90-771, Applied Econometrics II
Heinz School, Carnegie Mellon University
Spring, 2004-5

Solution #2

Please use the MEPS data from the website for this assignment. We are curious about the determinants of having health insurance. For all questions, I am referring to a model with insurance as the LHS variable and age, sex, employment, and income as RHS variables.

1. **Please calculate an estimate of and a 95% CI for the effect of employment on the probability of having health insurance. Do this using a logit model, a probit model, and a linear probability model. Do the calculation for an average person.**

   First, observe that the request for a 95% confidence interval is a typo. We only know how to do this for the linear probability model.

   For the linear probability model, the effect of employment on the probability of insurance does not depend on the levels of the various RHS variables: we can just read if off a regression table. On page 6 of the SAS output are the results of estimating the LPM by OLS. We see that the estimated effect of going from being unemployed to employed is to increase the probability of being insured by about 1.9 percentage points. To make a 95% CI, we look at the covariance matrix of the estimates (on page 7) and pick off the variance of the coefficient on employed, 0.000115. The square root of this, 0.011, is the standard error of the coefficient on employed. So a 95% CI for the effect of employed on insured is $0.019 \pm 1.96(0.011)$ or 0.019 ± 0.022.

   To calculate the effect of employed going from 0 to 1 in the probit and logit models, we must know what the levels of the other X variables are. The question says to use means of the other variables. On page 5, we see that the average person has sex=0.46, employed=0.76, age=41.4, and income in thousands = 25.

   So, using the logit model (on page 9), we can calculate the value of $\widehat{Y^*}$ for an average employed person:

$$\begin{aligned}\widehat{Y^*} &= 0.377 + 0.018 * 41.4 - 0.51 * 0.46 - 0.16 + 0.045 * 25 \\ &= 1.85\end{aligned}$$

   This gives a probability of being insured equal to:
   $P\{\text{insured}\} = \frac{exp(1.85)}{1+exp(1.85)} = 0.86$

Again using the logit model (on page 9), we can calculate the value of $\widehat{Y*}$ for an average unemployed person:

$$\begin{aligned}\widehat{Y*} &= 0.377 + 0.018 * 41.4 - 0.51 * 0.46 + 0.045 * 25 \\ &= 2.01\end{aligned}$$

This gives a probability of being insured equal to:

$P\{\text{insured}\} = \frac{exp(2.01)}{1+exp(2.01)} = 0.88$

According to the Logit model, getting a job reduces the average person's chance of being insured by about two percentage points.

Now, using the probit model on page 12, we again calculate the value of the index for an average employed person:

$$\begin{aligned}\widehat{Y*} &= 0.271 + 0.011 * 41.4 - 0.27 * 0.46 - 0.019 + 0.02 * 25 \\ &= 1.08\end{aligned}$$

This gives a probability of being insured of:

$P\{\text{insured}\} = \Phi(1.08) = 0.86$

Again, using the probit model on page 12, we calculate the value of the index for an average unemployed person:

$$\begin{aligned}\widehat{Y*} &= 0.271 + 0.011 * 41.4 - 0.27 * 0.46 + 0.02 * 25 \\ &= 1.1\end{aligned}$$

This gives a probability of being insured of:

$P\{\text{insured}\} = \Phi(1.1) = 0.86$

In conclusion, the LPM (with robust standard errors!) tells us that employment increases the probability of having insurance by a statistically insignificant and small two percentage points. The logit model tells us that employment decreases the probability of insurance by a statistically significant two percentage points. The probit model tells us that employment decreases the probability of having insurance by a statistically insignificant less than one percentage point.

2. **Let's compare the probit, logit and linear probability models. How often does the lpm predict probabilities outside the unit interval? How about logit and probit?**

   We know that logit and probit cannot ever predict outside the 0,1 interval. If we need confirmation, we can look at the results on pages 13 and 14.

Both of the Phats, for logit and probit have minimum values greater than zero and maximum values less than 1, so they are always between 0 and 1.

The LPM is different. Looking at page 8, we see that the Phat for the LPM is sometimes greater than 1: its maximum value is 1.32. I then defined a dummy variable Yhat01 which is equal to one if Yhat is outside the 0,1 interval (see the SAS program). Looking on page 8 again, we see that the average of Yhat01 is 0.043, so about 4.3% of the observations are outside the 0,1 interval.

3. **Consider the income effect. Please calculate the effect, for each person in the data, of a $1,000 increase in income on the probability of having health insurance. Take the average of this effect over all individuals in the data. How do the results for the logit, probit, and lpm compare?**

   For LPM, we can just look at the regression results on page 6. A $1,000 increase in income is a 1 unit increase in incomeK, so the effect of a $1,000 increase in income is to raise the probability of insurance by about 3.2 percentage points.

   For logit and probit, we proceed as we did in class, and these results are on pages 13 and 14. The logit and probit models yield a marginal effect of 5.8 percentage points and 4.7 percentage points, respectively.

   Notice, I divided income by 1000 to make this problem a little easier. If you prefer, you could do all these calculations with the original income variable and then multiply the marginal effects you calculate for a $1 increase in income by $1,000 to get the effect of a $1,000 increase in income.

4. **Using the logit model, test the null hypothesis that employment has no effect on the probability of having health insurance.**

   To do this, we look at the output on page 10. Employment has a p-value of about 0.03. Therefore, at a significance level of 10% or 5%, we would reject the null hypothesis of no effect of employment on insurance. At the 1% significance level, we would accept the null of no effect of employment on insurance.

5. **Now test the joint hypothesis that neither age nor sex has an effect on the probability of having health insurance using the probit model.**

   To do this, we compare the models on pages 11-12 and 15-6. We need to calculate 2 times the difference in the log-likelihoods. Helpfully, SAS calculates -2 times the log-likelihood for each model on pages 11 and 15. The difference between the two is 7965.6-7838.5=127.1. We compare this number to the chi-squared table with 2 degrees of freedom. 127 is a huge number for a chi-squared table, so we will reject this hypothesis at any reasonable significance level.