TSP-Problem Set 6:
Logit and Probit

Preparations

Please create a new folder for this exercise session with your name in directory T:. Then go to L:\Applied Econometrics\ex06 and copy the files into your folder.

1) Approval of mortgage loans

(Based on Wooldridge, Computer Exercise C7.8 and C17.2, pages )

Use the data in ex06_1.xls taken from the Boston Home Mortgage Disclosure Act (HMDA) data set for this exercise. The binary variable to be explained is approve, which is equal to one if a mortgage loan to an individual was approved. The key explanatory variable is white, a dummy variable equal to one if the applicant was white. The other applicants in the data set are black and Hispanic. To test for discrimination in the mortgage loan market, we will use three kinds of models: a linear probability model, a probit model and a logit model. The relevant variables are:

\[
\begin{align*}
\text{approve} & = 1 \text{ if mortgage loan was approved} \\
\text{white} & = 1 \text{ if applicant white} \\
\text{male} & = 1 \text{ if applicant male} \\
\text{married} & = 1 \text{ if applicant married} \\
\text{dep} & \text{ number of dependents} \\
\text{unem} & \text{ unemployment rate by industry} \\
\text{vr} & = 1 \text{ if tract vacancy rate > MSA median} \\
\text{(MSA = Boston Metropolitan Statistical Area)}
\end{align*}
\]
mortlat1 one or two late payments
mortlat2 more than two late payments
chist consumer payments credit history:
\[ = 0 \text{ if accounts delinquent } \geq 60 \text{ days} \]
loanprc loan-to-price ratio
pubrec = 1 if filed bankruptcy
cosign = 1 if there is a cosigner
sch = 1 if > 12 years schooling
hrat housing exp, % total inc
obrat other oblgs, % total inc

(a) If there is discrimination against minorities, and the appropriate factors have been controlled for, what is the sign of the coefficient on white going to be in the above-mentioned models?

(b) Regress approve on white using a simple linear regression model. Interpret the coefficient on white. Is it statistically significant? Is it practically large?

(c) As controls, add the variables hrat, obrat, loanprc, unem, male, married, dep, sch, cosign, chist, pubrec, mortlat1, mortlat2, and vr to the linear regression model. What happens to the coefficient on white? Is there still evidence of discrimination against nonwhites?

(d) Estimate a probit model of approve on white. Find the estimated probability of loan approval for both whites and nonwhites. How do these compare with the linear probability estimates?

(e) Now, add the variables hrat, obrat, loanprc, unem, male, married, dep, sch, cosign, chist, pubrec, mortlat1, mortlat2, and vr to the probit model. Is there statistically significant evidence of discrimination against nonwhites?

(f) Estimate the model from part (e) by logit. Compare the coefficient on white to the probit estimate.

(g) How would you compare the size of the discrimination effect between probit and logit?
2) Probit model for mothers’ labor force participation

Read in the data set ex06_2.raw into TSP. It contains selected variables of the German Socio-Economic Panel (SOEP) for German mothers for the year 2004. We will study mothers’ labor supply: participation and hours worked. The data set contains the following variables in the given order:

- *ifwork*: dummy=1 if mother working
- *hours*: working hours (0 if not working)
- *manearn*: partner’s labor earnings
- *married*: dummy=1 if married
- *kids*: number of children

(a) Estimate the influence of partner’s earnings, marriage and the number of children on the probability of working. In order to do so use a probit model with labor force participation as the dependent variable.

(b) **Marginal effect at the mean for a continuous variable**: What is the marginal effect of partner’s earnings on the probability of working for an average individual?

(c) **Marginal effect at the mean for a dummy variable**: What is the marginal effect of marriage on the probability of working for an average individual?

(d) **Average marginal effect of a continuous variable**: What is the average marginal effect of partner’s earnings on the probability of working?

(e) **Average marginal effect of a dummy variable**: What is the average marginal effect of marriage on the probability of working? Keep in mind that a dummy variable only takes on values 0 and 1.

(f) **Marginal effect of a discrete variable**: What is the marginal effect of the number of children on the probability of working? Differentiate between the marginal effect at the mean and the average marginal effect. Estimate the effects in analogy to the procedure for a dummy variable.
### 3.1) Appendix I: TSP-commands

<table>
<thead>
<tr>
<th>Command</th>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>probit</code></td>
<td><code>probit (options) y c x1 x2 x3 ;</code></td>
<td>When estimating a binary logit model, add the option <code>nchoice=2</code>; otherwise, TSP will run a multinomial logit estimation.</td>
</tr>
<tr>
<td><code>logit</code></td>
<td><code>logit (options) y c x1 x2 x3 ;</code></td>
<td></td>
</tr>
</tbody>
</table>


### 3.2) Appendix II: Marginal Effects in the Probit model

<table>
<thead>
<tr>
<th>Continuous variable</th>
<th>Marginal effect at the mean</th>
<th>Average marginal effect (AME)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi(x'_i \beta) \beta_k )</td>
<td>( \phi(\bar{x}' \beta) \beta_k )</td>
<td>( \frac{1}{n} \sum_{i=1}^{n} \phi(x'_i \beta) \beta_k )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dummy variable</th>
<th>Marginal effect at the mean</th>
<th>Average marginal effect (AME)</th>
</tr>
</thead>
</table>
| \( \Phi(\beta_1 + \beta_2 x_{2i} + \cdots + \beta_{k-1} x_{k-1,i} + \beta_k) \)  
\( -\Phi(\beta_1 + \beta_2 x_{2i} + \cdots + \beta_{k-1} x_{k-1,i}) \) | \( \Phi(\bar{x}' \beta | x_k = 1) - \Phi(\bar{x}' \beta | x_k = 0) \) | \( \frac{1}{n} \sum_{i=1}^{n} [\Phi(x'_i \beta | x_k = 1) - \Phi(x'_i \beta | x_k = 0)] \) |
| \( \Phi(x'_i \beta | x_k = 1) - \Phi(x'_i \beta | x_k = 0) \) | \( Pr(y_i = 1|x_i, x_k = 1) - Pr(y_i = 1|x_i, x_k = 0) \) |  |

<table>
<thead>
<tr>
<th>Discrete variable</th>
<th>Marginal effect at the mean</th>
<th>Average marginal effect (AME)</th>
</tr>
</thead>
</table>
| \( \Phi(\beta_1 + \beta_2 x_{2i} + \cdots + \beta_{k-1} x_{k-1,i} + \beta_k(c_k + 1)) \)  
\( -\Phi(\beta_1 + \beta_2 x_{2i} + \cdots + \beta_{k-1} x_{k-1,i} + \beta_k c_k) \) | \( \Phi(\bar{x}' \beta | x_k = c_k + 1) - \Phi(\bar{x}' \beta | x_k = c_k) \) | \( \frac{1}{n} \sum_{i=1}^{n} [\Phi(x'_i \beta | x_k = c_k + 1) - \Phi(x'_i \beta | x_k = c_k)] \) |
| \( \Phi(x'_i \beta | x_k = c_k + 1) - \Phi(x'_i \beta | x_k = c_k) \) | \( Pr(y_i = 1|x_i, x_k = c_k + 1) - Pr(y_i = 1|x_i, x_k = c_k) \) |  |