Applied Microeconometrics
Chapter 2

Models with binary dependent variables
1. Introduction to the Probit Model
2. Estimation
3. A Practical Application
4. Coefficients and Marginal Effects
5. Goodness-of-Fit Measures
6. Hypothesis Tests
7. Probit vs. Logit
1. Introduction to the Probit model

Recall our example from the introduction:

- **Binary** choice variable: voting yes-no \( y \in \{0, 1\} \)
- Explanatory variable: household income \( x \in \mathbb{R}^+ \)
Introduction to the Probit model – latent variables

- We aim to model the probability that the observed binary variable takes one of its values conditional on $x$, such as

$$ p = P(y_i = 1 | x) $$

where $0 \leq p \leq 1$

- We need to derive this probability to estimate the model by maximum likelihood
Introduction to the Probit model – latent variables

• We think of the process generating observations on discrete outcome $y$ as driven by an unobserved (latent) variable $y^*$ which can take all values in $(-\infty, +\infty)$.

• Example: $y^*$ = net utility from labour income, $y$ = observed labour market participation

• the underlying model is in terms of the latent variable and is linear

$$y_i = \begin{cases} 1, & y_i^* > 0 \\ 0, & y_i^* \leq 0 \end{cases}$$

$$y_i^* = x_i'\beta + \varepsilon_i$$
Introduction to the Probit model – latent variables

Probit is based on the latent model:

\[ P(y_i = 1 | x) = P(y_i^* > 0 | x) \]
\[ = P(x_i' \beta + \varepsilon_i > 0 | x) \]
\[ = P(\varepsilon_i > -x_i' \beta | x) \]
\[ = 1 - F(-x_i' \beta) \]

Assumption: Error terms are independent and normally distributed:

\[ P(y_i = 1 | x) = 1 - \Phi\left(-\frac{x_i' \beta}{\sigma}\right), \sigma = 1 \]
\[ = \Phi(x_i' \beta) \quad \text{because of symmetry} \]
Background on probability distribution functions (PDF)

- PDF: probability distribution function $f(x)$
- Example: Normal distribution:
  
  $\phi(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{(x-\mu)^2}{\sigma^2} \right)}$

- Example: Standard normal distribution: $N(0,1)$, $\mu = 0$, $\sigma = 1$
  
  $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
Notation and statistical foundations – CDF

- CDF: cumulative distribution function $F(x)$
- Example: Standard normal distribution:

$$
\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx
$$

- The cdf is the integral of the pdf. It is bounded between 0 and 1, as required
2. Estimation

- The probability of choosing $y_i = 1$ is
  $\Phi(x_i'\beta)$

- Similarly, the probability of choosing $y_i = 0$ is
  $1 - \Phi(x_i'\beta)$

- Combining these, the likelihood of observing unit $i$ in the state actually chosen is

$$L_i(x_i, \beta) = \Phi(x_i'\beta)^{y_i} \left(1 - \Phi(x_i'\beta)\right)^{1-y_i}$$
Derivation of the log likelihood function

- Taking the product over all units in the sample $i = 1, \ldots, n$ gives the likelihood function

$$L(y | x, \beta) = \prod_i \Phi(x_i \beta)^{y_i} \left[ 1 - \Phi(x_i \beta) \right]^{(1-y_i)}$$

$$= \prod_i \Phi_i^{y_i} (1 - \Phi_i)^{1-y_i}$$

- It is more convenient to use the log likelihood function:

$$\ln L = \sum_i y_i \ln \Phi_i + (1 - y_i) \ln (1 - \Phi_i)$$
The ML principle

- The principle of ML: Which value of $\beta$ maximizes the probability of observing the given sample?

$$\frac{\partial \ln L}{\partial \beta} = \sum_i \left[ \frac{y_i \varphi_i}{\Phi_i} + \frac{(1-y_i)(-\varphi_i)}{1-\Phi_i} \right] x_i$$

$$= \sum_i \left[ \frac{y_i - \Phi_i}{\Phi_i (1-\Phi_i)} \varphi_i \right] x_i = 0$$

- Usually, use $k$ explanatory variables rather than one

- The gradient vector $\frac{\partial \ln L(\theta)}{\partial \theta}$ is also called the score vector
Distribution of the ML estimator

• Under certain regularity conditions (see Cameron / Trivedi, p. 142) the MLE defined by $\frac{\partial \ln L(\theta)}{\partial \theta} = 0$ is consistent for $\theta_0$ and

$$ \sqrt{n}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} N\left[0, -A_0^{-1}\right] $$

where $A_0 = \text{plim} n^{-1} \left. \frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta'} \right|_{\theta_0}$

• Then, the asymptotic distribution of the MLE can be written as

$$ \hat{\theta}_{ML} \xrightarrow{a} N\left[0, - E\left[\left(\frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta'}\right)\right]^{-1}\right] $$
Derivation of the MLE

• It can be shown that the likelihood function for the Probit model is globally concave \( \rightarrow \) there exists only one maximum of the likelihood function

• However, the first-order conditions \( \frac{\partial \ln L(\theta)}{\partial \theta} = 0 \) cannot be solved analytically

• Hence, need to find numerical solutions

• Mostly used: Newton-Raphson Algorithm
Newton-Raphson Algorithm

- Iterative procedure: from an estimate in the s-th step, apply a rule that finds the next-step estimate
- The rule must be chosen such that it ensures a move towards the maximum
- Process stops if the distance between steps s and s+1 becomes very small
Newton-Raphson Algorithm

- In the Newton-Raphson case, the rule is
  \[ \hat{\theta}_{s+1} = \hat{\theta}_s - H_s^{-1} g_s \]

where \( g_s \) is the gradient \( g_s = \frac{\partial \ln L(\theta)}{\partial \theta}|_{\theta_s} \) derived from step \( s \) and

\[ H_s = \left. \frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta'} \right|_{\theta_s} \]

- Intuition: if the score is positive, need to increase \( \theta \) in order to get closer to maximum (note that \( H_s \) is always negative, as claimed previously).
Newton-Raphson Algorithm

Figure 8.2. Direction of step follows the slope.

Figure 8.3. Step size is inversely related to curvature.

Taken from:
K. Train (2003), Discrete Choice Methods with Simulation, Cambridge University Press
http://elsa.berkeley.edu/books/choice2.html

(Chapter on numerical maximisation highly recommended!)
Newton-Raphson Algorithm

What happens if the likelihood function is not globally concave?

Figure 8.6. NR in the convex portion of LL.

Taken from:
K. Train (2003), Discrete Choice Methods with Simulation, Cambridge University Press
http://elsa.berkeley.edu/books/choice2.html
(Chapter on numerical maximisation highly recommended!)
A Practical Application

- Analysis of the effect of a new teaching method in economic sciences

- Data:

<table>
<thead>
<tr>
<th>Obs.No.</th>
<th>GPA</th>
<th>TUCE</th>
<th>PSI</th>
<th>Grade</th>
<th>Obs.No.</th>
<th>GPA</th>
<th>TUCE</th>
<th>PSI</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.66</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>2.75</td>
<td>25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2.89</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>2.83</td>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3.28</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>3.12</td>
<td>23</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2.92</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>3.16</td>
<td>25</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>21</td>
<td>0</td>
<td>1</td>
<td>21</td>
<td>2.06</td>
<td>22</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2.86</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>3.62</td>
<td>28</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2.76</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>2.89</td>
<td>14</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>2.87</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>3.51</td>
<td>26</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>3.03</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>3.54</td>
<td>24</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>3.92</td>
<td>29</td>
<td>0</td>
<td>1</td>
<td>26</td>
<td>2.83</td>
<td>27</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>2.63</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>3.39</td>
<td>17</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>3.32</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>2.67</td>
<td>24</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>3.57</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>3.65</td>
<td>21</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>3.26</td>
<td>25</td>
<td>0</td>
<td>1</td>
<td>30</td>
<td>4</td>
<td>23</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>3.53</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>3.1</td>
<td>21</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>2.74</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>2.39</td>
<td>19</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Application – Variables

- **Grade**
  Dependent variable. Indicates whether a student improved his grades after the new teaching method PSI had been introduced (0 = no, 1 = yes).

- **PSI**
  Indicates if a student attended courses that used the new method (0 = no, 1 = yes).

- **GPA**
  Average grade of the student

- **TUCE**
  Score of an intermediate test which shows previous knowledge of a topic.
Application – Estimation

Estimation results of the model (output from Stata):

```
. probit grade psi tuce gpa

Iteration 0:  log likelihood =  -20.59173
Iteration 1:  log likelihood =  -13.315851
Iteration 2:  log likelihood =  -12.832843
Iteration 3:  log likelihood =  -12.818826
Iteration 4:  log likelihood =  -12.818803

Probit estimates

Number of obs = 32
LR chi2(3)     =  15.55
Prob > chi2    =  0.0014
Pseudo R2      =  0.3775

Log likelihood =  -12.818803

                      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------------------------------------------------------------------------------------
   grade               
      psi             1.426332   .595037    2.40   0.017         .2600814       2.592583
      tuce            .0517289   .0839001   0.62   0.529       -.1126927       .2161506
       gpa             1.62581   .6938818    2.34   0.019        .2658269       2.985794
     _cons            -7.452332   2.542467   -2.93   0.003       -12.43546       -2.469177
```

Application – Discussion

• ML estimator: Parameters were obtained by maximization of the log likelihood function. Here: 5 iterations were necessary to find the maximum of the log likelihood function (-12.818803)

• Interpretation of the estimated coefficients:
  • Unlike in OLS, estimated coefficients cannot be interpreted as the quantitative influence of the rhs variables on the probability that the lhs variable takes on the value one.
  • This is due to non-linearity and using the standard normal distribution for normalisation.
Coefficients and marginal effects

• The marginal effect of a rhs variable is the effect of an infinitesimal change (dummy variables: unit change) of this variable on the probability $P(Y = 1|X = x)$, given that all other rhs variables are constant:

$$\frac{\partial P(y_i = 1| x_i)}{\partial x_i} = \frac{\partial E(y_i | x_i)}{\partial x_i} = \varphi(x_i' \beta) \beta$$

• Recap: The slope parameter of the linear regression model measures directly the marginal effect of the rhs variable on the lhs variable.
Coefficients and marginal effects

- The marginal effect depends on the value of the rhs variable.
- Therefore, there exists an individual marginal effect for each person of the sample:
Coefficients and marginal effects – Computation

- Two different types of marginal effects can be calculated:
  - Average marginal effect
    Stata command: `margin`

```
Marginal effects on Prob(grade==1) after probit

| grade | Coef.    | Std. Err. | z   | P>|z|  | [95% Conf. Interval] |
|-------|----------|-----------|-----|------|----------------------|
| gpa   | 0.3637083| 0.1129461 | 3.22| 0.001| 0.1424101 - 0.5851586|
| tuce  | 0.011476 | 0.0184085 | 0.62| 0.533| -0.024604 - 0.047556 |
| psi   | 0.3737518| 0.1399912 | 2.67| 0.008| 0.0993741 - 0.6401295|
```

- Marginal effect at the mean:
  Stata command: `mfx compute`
Coefficients and marginal effects – Computation

• Principle of the computation of the average marginal effects:

• Average of individual marginal effects
Coefficients and marginal effects – Computation

- Computation of average marginal effects depends on type of rhs variable:
  - Continuous variables like TUCE and GPA:
    \[ AME = \frac{1}{n} \sum_{i=1}^{n} \varphi(x_i \beta) \beta \]
  - Dummy variable like PSI:
    \[ AME = -\frac{1}{n} \sum_{i=1}^{n} \left[ \Phi(x_i \beta | x_i^k = 1) - \Phi(x_i \beta | x_i^k = 0) \right] \beta \]
Coefficients and marginal effects – Interpretation

- Interpretation of average marginal effects:
  - Continuous variables like TUCE and GPA:
    A change of TUCE or GPA of size 1 changes the probability that the lhs variable takes the value one by X%.
  - Dummy variable like PSI:
    A change of PSI from zero to one changes the probability that the lhs variable takes the value one by X%.
Coefficients and marginal effects – Interpretation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated marginal effect</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>0.364</td>
<td>If the average grade of a student goes up by size 1, the probability for the variable grade taking the value one rises by 36.4%.</td>
</tr>
<tr>
<td>TUCE</td>
<td>0.011</td>
<td>As with GPA, with an increase of 1.1%.</td>
</tr>
<tr>
<td>PSI</td>
<td>0.374</td>
<td>If the dummy variable changes from zero to one, the probability for the variable grade taking the value one rises by 37.4%.</td>
</tr>
</tbody>
</table>
Coefficients and marginal effects – Significance

• Significance of a coefficient: test of the hypothesis whether a parameter is significantly different from zero.

• The decision problem is similar to the t-test, whereas the probit test statistic follows a standard normal distribution. The z-value is equal to the estimated parameter divided by its standard error.

• Stata computes a p-value which shows directly the significance of a parameter:

<table>
<thead>
<tr>
<th></th>
<th>z-value</th>
<th>p-value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>3.22</td>
<td>0.001</td>
<td>significant</td>
</tr>
<tr>
<td>TUCE:</td>
<td>0.62</td>
<td>0.533</td>
<td>insignificant</td>
</tr>
<tr>
<td>PSI:</td>
<td>2.67</td>
<td>0.008</td>
<td>significant</td>
</tr>
</tbody>
</table>
Coefficients and marginal effects

- Only the average of the marginal effects is displayed.
- The individual marginal effects show large variation:

```
Descriptive statistics for individual marginal effects

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>gpa</td>
<td>0.36379</td>
<td>0.21358</td>
<td>0.06783</td>
<td>0.64807</td>
</tr>
<tr>
<td>tuce</td>
<td>0.01148</td>
<td>0.00687</td>
<td>0.00209</td>
<td>0.02063</td>
</tr>
<tr>
<td>psi</td>
<td>0.37375</td>
<td>0.12878</td>
<td>0.06042</td>
<td>0.51959</td>
</tr>
</tbody>
</table>
```

**Stata command**: `margin, table`
Coefficients and marginal effects

• Variation of marginal effects may be quantified by the confidence intervals of the marginal effects.

• In which range one can expect a coefficient of the population?

• In our example:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated Coefficient</th>
<th>Confidence Interval (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>0.364</td>
<td>-0.055 - 0.782</td>
</tr>
<tr>
<td>TUCE</td>
<td>0.011</td>
<td>-0.002 - 0.025</td>
</tr>
<tr>
<td>PSI</td>
<td>0.374</td>
<td>0.121 - 0.626</td>
</tr>
</tbody>
</table>
Coefficients and marginal effects

- What is calculated by \texttt{mfx}?
- Estimation of the marginal effect at the sample mean.
Goodness of fit

- Goodness of fit may be judged by McFaddens Pseudo $R^2$.
- Measure for proximity of the model to the observed data.
- Comparison of the estimated model with a model which only contains a constant as rhs variable.
  - $\ln \hat{L}(M_{Full})$: Likelihood of model of interest.
  - $\ln \hat{L}(M_{Intercept})$: Likelihood with all coefficients except that of the intercept restricted to zero.
  - It always holds that $\ln \hat{L}(M_{Full}) \geq \ln \hat{L}(M_{Intercept})$
Goodness of fit

- The Pseudo $R^2$ is defined as:

$$PseudoR^2 = R^2_{McF} = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Intercept})}$$

- Similar to the $R^2$ of the linear regression model, it holds that $0 \leq R^2_{McF} \leq 1$

- An increasing Pseudo $R^2$ may indicate a better fit of the model, whereas no simple interpretation like for the $R^2$ of the linear regression model is possible.
Goodness of fit

- $R^2_{McF}$ increases with additional rhs variables. Therefore, an adjusted measure may be appropriate:

$$PseudoR^2_{adjusted} = \frac{\bar{R}^2_{McF}}{\ln \hat{L}(M_{Full}) - K} = 1 - \frac{\ln \hat{L}(M_{Full}) - K}{\ln \hat{L}(M_{Intercept})}$$

- Further goodness of fit measures: $R^2$ of McKelvey and Zavoinas, Akaike Information Criterion (AIC), etc. See also the Stata command `fitstat`. 
Hypothesis tests

- Likelihood ratio test: possibility for hypothesis testing, for example for variable relevance.

- Basic principle: Comparison of the log likelihood functions of the unrestricted model ($\ln L_U$) and that of the restricted model ($\ln L_R$)

- Test statistic: $LR = -2 \ln \lambda = -2(\ln L_R - \ln L_U)$
  \[ \lambda = \frac{L_R}{L_U} \quad 0 \leq \lambda \leq 1 \]

- The test statistic follows a $\chi^2$ distribution with degrees of freedom equal to the number of restrictions.
Hypothesis tests

- Null hypothesis: All coefficients except that of the intercept are equal to zero.
- In the example: LR $\chi^2(3) = 15.55$
- Prob > chi2 = 0.0014
- Interpretation: The hypothesis that all coefficients are equal to zero can be rejected at the 1 percent significance level.
The Logit model

- Binary dependent variable: \[ y = \begin{cases} 1 \\ 0 \end{cases} \]

- Let \( P(y_i = 1 \mid x) = F(x_i' \beta) \)
  (as in the case of Probit)

- In the Logit model, \( F(.) \) is given the particular functional form:
  \[ P(y_i = 1) = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)} \]
• The model is called Logit because the residuals of the latent model are assumed to be distributed standard logistic.
Notation and statistical foundations – distributions

- **Standard logistic distribution:**
  \[ f(x) = \frac{e^x}{(1 + e^x)^2}, \mu = 0, \sigma^2 = \frac{\pi^2}{3} \]

- **Exponential distribution:**
  \[ f(x) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}}, & x \geq 0 \\ 0, & x < 0 \end{cases}, \theta > 0, \mu = \theta, \sigma^2 = \theta^2 \]

- **Poisson distribution:**
  \[ f(x) = \frac{e^{-\theta} \theta^x}{x!}, \mu = \theta, \sigma^2 = \theta \]
PDF Probit vs. Logit

- PDF of Probit:  
- PDF of Logit:
CDF Probit vs. Logit

- $F(z)$ lies between zero and one
- CDF of Probit: $z = x_i \beta$

CDF of Logit:
Estimation output

The Logit model is implemented in all major software packages, such as Stata:

```
. logit grade psi tuce gpa

Iteration 0:  log likelihood = -20.59173
Iteration 1:  log likelihood = -13.496795
Iteration 2:  log likelihood = -12.929188
Iteration 3:  log likelihood = -12.889941
Iteration 4:  log likelihood = -12.889633
Iteration 5:  log likelihood = -12.889633

Logit estimates

Number of obs   =       32
LR chi2(3)      =      15.40
Prob > chi2     =     0.0015
Log likelihood  = -12.889633
Pseudo R2       =      0.3740

             | Coef.     Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+-----------------------------------------------
      grade   |                                    
        psi   |  2.378688   1.064564     2.23   0.025    .29218   4.465195
        tuce   |   .0951577   .1415542     0.67   0.501   -.1822835   .3725981
        gpa     |  2.826113   1.262941     2.24   0.025    .3507938   5.301492
     _cons    | -13.02135   4.931325    -2.64   0.008  -22.68657  -3.35613
```
Coefficient magnitudes

Coefficient Magnitudes differ between Logit and Probit:

<table>
<thead>
<tr>
<th></th>
<th>Probit</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>gpa</td>
<td>1,626</td>
<td>2,826</td>
</tr>
<tr>
<td>tuce</td>
<td>0,052</td>
<td>0,095</td>
</tr>
<tr>
<td>psi</td>
<td>1,426</td>
<td>2,379</td>
</tr>
</tbody>
</table>

This is due to the fact that in binary models, the coefficients are identified only up to a scale parameter.
Coefficient magnitudes

- Coefficient magnitudes can be made comparable by standardizing with the variance of the errors:
  - with logarithmic distribution: $\text{Var} = \pi^2/6$
  - with standard normal distribution: $\text{Var} = 1$

- Approximative conversion of the estimated values using

$$\frac{1}{\sqrt{\pi^2/6}} \approx 0.61$$
Marginal effects

For interpretation we have to calculate the marginal effects of the estimated coefficients (as in the Probit case)

| grade | Coef.    | Std. Err. | z   | P>|z|  | [95% Conf. Interval] |
|-------|----------|-----------|-----|------|---------------------|
| gpa   | .3682795 | .1088308  | 3.38| 0.001| .1549751            | .581584 |
| tuce  | .0122101 | .0177941  | 0.69| 0.493| -.0226656           | .0470859 |
| psi   | .3575152 | .1420034  | 2.52| 0.012| .0791936            | .6358367 |

Interpretation of the marginal effects analogous to the Probit model