Professor R.L. Oaxaca — Assignment #1
Econ 522b — Fall 2009

Due Thursday, September 17
(20 points)

This assignment is on discrete dependent variable models. The necessary data are contained in both the Excel file 'dat101.xls' and the STATA file 'dat101.dta' available at http://www.u.arizona.edu/~rlo/. Be sure to attach the supporting computer print out to the completed assignment and make clear where your answers are shown.

The data for this exercise are a national sample of adults between the ages of 33 and 41 in 1998 and are drawn from the NLSY79 data set.

For purposes of the exercises below, restrict your sample to the subset for whom $ntinc98_i > 0$ where $ntinc98$ is household net income ($1,000's).

Let $\mathrm{I}_i = \beta_0 + \beta_1 emp98_i + \beta_2\, black_i + \beta_3 other_i + \beta_4 female_i + \beta_5 ntinc_i + \beta_6\,(ntinc)_i^2 + \beta_7\,(ntinc)_i^3,\ i = 1, ..., T$

where $\mathrm{I}(\cdot)$ is an index function, $emp98$ is an indicator variable for being currently employed, $black$ and $other$ are indicator variables for nonwhites, $female$ is an indicator variable for females, and $ntinc = (ntinc98)\text{x}10^{-2}$.

$$hins98_i = 1 \quad \text{if the individual is currently covered by health insurance}$$
$$= 0 \quad \text{otherwise}$$

1. Estimate the following models and determine the marginal effects of the explanatory variables on Prob($hins98_i = 1$) (evaluate at the sample mean for continuous variables and (1,0) for the binary regressors):

   a. The LP model $hins98_i = \mathrm{I}_i + \varepsilon_i$ by OLS and use the White procedure to correct the estimated standard errors.

   b. Prob($hins98_i = 1$) = Prob($\varepsilon_i > -\mathrm{I}_i$) where $\varepsilon_i \sim N(0,1)$

   c. Prob($hins98_i = 1$) = Prob($\varepsilon_i > -\mathrm{I}_i$) where $\varepsilon_i \sim Logistic(0, \pi^2/3)$

2. For all three probability models determine the following:

   a. For each individual calculate the estimated probability that $hins98_i = 1$ and average these over the sample. Is this average equal to the sample mean for $hins98$? Explain.

   b. Calculate the estimated probability that $hins98_i = 1$ at the sample mean characteristics of the explanatory variables. Is this average equal to the sample mean for $hins98$? Explain.

c. Generate a binary prediction of whether or not $hins98_i = 1$ according to $\hat{P}_i = 1[F(\hat{\text{I}}_i) > 0.5]$, where $F(\cdot)$ is the CDF of the corresponding probability model. Determine the goodness of fit of each model according to its predictive accuracy rate and according to the R^2 (psuedo R^2 as defined as the squared correlation between $hins98$ and the estimated probability that $hins98_i = 1$ in the cases of the probit and logit models). Does the goodness of fit evidence allow one to choose the best model among the three competing probability models? Explain.

3. Estimate the LPM using the SLS estimator discussed in class and use the White procedure to correct the estimated standard errors.

4. Estimate the LPM using the FGLS SLS estimator.