ECO671, Spring 2005, Midterm 2
Prof. Bill Even

1. (25 points) Knapp and Seakes (1994)[1] examine the factors that influence whether a student defaults on a college student loan. Below is a summary of 4 probit models that they estimated where the dependent variable is one if the student defaults on the loan and is zero otherwise. **Family status** is a dummy that equals one if the student is from a two-parent family; **graduation** is a dummy that equals one if the person finished his/her degree. The other variable names are sufficiently descriptive for your task here.



TABLE 2.—PROBIT EQUATION COEFFICIENT ESTIMATES AND *t* STATISTICS

| | Equation Number | | | | |
|---|---|---|---|---|---|
| Variable | 1 | 2 | 3 | 4 | 5 |
| Constant | −0.608<br>(−5.587) | −1.005<br>(−1.942) | −0.973<br>(−4.816) | −1.146<br>(−3.915) | −1.375<br>(−2.281) |
| Family Status | −0.259<br>(−2.251) | −0.203<br>(−1.701) | −0.256<br>(−2.216) | −0.258<br>(−2.217) | −0.208<br>(−1.737) |
| Graduation | −0.943<br>( 9.186) | −1.021<br>( 9.249) | −1.045<br>( 8.994) | −1.035<br>( 8.823) | −1.090<br>( 8.964) |
| Parent's Income | −0.016<br>(−3.724) | −0.017<br>(−3.785) | −0.017<br>(−3.886) | −0.017<br>(−3.806) | −0.018<br>(−3.836) |
| Race: Black | 0.902<br>(7.865) | 0.820<br>(5.733) | 0.941<br>(8.069) | 0.857<br>(6.894) | 0.834<br>(5.800) |
| Gender: Female | — | — | −0.057<br>(−0.602) | −0.078<br>(−0.794) | 0.089<br>(−0.872) |
| Loan Amount | — | — | 0.020<br>(1.090) | 0.024<br>(1.221) | 0.026<br>(1.283) |
| Total College Cost | — | — | 0.053<br>(1.984) | 0.111<br>(2.442) | 0.085<br>(1.381) |
| Type College: Two year | — | — | — | 0.018<br>(0.118) | — |
| Type College: Private | | | | −0.350<br>(−1.745) | — |
| School Size | — | — | — | −0.015<br>(−0.879) | — |
| School dummies | no | yes | no | no | yes |
| log(*L*) | −436.00 | −423.28 | −432.62 | −430.37 | −420.84 |
| *K* | 5 | 30 | 8 | 11 | 33 |
| $\chi^2(K-1)$ | 283.07 | 308.50 | 289.82 | 294.32 | 313.39 |

a. Using specification (1), estimate the probability of default on a student loan for a person who has graduated, comes from a one-parent family (family status=0), whose race is not black, and whose parent's income is $50,000 **(note: parent's income is measured in $1000s of dollars in this problem --- so $50,000 is measured as 50 in the regression).** Show the basic steps involved in your calculation.

b. For the person described in (a), what is the marginal effect of an extra $1000 of parental income on the probability of default? Show the basic steps involved in your calculation.

c. Notice that at the bottom of the table there is a row indicating whether "school dummies" are included. This indicates whether the specification includes a set of dummies indicating which of 26 different schools the student attended. Also, K is the number of explantory variables in the specification, and log(L) is the log of the likelihood function. Based on the information provided, **construct a test-statistic of the null hypothesis that there is no difference in the probabilty of default across the 26 schools the students attended.** Show how you construct your statistic, describe its distribution, and whether you would reject the null at the .05 level.

---

[1] Laura Greene Knapp and Terry G. Seaks. "An Analysis of the Probabilty of Default on Federally Guaranteed Student Loans." *Review of Economics and Statistics*, August 1992.

d. Suppose that you want to test that all coefficients in specification (1) above are identical for men and women. Explain how you would test this hypothesis. **Describe the models you would estimate, how you would construct the test statistic, the distribution of the test statistic (including degrees of freedom) and under what conditions you would reject the null hypothesis.**

---

2. (30 points) Using data from the March 2001 Current Population Survey, I constructed an indicator of employment status:

emp=1 implies the person is employed.
emp=2 implies the person is unemployed.
emp=3 implies the person is out of the labor force.

I estimated a multinomial logit model of employment status as a function of age, sex, and education. Female is a dummy variable that equals one if the person is female (0 otherwise). The four education dummies are as follows:

ed1=1 if less than 12 years of education; 0 otherwise.
ed2=1 if 12 years of education; 0 otherwise.
ed3=1 if 13-15 years of education; 0 otherwise.
ed4=1 if 16 or more years of education; 0 otherwise.

The results of the multinomial logit model are as follows:

| | Coefficient | t-statistic |
|---|---|---|
| empl=2 | | |
| female | 0.25 | 2.28 |
| age | -0.01 | -2.24 |
| ed2 | -0.83 | -5.05 |
| ed3 | -0.79 | -4.93 |
| ed4 | -0.96 | -5.72 |
| _cons | -1.63 | -9.11 |
| empl=3 | | |
| female | 0.67 | 9.57 |
| age | -0.01 | -6.21 |
| ed2 | -1.45 | -15.91 |
| ed3 | -1.64 | -17.75 |
| ed4 | -2.55 | -21.72 |
| _cons | 0.24 | 2.3 |

**The coefficients for empl=1 are normalized to zero.**

a. Using the above results, estimate the probability that a 40 year old male is employed. Briefly describe how you derived your answer.

b. Based on the above output, provide an interpretation of the female coefficient for empl=3 in terms of the implications it has for the probability of a particular employment outcome. Be as precise as possible in your interpretation.

c. If the coefficients for empl=2 were normalized to zero, what would the coefficient on female be for
   i. empl=1
   ii. empl=3

No need to provide an explanation. Just provide the answer.

ECO671, Spring 2005, Midterm 2
Prof. Bill Even

After estimating the multinomial logit model, I ran mfx to calculate marginal effects. The commands and results are listed below:

```
mfx, predict(p outcome(1));
    y  = Pr(empl==1) (predict, p outcome(1))
      = .776

variable |                dy/dx
female*|          -.096
age |              .002
ed2*|              .188
ed3*|              .203
ed4*|              .273

mfx, predict(p outcome(2));
          y  = Pr(empl==2) (predict, p outcome(2))
             =  .05988267
variable |                dy/dx
female             .007
age |              -.001
ed2*|              -.030
ed3*|              -.027
ed4*|              -.029

Marginal effects after mlogit
    y  = Pr(empl==3) (predict, p outcome(3))
      =  .16432447
variable          dy/dx
female            .089
age               -.002
ed2               -.158
ed3               -.175
ed4               -.243
```

d. The numbers .776 and -.096 are highlighted in the above output. Provide an interpretation of each number in terms of the implications for the probability of being in a particular employment category (i.e. employed, unemployed, or out of the labor force). Be as precise as possible in your interpretation.

e. Compared to a high school drop out, how much more or less likely is a college graduate to be:
   i.. out of the labor force?
   ii.. unemployed?
   No explanation required here. Just give the numbers.

ECO671, Spring 2005, Midterm 2
Prof. Bill Even

3. (30 points) In a 1998 working paper, Glenn Platt and Mark McBride examined the determinants of franchising. When a firm establishes a franchise, it sells the rights to operating one of its outlets to another owner/operator. The dependent variable in one of their models is the percentage of a firm's outlets that are franchised. **The dependent variable ranges between 0 and 1.** For simplicity, however, ignore the issue of truncation at 1 (relatively few firms are at the upper limit anyway). The independent variables are defined as follows:
Agefirm is age of the firm.
Agefran is years since the firm started franchising.
EMFT is number of full time employees per outlet.
EMPT is number of part time employees per outlet.
Royal2 is the royalty rate charged to franchises.
Nonroyal is a dummy if they don't charge a royalty.
Totinv is the total investment to open an outlet.
Dispers is the number of states with outlets.
Dense is the percentage of a firm's outlets in the state that has the most outlets.
Finaid1 is a dummy if the firm offers financial aid to potential franchisers.
Pass is a dummy if the firm allows passive ownership (meaning that the owner does not have to be the on-site manager).

The table below gives both ols and tobit coefficient estimates (b), t-statistics (t), the mean of the independent variables (xbar), and the product of xbar and b (xbar*b). At the bottom of the xbar*b column, the sum of the column is reported. The standard deviation of the error in the tobit equation is also reported below the tobit coefficients.

| | OLS | | | | Tobit | | |
|---|---|---|---|---|---|---|---|
| | b | t | xbar | xbar*b | b | t | xbar*b |
| Constant | 0.769 | 36.629 | 1.000 | 0.769 | 0.777 | 27.628 | 0.777 |
| AGEFIRM | -0.004 | -10.503 | 17.810 | -0.070 | -0.005 | -10.042 | -0.090 |
| AGEFRAN | 0.038 | 21.465 | 11.450 | 0.435 | 0.049 | 19.598 | 0.559 |
| $AGEFRAN^2$ | -0.001 | -14.669 | 229.500 | -0.252 | -0.001 | -13.128 | -0.329 |
| $AGEFRAN^3$ | 0.000 | 11.188 | 6792.000 | 0.062 | 0.000 | 9.784 | 0.083 |
| EMFT | -0.003 | -6.532 | 4.672 | -0.014 | -0.004 | -6.011 | -0.016 |
| EMPT | -0.003 | -6.291 | 5.005 | -0.014 | -0.003 | -5.907 | -0.017 |
| ROYAL2 | -0.003 | -2.528 | 4.971 | -0.013 | -0.003 | -2.268 | -0.015 |
| NONROYAL | -0.008 | -0.578 | 0.087 | 0.000 | 0.016 | 0.814 | 0.001 |
| TOTINV/1000 | -0.004 | -1.755 | 0.273 | 0.000 | -0.004 | -1.762 | -0.001 |
| DISPERS | 0.000 | 0.958 | 13.380 | 0.005 | 0.000 | 1.492 | 0.010 |
| DENSE1 | -0.214 | -14.044 | 0.488 | -0.104 | -0.248 | -12.168 | -0.121 |
| FINAID1 | 0.037 | 3.373 | 0.141 | 0.005 | 0.038 | 2.538 | 0.005 |
| PASS | 0.005 | 0.595 | 0.725 | 0.003 | 0.006 | 0.584 | 0.004 |
| sum | -- | -- | -- | 0.81 | -- | -- | 0.85 |
| std deviation of error | | | | | 0.31 | | |

a. Compare the ols and tobit coefficients for the variables above. Is there any consistent pattern in terms of the relative size of the coefficients? Why should you have expected this pattern?
b. Based on the ols estimates, what is the expected fraction of outlets that are franchised for the firm with average characteristics? Explain your work.
c. Based on the tobit estimates, what is the expected fraction of outlets that are franchised for the firm with average characteristics? Explain your work.
d. Based on the tobit estimates, what is the probability of having between 25 and 75 percent of outlets that are franchises for the firm with average characteristics? Explain your work.
e. Based on the tobit estimates, what is the marginal effect of an additional full-time employee per outlet (EMFT) on the expected percent of outlets franchised? Explain your work.

ECO671, Spring 2005, Midterm 2
Prof. Bill Even

4. (15 points) Suppose that you estimate a logit model of whether a person owns a home or not as a function of whether he or she is single. In the data, 6,000 of 10,000 single people own a home, and 18,000 of 20,000 married people own a home.

Suppose the logit model is specified as

$$P_i = \frac{\exp(\alpha+S_i\beta)}{1+\exp(\alpha+S_i\beta)}$$

where $P_i$ is the probability that a person owns a home and $S_i$ is a dummy variable indicating that the person is single.

a. Write out the likelihood function for the logit model and data described.

b. Derive the maximum likelihood estimators of $\alpha$ and $\beta$. Show your work and provide numerical estimates of both parameters.

c. Provide a likelihood ratio test of the null hypothesis that home ownership is equal for single and married people. Indicate whether you reject the null at the .05 level of significance.

d. [Bonus credit -- 5 points] If 0 of 10,000 single own a home and 18,000 of 20,000 married people own a home, what would be the maximum likelihood estimate of $\alpha$ and $\beta$? Explain [You may refer to work in (a) and (b) to get your answer here.]