the above tobit model and the direction of the bias. (Note: you can use the same logic that we used for linear models in describing the source and direction of the bias.)

Eco671 Final Exam, Spring 2007
Professor Bill Even

3. (25 points) I used March 2005 CPS data to estimate a log-wage (natural log) equation as a function of a person's years of schooling and its square, age and its square, a dummy indicating whether the person is female, and a dummy indicating whether the person is a union member. Lambda represents the correction factors included in the second stage of the treatment effects model. **The model is estimated as a "treatment effects" model** using the two-step process discussed in class. The treatment is union membership. The top half of the table reveals the coefficients for the log-wage equation. The bottom half of the table reveals the coefficients for the probit model of union membership that are used to generate the relevant values of lambda.

The rationale for estimating the wage equation with union membership as a treatment effects model is that unionism is not randomly determined. Workers have control over where they work and thus unobservables that influence wages could also influence whether they seek or avoid a unionized job

	(1)	(2)	(3)	(4)
	Treatment Effects			
	Coefficient	t-statistic	X	X*b
Intercept	1.06	16.37	1	1.06
Years of schooling	-0.02	-2.31	12	-0.24
(Years of schooling)^2	0	9.85	144	0.5
age	0.05	26.8	30	1.57
age^2	0	-23.4	900	-0.41
female	-0.17	-16.51	1	-0.17
union	0.35	6.5	--	--
lambda	-0.08	-2.63		
	Probit model for union membership			
Intercept	-4.49	-11.35	1	-4.49
Years of schooling	0.17	3.61	12	2.05
(Years of schooling)^2	0	-2.81	144	-0.73
age	0.1	12.11	30	3.07
age^2	0	-10.45	900	-0.94
female	-0.39	-10.64	0	0

In column (3), I present the characteristics of a person (call her **Sally**) as a vector. In column (4), I present the value of Sally's characteristics times the corresponding coefficients.

a. Based on the information provided, is there positive or negative selection into unions? How did you conclude this and what does this tell you about the unobservable characteristics of such people? Give an example of an unobservable characteristic that might fit this situation.

b. If the log-wage equation had been estimated by OLS instead of the treatment effects model, would you expect that the coefficient on unionism would be larger or smaller than estimated in the treatment effects model? Justify your answer.

c. Suppose that you have no information on whether Sally is a union member. Provide the best prediction of her log-wage based on the characteristics given in the table above if you assume that she would not be in a union. Provide a brief description of your calculations.

d. Suppose that you have been told that Sally is in a union. Provide the best prediction of her log-wage based on the characteristics given in the table above and the knowledge that she joined a union. Provide a brief description of your calculations.

e. Suppose Sally and Sue are identical in all respects, except that you've been told that Sally joined a union and Sue did not. What is your best prediction of the difference between their wages? Provide a brief description of your calculations.

f. What is your best estimate of the effect of union membership on the log-wage for Sally or Sue? Explain.

4. (30 points) In a recent article, Yankow (2006)[2] examines why workers earn more in big cities. There are numerous hypotheses for this fact including (1) big cities must pay more as a compensating difference for negative attributes like crime and pollution, (2) big cities must pay more to compensate for higher housing costs, or (3) "better" workers are attracted to big cities.

To examine this question, Janklow estimates wage equations using the National Longitudinal Survey. He controls for the standard set of worker characteristics (e.g. education, experience, race, experience, tenure with current employer, etc.) and also includes dummy variables indicating whether the worker lives in a big city or a small city (reference group is outside of a city). The results of the regressions are presented in table 2 below.

Specifications (1) through (6) are all OLS models of the log-wage rate. With the smallest number of controls for worker characteristics, the big-city coefficient is .220. It falls to .190 when education (measured as highest graded attended -- HGA) is added in column (2). It falls even further as the additional controls are added in columns (3) through (6). [AFQT=score on Air Force Qualifying Test (AFQT) which is a measure of IQ; dummy for being married; years of tenure with current employer; dummy variable for union membership; and dummy variable for government employee.]

a. Given that the coefficient on "big city" shrinks as these other controls are added, what does this say about the types of workers that are attracted to big cities in terms of education, AFQT, etc.? Justify your answer.

b. When individual fixed effects are added, the coefficiint on "big city" drops from .176 (column 5) to .054 (column 8). What does this say about the type of workers that are drawn to big cities? Justify your conclusion.

c. When individual fixed effects are added, the coefficieint on government rises from -.050 (col. 6) to -.000 (col 8). What does this say about the type of workers that are drawn to government jobs? Justify your conclusion.

Suppose someone suggests that you estimate this model as a random effects model instead of a fixed effects model.

d. What are the advantages of a random effects model as compared to a fixed model?
e. What is the potential disadvantage of a random effects model?
f. How could you test whether the random effects model would be "better" than the fixed effects model?

g. Define the vector X as the list of controls used in specification (8) and b as the corresponding vector of coefficients. Notice that this excludes some of the variables in the earlier OLS specifications. Explain how you would predict the wage rate for person i with characteristics X_i using the fixed effects estimates. [Be careful ... remember that the fixed effects model washes out the individual specific intercepts.]

---

[2] Yankow, Jeffrey J. 2006. "Why Do Cities Pay More? An Empirical Examination of Some Competing Theories of the Urban Wage Premium." *Journal of Urban Economics* 60, no. 2: 139-161.

Table 2
Effect of urban residence on wages

Variable	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)	Fixed (7)	Fixed (8)	Exper. (9)	Exper. (10)
Big City	0.220[a]	0.190[a]	0.187[a]	0.187[a]	0.176[a]	0.175[a]	0.050[b]	0.054[a]	0.049[b]	0.052[a]
	(0.017)	(0.015)	(0.015)	(0.012)	(0.012)	(0.012)	(0.022)	(0.020)	(0.022)	(0.020)
Small City	0.095[a]	0.069[a]	0.075[a]	0.082[a]	0.078[a]	0.076[a]	0.033	0.038[c]	0.026	0.032
	(0.019)	(0.018)	(0.017)	(0.014)	(0.014)	(0.014)	(0.023)	(0.021)	(0.023)	(0.021)
Black	−0.300[a]	−0.253[a]	−0.163[a]	−0.087[a]	−0.098[a]	−0.101[a]				
	(0.016)	(0.014)	(0.016)	(0.013)	(0.013)	(0.013)				
Hispanic	−0.161[a]	−0.064[a]	−0.019	−0.001	−0.005	−0.006				
	(0.017)	(0.017)	(0.017)	(0.014)	(0.014)	(0.014)				
Exper.	0.012[a]	0.042[a]	0.042[a]	0.020[a]	0.019[a]	0.020[a]	0.065[a]	0.044[a]		
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)		
(Exper. sq.)/10	−0.012[a]	−0.012[a]	−0.013[a]	−0.007[a]	−0.007[a]	−0.007[a]	−0.022[a]	−0.016[a]		
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)		
HGA		0.085[a]	0.058[a]	0.044[a]	0.042[a]	0.041[a]				
		(0.003)	(0.004)	(0.003)	(0.003)	(0.003)				
AFQT			0.004[a]	0.003[a]	0.003[a]	0.003[a]				
			(0.000)	(0.000)	(0.000)	(0.000)				
AFQT miss.			0.008	0.017	0.017	0.019				
			(0.030)	(0.026)	(0.025)	(0.025)				
Married				0.121[a]	0.117[a]	0.115[a]		0.053[a]		0.054[a]
				(0.009)	(0.009)	(0.009)		(0.008)		(0.008)
Tenure				0.060[a]	0.056[a]	0.054[a]		0.039[a]		0.025[a]
				(0.003)	(0.003)	(0.003)		(0.003)		(0.001)
(Tenure sq.)/10				−0.027[a]	−0.026[a]	−0.025[a]		−0.020[a]		−0.029[a]
				(0.002)	(0.002)	(0.002)		(0.002)		(0.003)
Govt.				−0.023[b]	−0.050[a]	−0.050[a]		−0.000		0.007
				(0.012)	(0.011)	(0.011)		(0.009)		(0.009)
Union					0.184[a]	0.175[a]		0.119[a]		0.118[a]
					(0.010)	(0.010)		(0.009)		(0.009)
Union miss.					0.032[b]	0.039[a]		0.034[a]		0.032[a]
					(0.015)	(0.015)		(0.012)		(0.012)
Industry	No	No	No	Yes	Yes	Yes	No	Yes	No	Yes
Occupation	No	No	No	Yes	Yes	Yes	No	Yes	No	Yes
Firm size	No	No	No	No	No	Yes	No	No	No	No
R^2	0.16	0.26	0.28	0.41	0.43	0.43	0.13	0.20	0.07	0.14

Notes. Standard errors are in parentheses. Dependent variable is log of regional CPI-deflated hourly wage. *Big City* refers to residence in an MSA with population > 1 million residents, *Small City* refers to residence in an MSA with population > 250,000 but less than 1 million residents. All regressions include year dummies. Standard errors are corrected for heteroskedasticity and within-person correlation.

[a] Significant at the 1% level.
[b] Idem, 5%.
[c] Idem, 10%.