Gov 2001: Final Exam

Due Thursday, May 14th by 5 p.m.

Instructions: Please complete the exam online via Canvas and submit your R code in the submission box at the end of the exam. Late exams will not be accepted. Collaboration is not permitted for the final exam. Please email any questions about the exam to both Teaching Fellows and do not discuss the exam online on Canvas. If we are capable of answering the question, we will post an announcement online via Canvas so that all exam-takers have access to the same clarifying information.

Problem 1

In this problem, you will evaluate whether employment-related health insurance affects the probability that individuals change jobs. You will use data collected from a survey of married men to evaluate whether "job lock" occurs. Job lock refers to the problem of employees being constrained from seeking new employment opportunities because they risk losing existing employer-provided health insurance. The data are in R workspace format and in the `employmentdata.rdata` file. Below is a brief description of the relevant variables:

- chjob: whether the worker voluntarily changed jobs (1=yes; 0=otherwise)
- hi: whether the worker had employment-related health insurance in the original job (1=yes; 0=otherwise)
- othhi: whether the worker had other sources of health insurance available while at the original job (1=yes; 0=otherwise)
- lhwage: log of the worker's hourly wage at the original job
- winc: annual income of worker's wife in dollars
- nfam: number of individuals in the worker's family
- educ: worker's number of years of education
- exper: worker's number of years of work experience in the labor force
- afram: whether the worker is African American (1=yes; 0=otherwise)

1.A

Estimate the probability of voluntarily changing jobs (chjob) for workers with employment-related health insurance (hi) using logistic regression and the following covariates: othhi, an interaction between hi and othhi, lhwage, winc, nfam, educ, exper, afram. You may use a canned regression routine to answer this question. Report your coefficient estimates and standard errors below.

1.B

Calculate:

a) The expected probability of voluntarily changing jobs for a white married man with employer-provided health insurance but no other sources of health insurance available, holding all other covariates at their mean values. Report a 95% confidence interval (using simulation) for your estimate.

b) The expected probability of voluntarily changing jobs for a white married man with employer-provided health insurance and other sources of health insurance available, holding all other covariates at their mean values. Report a 95% confidence interval (using simulation) for your estimate.

c) The expected probability of voluntarily changing jobs for a white married man without employer-provided health insurance and no other sources of health insurance available, holding all other covariates at their mean values. Report a 95% confidence interval (using simulation) for your estimate.

d) The expected probability of voluntarily changing jobs for a white married man without employer-provided health insurance but other sources of health insurance available, holding all other covariates at their mean values. Report a 95% confidence interval (using simulation) for your estimate.

1.C

Calculate:

a) The expected difference in probability of voluntarily changing jobs between a white married man with employer-provided health insurance, but no other coverage and a white married man without employer-provided health insurance and no other coverage, holding all other covariates at their mean values. Report a 95% confidence interval (using simulation) for your estimate.

b) The expected difference in probability of voluntarily changing jobs between a white married man without employer-provided health insurance and no other coverage and a

white married man without employer-provided health insurance, but with other coverage holding all other covariates at their mean values. Report a 95% confidence interval (using simulation) for your estimate.

1.D

What do the results from 1.B and 1.C reveal about the existence or non-existence of job lock? How does having employer-provided heath insurance affect the probability of voluntarily changing jobs? How does having other sources of insurance moderate that relationship? Is this consistent with the job lock hypothesis?

Problem 2: Tobit Regression

The tobit regression model is used for continuous dependent variables that are censored from below. Tobit regression estimates a normal linear regression model where our data is observed for values greater than some cutoff point c. If the values are less than c, then we only observe c. The model is as follows:

Let Y^* be a latent (unobserved) dependent variable with the stochastic component

$$Y_i^* \sim \text{Normal}(\mu_i, \sigma^2)$$

where μ_i is a vector of means and σ^2 is a scalar variance parameter. Y_i^* is not directly observed. Rather, we observe Y_i which is defined as

$$Y_i = \begin{cases} Y_i^* & \text{if } Y_i^* > c \\ c & \text{if } Y_i^* \leq c \end{cases}$$

where c is the lower bound below which Y_i^* is censored.

The systematic component is given by

$$\mu_i = x_i\beta$$

where x_i is a vector of covariates and β is the vector of coefficients.

A)

i) Let d_i be an indicator variable for each observation i where $d_i = 1$ if $y_i > 0$ and $d_i = 0$ if $y_i = c$. Given n independent and identically distributed observations Y_i paired with covariates X_i, write the probability of the data $\mathbf{Y} = \{Y_1, \ldots, Y_n\}$ given $\mathbf{X} = \{X_1, \ldots, X_n\}$ and the parameters: $Pr(\mathbf{Y}|\mathbf{X}, \beta, \sigma^2)$. Simplify as much as possible. Note that you can use the notation $\Phi(a)$ to denote the normal CDF, where $\Phi(a) = P(X \leq a)$.

ii) Using your result in i), derive the log-likelihood for the tobit model.

B) Write a function in R that evaluates the log-likelihood. `pnorm()` may be helpful for this function.

C) Load the `tobin` dataset from the `Zelig` package using `data(tobin)`. Using your log-likelihood function and `optim()`, run a tobit regression where `durable` is the dependent variable and `age` and `quant` are the independent variables. In this case, the dependent variable is censored at $c = 0$. Report estimates and standard errors for the βs and report an estimate of σ in a table. Do not use a canned tobit routine for this part.

D) Generate a histogram of 10,000 expected values of Y_i with `quant` at it's 20th percentile value and `age` at its mean. Set your seed to 12345 and do not set the seed again for the rest of Problem 2.

E) Generate a histogram of 10,000 expected values of Y_i with `quant` at it's 80th percentile value and `age` at its mean.

F) Setting `age` at its mean, generate a histogram of 10,000 first differences in the expected value of Y_i going from the 20th percentile of `quant` to the 80th percentile of `quant`.

If you were unable to complete part b), you may use the estimates from a canned tobit routine (such as the `tobit()` in the `AER` package. (Note: You should draw your σ parameter on the log (unbounded) scale when drawing the simulated parameters)

G) What is the average first difference and its standard error?