AAEC6311 – Applied Econometrics II

Spring 2010

Instructor: Eric Belasco

Exam 1

Name: _____

You have one hour and 20 minutes to complete the exam. There are three questions. Point values will sum to 70, with 5 points associated with each question. **USE YOUR TIME WISELY**. Also, I will address clarifying questions, but not substantive questions. You are allowed to use one 3x5 note sheet (front and back), a calculator, and a pencil/pen. Good luck!

Question 1 (25 points). The new research initiative from the USDA provides you with funding to evaluate the impact of taxes on smoking. You collect state level data from 1995. You assume that the number of cigarette packs per capita (y) can be distributed as a gamma distribution,

$$f(y_i|\lambda_i) = \frac{y_i}{\lambda_i^2}\exp\left(-\frac{y_i}{\lambda_i}\right)$$

where $\lambda_i = \exp(x_i\beta)\ /2$. In this case, you include real cigarette prices (*rprice*), real income (*rincome*), and real cigarette taxes (*ctax*) as independent variables (all of which are averages at the state level).

(a) (5 points) Derive the log-likelihood function necessary for computing the Maximum Likelihood Estimates for β.

(b) (5 points) Can you solve for an explicit expression of $\hat{\beta}_{MLE}$ in terms of x and y? Are you still able to estimate β if an explicit expression doesn't exist? (*Hint: An explicit expression of y, in terms of x is* $y = f(x) = x^2 + 4$)

Table 1. Gamma regression results

	Estimate	SE	t-value
constant	5.735	0.743	7.722
rincome	0.068	0.051	1.317
rprice	-0.020	0.008	-2.341
rctax	-0.037	0.045	-0.823
Likelihood Ratio Index (LRI)	0.039		
LL	260.181		

Note: $\frac{1}{N}\sum_{i=1}^{N} \exp(x_i\beta) = 68.651$

(c) (5 points) One can show that according to a gamma distribution,

$$E[y_i|x_i] = \exp(x_i\beta)$$

Given this specification, we can evaluate the marginal impact of x on y. Using table 1 above, interpret the marginal impact of an increase in the independent variables on the predicted value of y (at the mean predicted value of y). Are these results inline with your expectations?

(d) (5 points) What is the impact from a proposed tax on cigarettes of $1.50 per pack on the amount of cigarettes consumed per capita?

(e) (5 points) Given the results from the LRI and Log likelihood value (LL) from the model described above, derive a joint hypothesis test of statistical significance for the three independent variables (H_0: $\beta_1 = \beta_2 = \beta_3 = 0$). What test do you use? How do you conclude statistical significance? **Be as specific as possible.**

Question 2 (35 points). You estimate the impact of demographic factors on the participation in the work force. To this end you use three models currently in your econometric toolbox (Linear Probability Model, Logit, and Probit). You use the following:

$$\text{inlf}_i = x_i\beta + \varepsilon_i$$

where *inlf* is equal to 1 if the individual is in the workforce; and 0 otherwise. You collect data resulting in 753 female individuals with the following variables of interest:

Educ (number of years in education, eg, high school graduate = 12),
Exper (Years of work experience),
Age (in years),
Kidslt6 (number of kids under the age of 6 years old).

Table 2. Binary choice model results.

Dependent variable: inlf

	LPM		Logit		Probit	
	Est	t-value	Est	t-value	Est	t-value
constant	0.679	4.885	0.918	1.219	0.563	1.257
Educ	0.032	4.574	0.181	4.519	0.108	4.595
Exper	0.040	7.086	0.208	6.538	0.125	6.704
Exper^2	-0.001	-3.170	-0.003	-3.030	-0.002	-3.080
Age	-0.018	-7.762	-0.097	-7.200	-0.058	-7.443
Kidslt6	-0.269	-8.051	-1.446	-7.228	-0.871	-7.483
AIC			822.65		821.39	
Pseudo R-squared	0.264		0.220		0.221	

Note: $\sum_{i=1}^{N} \Phi(x_i\beta) = 0.303$ and $\sum_{i=1}^{N} \Lambda(\mathrm{x_i}\beta) = 0.225$.

(a) (5 points) Based on AIC and Pseudo-R-squared measures which model is preferred between the Probit and Logit models? Do the two measures agree on a superior model?

(b) (5 points) Interpret the impact from an additional year of education for the average individual in the sample based on the results from the LPM, Logit, and Probit results shown above.

(c) (5 points) An individual asks for your advise. Upon graduating with an undergraduate degree (Educ =16), a female collegue is interested in one more year of education. This individual has no kids, no work experience, and is 22 years old. What is the marginal impact on the probability of her finding a job after one more year of schooling based on the three models. Two helpful results to have in your analysis are

$$x_{col}\hat{\beta}_{logit} = 1.683$$
$$x_{col}\hat{\beta}_{probit} = 1.012$$

where x_{col} is a vector with elements (1,16,0,0,22,0) and $\hat{\beta}_{probit}$ ($\hat{\beta}_{logit}$) contains the estimated parameters from the Probit (Logit) results in Table 2.

(d) (5 points) Which result (LPM, Logit, or Probit) would you advise your collegue is the most accurate information regarding her situation? Justify your answer.

In order to obtain a more complete evaluation of the data, you include lhours, $lhours = \log(1 + hours)$, where *hours* is the number of weekly hours of labor. Results are shown below.

Table 3. Outcome equation results for the Tobit and Heckman models

Dependent variable: lhours

	Tobit		Heckman (Outcome Eqn)	
	Est	t-value	Est	t-value
constant	0.518	2.154	8.108	14.193
Educ	0.058	4.695	-0.090	-2.077
Exper	0.074	7.206	0.013	0.252
$Exper^2$	-0.001	-3.594	0.000	-0.288
Age	-0.032	-7.681	0.008	0.367
Kidslt6	-0.498	-7.683	-0.118	-0.345

(e) (5 points) Given that not all observations participate in the labor market, you will have a censored dataset. Explain censoring and discuss the issues around using OLS to estimate this relationship.

(f) (5 points) You decide the run the Tobit and Heckman models in order to characterize the relationship between demographic variables and the number of hours provided to the labor market. Discuss the major differences in assumptions between these two models.

(g) (5 points) Recall that the Heckman model uses a two-step process where the Probit results in Table 2 are essentially used to characterize participation in the labor market, then for those that are in the labor market, the parameter estimates provided in Table 3 are estimated. One questionable result is that around the impact on education. First, the estimate from the Tobit (0.058) and Heckman (-0.090) are opposite signs. Second, the Probit portion of the Heckman model (0.108) is opposite in sign to the Outcome equation. Explain why this might be?

Question 3 (10 points). Assume you obtain data on modes of transportation to work for the metropolitian area of San Francisco. Modes of transportation include car, train, and bus. If we assume the basic random utility model, we can assume that individuals maximize the following utility function:

$$U_{ij} = X_{ij}\beta + \varepsilon_{ij}$$

where X contains information about the costs and transportation time for each travel model. City officials are funding research to evaluate the impact from a new subway system in the city.

(a) (5 points) We know that the conditional Logit model imposes the Independence from Irrelevant Alternatives (IIA) assumption. Define this assumption as specifically as possible

(b) (5 points) Discuss the IIA assumption as it pertains to this situation where a new mode of transportation is added to this mix of current alternatives. Is this assumption realistic? Explain why or why not.