Problem Set #5: Assessing Job Lock

A key component in recent debates over health care reform is how best to provide universal health insurance. This question has become more urgent as the employer-provided health system has become increasingly expensive and untenable for those corporations that provide health benefits. Currently, employment and health coverage are highly correlated with one another. When a person loses his or her job or works for a small employer (or a large one like Wal-Mart) who does not provide health insurance, that individual and usually that individual’s entire family lacks health insurance. The idea behind universal coverage is that everyone would have health insurance regardless of employment status.

One concern about the current lack of universal health insurance is that people may not be able to change jobs easily if their current employer provides health insurance, since an alternative job might not provide health insurance. Health insurance is essentially a part of the total compensation the employee receives, and similar wages without comparable insurance would be considered lower total compensation, and would therefore be undesirable from the employee’s perspective. Alternatively, if there are costs to changing jobs other than losing insurance (e.g., having to change physicians or having one’s insurance premium rise due to a condition that a new insurance company considers a high risk, etc.) then the existence of health insurance in the current job may affect the frequency with which a person changes jobs. In other words, the existence of health insurance may “lock in” the employee with their current employer. This is known as “job lock.”

In this assignment you will analyze the importance of job lock. Specifically you will examine how the existence of employment-related health insurance affects the probability of changing jobs. To do this you will use data on individuals, including whether or not each person changed jobs and whether that individual had employment-related health insurance or other sources of health insurance. Note that what you are trying to explain is whether the individuals changed jobs (i.e., the dependent variable is 1 if the person changed jobs and 0 otherwise). You will therefore need to use techniques appropriate for qualitative dependent variables. The data are in a Stata dataset called joblock.dta and are described at the end of this assignment.
Part I: Data Analysis

1. Estimate a linear probability model where the dependent variable is chjob and the independent variables are a constant, hi, log(hwage), hi*othhi, othhi, educ, exper, union, nfam, and winc. What evidence is there for job lock in this specification?

2. Use probit to answer the following questions:

   a) Estimate a probit model where the dependent variable is chjob and the independent variables are a constant, log(hwage), union, hi, othhi, educ, exper, nfam, and winc. Is there evidence of significant job lock? Explain.

   b) Now add othhi*hi to your list of independent variables from (2a) and re-estimate the probit. Has your estimate of the effect of employment-related health insurance changed? How has it changed, and how does this change relate to what you would expect (make specific reference to what the coefficient on the othhi*hi variable means in a probit model)?

   c) Do the estimation in (2b) above using a logit model instead of a probit model. Do you find evidence of significant job lock?

   d) Are nfam and winc jointly significant? Manually calculate a likelihood ratio test to answer this question. Then compare your manual answer to Stata’s.

3. Summarize your results to all parts of this question in the table attached, and in addition show your calculations (you may do them by hand or by using Stata or Excel; there are Stata and Excel hints for doing the calculations on the next page, and you will find that this is much quicker than doing it by hand). Remember: you will be using one regression to find probabilities in multiple situations. Run only one regression for each of type of estimation (i.e., one linear probability model, one probit model, and one logit model).

   a) For the specifications of the linear probability, probit, and logit models with independent variables constant, log(hwage), hi, othhi, hi*othhi, educ, exper union, nfam, and winc, compute the probability of changing jobs for a married man (all individuals in this sample are married men, unfortunately) with employment-related health insurance, no union membership, 12 years of education, 10 years of experience, no other sources of health insurance, and who earns $280 per week and works 40 hours per week, with 2 children and a wife who earns $20,000 per year. Also compute the probability of changing jobs for a person with no employment-related health insurance, but who has other sources of health insurance.

   b) Compute the probabilities for a person with the same characteristics except with no employment-related health insurance and no other source of health insurance (i.e., even if the wife has income, assume she has no health insurance that provides coverage for her husband) and also compute the probabilities for a person with employment-related health insurance and additional sources of health insurance (for example if the wife does have health insurance that covers the husband).

   c) Compute the difference in the probability of changing jobs for a person with and without employment-related health insurance, and with no other sources of health insurance (and all other demographic characteristics as listed earlier in this question).
d) Compute the difference in the probability of changing jobs for a person with and without employment-related health insurance who does have other sources of health insurance (and all other demographic characteristics as listed earlier in this question).

e) Compute the difference in the probability differences in (c) and (d). What does this resulting number represent (i.e., say in words what this number tells you about the probability of changing jobs under particular conditions).

f) Using the table you have constructed, as well as the coefficient estimates from these specifications, what do you conclude about the existence of job lock? Explain.

**Part II: Stata and Excel Hints**

Probit and logit produce estimates of coefficients that look much like those estimated by ordinary least squares but cannot be interpreted in the same way. Logit and probit are non-linear estimation techniques. Coefficients found using logit and probit do not represent rates of change in the probability, as is true of the linear probability model. The only way to assess how much a change in a variable (for instance, years of education) affects the probability of “success” is to do “simulations.” Rates of change depend on the values that are set for all other independent variables. In order to do the analysis required for question 3, you will need to use either Stata or Excel to do a series of simulations. My preferred approach is to use Excel, because it is somewhat more flexible, but either approach will yield the same outcomes (to at least 4-6 decimal places).

**Option 1: Use Stata**

The easiest Stata alternative is to use the `prvalue` command after you execute a regression. The `prvalue` command has the further benefit of providing a 95% confidence interval around the predicted probability after you run your model. However, `prvalue` may not be installed on your machine. To find out, simply run a logit or probit model and then type in `prvalue`. If you have it, the command will execute. If you get an error, then run this command in Stata:

```
net from http://www.indiana.edu/~jslsoc/stata/
```

A list of potential files to install will pop up. You want to install `spost9_ado` if you are using Stata 9.0. When you click on this option a new window will appear. On the right side of the window there will be the following text: (click here to install). Click on this link to install the command.

To use `prvalue` you need to tell it which values to use for each of the independent variables. For instance, if you estimated the following model:

```
logit chjob black lnhwage winc
```

you could then predict the probability of changing jobs for an African-American with a log hourly wage of 1.65 and wife’s annual income of 26,000 by using the following command:

```
. prvalue, x(black=1 lnhwage=1.65 winc=26000)
```

Here is the output:
logit: Predictions for chjob

Confidence intervals by delta method

95% Conf. Interval

Pr(y=1|x): 0.1514 [ 0.1108, 0.1920]
Pr(y=0|x): 0.8486 [ 0.8080, 0.8892]

black lnhwage winc
x= 1 1.65 26000

If you fail to specify a value for every independent variable in the regression, the command will use the means of the data instead. For instance, we could just look at the results by specifying a value for black:

. prvalue, x(black=1)

logit: Predictions for chjob

Confidence intervals by delta method

95% Conf. Interval

Pr(y=1|x): 0.1040 [ 0.0766, 0.1313]
Pr(y=0|x): 0.8960 [ 0.8687, 0.9234]

black lnhwage winc
x= 1 2.2953879 11969.432

As the `summ` below demonstrates, the values `prvalue` selected for `lnhwage` and `winc` are the means of the data:

. summ lnhwage winc

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnhwage</td>
<td>2978</td>
<td>2.295388</td>
<td>0.554288</td>
<td>0.0582689</td>
<td>5.259109</td>
</tr>
<tr>
<td>winc</td>
<td>2978</td>
<td>11969.43</td>
<td>13878.64</td>
<td>1</td>
<td>274310</td>
</tr>
</tbody>
</table>

If you cannot install `prvalue` there is another procedure. You can make use of the fact that Stata saves the estimated coefficient values after you run a linear regression, a probit, or a logit. The estimated coefficient for each independent variable is saved in something called `-coef[varname]` where `varname` is the name of the variable whose coefficient you want to access. To do the calculations, then, you just need to run the regression (linear probability or probit or logit), and then do the computation as follows:

```
gen predval = _coef[_cons] + _coef[var1]*(value for var1) + _coef[var2] *(value for var2)
```

For example, to get a predicted probability for a linear probability model with independent variables `hi`, `educ`, and `exper`, where we want the predicted probability for `hi = 1 educ = 12, and exper=10` type the following:

```
reg chjob hi educ exper
```
gen linpred = _coef[_cons] + _coef[hi]*1 + _coef[educ]*12 + _coef[exper]*10

tabulate linpred

The first line estimates the coefficients. The second line computes a predicted probability for hi = 1, educ = 12, and exper = 10, and the third line prints the result.

For the probit and logit predicted probabilities, the above procedure will get you a value of Z (also called the “index”), but you must still convert this into a predicted probability, using either the standard normal or logistic CDF. For probit, Stata can do this for you:

probit chjob hi educ exper

gen probpred = _coef[_cons] + _coef[hi]*1 + _coef[educ]*12 + _coef[exper]*10

gen prpred = normprob(probpred)

tabulate prpred

The first line estimates the parameters. The second line computes the value of Z for hi=1, educ= 12, exper=10 The third line computes F(Z) where F is the standard normal CDF. The fourth line prints the result. Make sure you understand how Stata is computing normprob(probpred) (i.e., make sure you could get the same number from a standard normal table).

For the logit model, you can compute F(Z) yourself

logit chjob hi educ exper

gen logpred = _coef[_cons] + _coef[hi]*1 + _coef[educ]*12 + _coef[exper]*10

gen lopred = 1/(1+exp(-logpred))

tabulate lopred

Note: you must do the calculations as the next step after estimating the parameters. If you estimate a new model, the coefficients from the previous model will no longer be saved.

To be valid, you must use every coefficient in the regression. For instance, you cannot find the probability by first estimating the logit specified above and then calculating the probability using the hi and educ variables. If you use only a subset of your regressors in the calculation of predicted probability, you will get the wrong answer.

Option 2: Use Excel

To use Excel, you must first transfer at least the coefficients to a spreadsheet; you may also wish to transfer the means for each of the independents. There may be more efficient ways, but I usually do the following:
1. Run the estimation, summary, etc. in Stata.
2. Select the table of coefficients, standard errors, and so forth in Stata’s “Results” window (i.e., drag through the information you want in Excel while holding down the right mouse button).
3. Select “Copy Table” (shortcut: Shift-Control-C) from Stata’s Edit menu
4. Switch to Excel; select the point at which you wish to paste the numbers, and then paste.
5. Enter the values for the independent variables in another column. Again: you must enter a value for each independent that is in the regression.
6. Multiple the coefficient cell by the value cell. This is the “Xbeta” column.
7. Sum the Xbeta column.

For the linear probability model, the sum of the Xbeta column is the predicted probability. For logit and probit, you must do an additional step. The sum of the Xbeta column is actually the Z score or “index.” The index is not the predicted probability for logit and probit. To find the predicted probability, you need to do one additional calculation to convert the index into a probability:

For probit, use the following function if the index is in cell A12: \( =NORMDIST(A12,0,1,1) \)

For logit, use the following function if the index is in cell A12: \( =1/(1+\exp(-A12)) \)

### TABLE I

*Probabilities of changing jobs*

<table>
<thead>
<tr>
<th></th>
<th>Linear Probability</th>
<th>Probit</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>With hi, No, othhi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No hi, No othhi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With hi, With othhi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No hi, With othhi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff. between hi and no hi, with othhi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff. between hi and no hi, without othhi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff between hi and no hi, between othhi and no othhi</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Description

joblock.dta

The data come from the 1987 National Medical Expenditure Survey conducted by the AHCPR (Agency for Health Care Policy Research). This survey of approximately 14,000 households (38,446 individuals) collected information about health status, health insurance, and medical care utilization in 1987 through a series of 4 interviews. Additionally, several questions relating to employment were asked which make it possible to determine whether an individual changed jobs between the first and fourth interviews. The sample here is restricted to married men ages 20-55 who were employed but not self-employed at the first interview, were not in the military, and were married to the same individual at the time of the first and fourth interviews. (Note: these restrictions are made in part to make the sample size more tractable for this assignment, not because the additional data are not interesting).

The variable describing job changes is restricted to voluntary job changes only, where the job change occurred between the first and fourth interviews. The restriction to voluntary job changes is made because we are interested in job lock, i.e., whether a person does not WANT to change jobs because the current job provides health insurance. The other information regarding employment was collected during the first interview, and therefore refers to the initial employment. The job change variable equals one if a person changed jobs (voluntarily) and zero otherwise. The health insurance variable equals one if the individual had an employment-related health insurance plan, which was typically part of the individual employee’s benefit plan and generally involved some copayment contribution by the individual (i.e., was not entirely paid for by the employer). The variable othhi equals one if the person had an alternative source of health insurance coverage (usually health insurance provided by his wife’s employer) which could be used in place of the individual’s own insurance should he decide to quit his job.

Variable Definitions

chjob  = 1 if worker changed jobs voluntarily between interview, = 0 otherwise
black  = 1 if black, = 0 otherwise
educ  = number of years of education
exper  = number of years of work experience (in the labor force, not necessarily at the current job)
hi  = 1 if the worker has employment-related health insurance, = 0 otherwise
hwage  = hourly wage in the original job
union  = 1 if a union member in the original job, = 0 otherwise
nfam.  = number of individuals in the family
othhi  = 1 if other sources of health insurance coverage are available, = 0 otherwise
winc  = wife’s annual income (in $)
intdif  = interval in days between first and fourth interviews
ind  = industry code for original job
occ  = occupational code for original job