Computer Tutorial 6

TOPIC: **Estimating and Interpreting Probit Models with *Stata*: Introduction**

DATA: **mroz.raw** (a text-format (ASCII) data file)

TASKS: ***Computer Tutorial 6*** demonstrates how to use *Stata* to compute ***maximum likelihood (ML) estimates of probit models*** for binary dependent variables. It also introduces you to ***hypothesis testing*** in probit models, and demonstrates how to compute the ***marginal effects*** of both continuous and binary explanatory variables in probit models.

- The ***Stata* commands** that constitute the primary subject of this tutorial are:

probit	Used to compute ML coefficient estimates in probit models of binary dependent variables.
dprobit	Used to compute ML estimates of the marginal probability effects of explanatory variables in probit models.
mfx compute	Used after probit estimation to compute the marginal probability effects of individual regressors in probit models.
test	Used after probit estimation to compute ***Wald tests*** of linear coefficient equality restrictions on probit coefficients.
lincom	Used after probit estimation to compute and test the marginal effects of individual explanatory variables.

- The ***Stata* statistical functions** used in this tutorial are:

normden(*z*)	Computes ***values of the standard normal density function***.

NOTE: *Stata* commands are *case sensitive*. All *Stata command names* must be typed in the Command window in ***lower case* letters**.

❑ Preparing for Your *Stata* Session

Before beginning your *Stata* session, you will need to do two things.

1. Download the text-format data file **mroz.raw** and its description from the course web site. If necessary, consult *Computer Tutorial 1* to refresh your memory on how to do this.

2. Use Windows Explorer to copy the text-format data file **mroz.raw** to the *Stata* ***working directory*** on the C:-drive or D:-drive of the computer at which you are working.

- **On the computers in Dunning 350**, the default *Stata* working directory is usually either **C:\courses** or **C:\Data**.

- **On the computers in MC B109/B111**, the default *Stata* working directory is usually **D:\courses**.

❑ Start Your *Stata* Session

To start your *Stata* session, double-click on the **Intercooled Stata icon** on the Windows desktop.

After you double-click the **Intercooled Stata icon**, you will see the familiar screen of four *Stata* windows.

❑ Record Your *Stata* Session -- log using

To record your *Stata* session, including all the *Stata* commands you enter and the results (output) produced by these commands, make a text-format **.log** file named **452tutr6.log**. To open (begin) the log file **452tutr6.log**, enter in the Command window:

```
log using 452tutr6.log
```

This command opens a text-format (ASCII) file called **452tutr6.log** in the current *Stata* working directory.

Note: It is important to include the **.log** file extension when opening a log file; if you do not, your log file will be in smcl format, a format that only Intercooled Stata 7 can read. Once you have opened the **452tutr6.log** file, a copy of all the commands you enter during your *Stata* session and of all the results they produce is recorded in that **452tutr6.log** file.

❑ Load the Text-Format Data File *mroz.raw* into *Stata* -- infile

Before starting your *Stata* session, you downloaded the text-format (or ASCII-format) data file **mroz.raw** from the ECON452A web site and placed it in a directory or folder on the C:-drive of your computer. This section demonstrates how to input that data file into *Stata* using the **infile** command.

- To load the text-format data file **mroz.raw** into memory, type in the Command window either of the following two commands:

```
infile inlf hours kidslt6 kidsge6 age educ wage repwage hushrs
husage huseduc huswage faminc mtr motheduc fatheduc unem city
exper nwifeinc lwage expersq using mroz.raw
```

or

```
infile inlf hours kidslt6 kidsge6 age educ wage repwage hushrs
husage huseduc huswage faminc mtr motheduc fatheduc unem city
exper nwifeinc lwage expersq using mroz
```

The data file **mroz.raw** contains 753 observations on 22 variables.

- Using the variable definitions in the description file **mroz.des**, you may wish to assign variable labels to the variables in the data set. Use the **label variable** command for this purpose. To refresh your memory on the **label variable** command, consult *Computer Tutorial 1*.

- To save some keystrokes in this exercise, I have renamed some of the variables in the original **mroz.raw** data file. In particular, I have used the following new variable names: ed = educ; exp = exper; expsq = expersq. Use the **rename** command to assign the new names ed, exp and expsq to the original variable names educ, exper and expersq, respectively. Enter the following **rename** commands:

```
rename educ ed
rename exper exp
rename expersq expsq
```

- To save the data in memory as a *Stata*-format dataset named **mroz**, enter the following **save** command:

```
save mroz
```

This command saves on disk in the *Stata* working directory the *Stata*-format dataset **mroz.dta**, but leaves unchanged the original text-format data file **mroz.raw**.

❑ Familiarize yourself with the current data set -- *describe*, *summarize*, *codebook*

- **To summarize the contents of the current dataset**, use the **describe** command. Recall from *Computer Tutorial 1* that the **describe** command displays a summary of the contents of the current dataset in memory, which in this case is the *Stata*-format data file **mroz.dta**. Enter the command:

```
describe
```

- **To compute summary statistics for the variables in the current dataset**, use the **summarize** command. Recall from *Computer Tutorial 1* that the **summarize** command computes descriptive summary statistics for all *numeric* variables in the current dataset in memory. Enter the command:

```
summarize
```

- To display summary statistics only for the variables that are used in this exercise, enter the command:

```
summarize inlf nwifeinc ed exp expsq age kidslt6 kidsge6
```

Note that the variable **inlf** is a binary variable that takes only the two values 0 and 1. It is the dependent variable in the probit models estimated in this exercise.

- **To obtain more detailed summary statistics** on the variables used in this exercise, enter the following **codebook** command:

```
codebook inlf nwifeinc ed exp expsq age kidslt6 kidsge6
```

❑ Maximum Likelihood Estimation of Probit Models -- *probit*

A Generic Probit Model

$$Y_i^* = x_i^T\beta + u_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i \qquad (1)$$

where:

Y_i^* = a continuous real-valued index variable for observation i that is *unobservable*, or *latent*;

x_i^T = $(1\ X_{i1}\ X_{i2}\ \cdots\ X_{ik})$, a 1×K row vector of regressor values for observation i;

β = $(\beta_0\ \beta_1\ \beta_2\ \cdots\ \beta_k)^T$, a K×1 column vector of regression coefficients;

u_i = an iid $N(0, \sigma^2)$ random error term for observation i.

- The ***observable outcomes* of the binary choice problem** are represented by a **binary indicator variable Y_i** that is related to the unobserved dependent variable Y_i^* as follows:

$$Y_i = 1 \text{ if } Y_i^* > 0 \qquad (2.1)$$
$$Y_i = 0 \text{ if } Y_i^* \le 0 \qquad (2.2)$$

The **random indicator variable Y_i** represents the observed realizations of a binomial process with the following probabilities:

$$\Pr(Y_i = 1) = \Pr(Y_i^* > 0) = \Pr(x_i^T\beta + u_i > 0) \qquad (3.1)$$

$$\Pr(Y_i = 0) = \Pr(Y_i^* \le 0) = \Pr(x_i^T\beta + u_i \le 0) \qquad (3.2)$$

- The **binomial probabilities** $\Pr(Y_i = 1) = \Pr(Y_i^* > 0)$ and $\Pr(Y_i = 0) = \Pr(Y_i^* \leq 0)$ are represented in terms of the c.d.f. (cumulative distribution function) for the random variable u_i as follows:

$$\Pr(Y_i = 1) = \Pr(Y_i^* > 0) = \Phi(x_i^T \beta) \quad (4.1)$$
$$\Pr(Y_i = 0) = \Pr(Y_i^* \leq 0) = 1 - \Phi(x_i^T \beta) \quad (4.2)$$

where $\Phi(x_i^T \beta)$ is the c.d.f. for the standard normal distribution.

Model 1

The following is a model of married women's labour force participation, where the observed binary dependent variable ***inlf***$_i$ is defined as follows:

inlf$_i$ = 1 if the i-th married woman is in the employed labour force
= 0 if the i-th married woman is not in the employed labour force

The ***probit index function***, or ***regression function***, for model 1 is:

$$x_i^T \beta = \beta_0 + \beta_1 \text{nwifeinc}_i + \beta_2 \text{ed}_2 + \beta_3 \text{exp}_i + \beta_4 \text{exp}_i^2 + \beta_5 \text{age}_i + \beta_6 \text{kidslt6}_i + \beta_7 \text{kidsge6}_i$$

where

nwifeinc_i	= non-wife family income of the i-th woman (in thousands of dollars per year);
ed_i	= years of formal education of the i-th woman (in years);
exp_i	= years of actual work experience of the i-th woman (in years);
age_i	= age of the i-th woman (in years);
kidslt6_i	= number of children less than 6 years of age for the i-th woman;
kidsge6_i	= number of children 6 years of age or older for the i-th woman.

This binary dependent variable model is used extensively as an example in the Wooldridge textbook. See Wooldridge (2000): Sec. 7.5, pp. 232-36; Example 8.8, p. 271; and Example 17.1, pp. 537-39.

- To compute ML estimates of probit model 1 and display the full set of saved results, enter the following commands:

```
probit inlf nwifeinc ed exp expsq age kidslt6 kidsge6
estimate list
```

- To save as a *column* vector the ML probit coefficient estimates $\hat{\beta}$ for model 1 and display the results, enter the following **matrix** commands:

```
matrix bhat = e(b)'
matrix list bhat
```

❑ Marginal *index* effects of continuous explanatory variables -- *test* and *lincom*

Background

- The **ML probit coefficient estimates** $\hat{\beta}$ are the partial derivatives of the estimated **probit index function** $x_i^T\hat{\beta}$ with respect to the individual regressors:

$$\hat{\beta}_j = \frac{\partial x_i^T\hat{\beta}}{\partial X_{ij}}$$

- The ***estimated*** **probit index function for model 1** is written in general as:

$$x_i^T\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 \text{nwifeinc}_i + \hat{\beta}_2 \text{ed}_2 + \hat{\beta}_3 \text{exp}_i + \hat{\beta}_4 \text{exp}_i^2 + \hat{\beta}_5 \text{age}_i + \hat{\beta}_6 \text{kidslt6}_i + \hat{\beta}_7 \text{kidsge6}_i$$

- The **ML probit coefficient estimates** for model 1 are:

$$\hat{\beta}_1 = \frac{\partial x_i^T\hat{\beta}}{\partial \text{nwifeinc}_i}; \quad \hat{\beta}_2 = \frac{\partial x_i^T\hat{\beta}}{\partial \text{ed}_i}; \quad \hat{\beta}_3 = \frac{\partial x_i^T\hat{\beta}}{\partial \text{exp}_i}; \quad \hat{\beta}_4 = \frac{\partial x_i^T\hat{\beta}}{\partial \text{exp}_i^2}$$

$$\hat{\beta}_5 = \frac{\partial x_i^T\hat{\beta}}{\partial \text{age}_i}; \quad \hat{\beta}_6 = \frac{\partial x_i^T\hat{\beta}}{\partial \text{kidslt6}_i}; \quad \hat{\beta}_7 = \frac{\partial x_i^T\hat{\beta}}{\partial \text{kidsge6}_i}.$$

Notice that there is something very odd about the above interpretation of the probit coefficient estimates $\hat{\beta}_3$ and $\hat{\beta}_4$. What does it mean to vary exp_i while holding exp_i^2 constant, and vice versa?

The answer is that the marginal index effect of the explanatory variable exp_i in model 1 is not a constant, and therefore is not estimated by a single slope coefficient estimate.

- The **marginal *index* effects of the individual explanatory variables** on the latent (unobserved) dependent variable Y_i^* -- i.e., on the probit index function $x_i^T\beta$ -- equal their respective probit coefficients only for *continuous* variables that enter the index function *linearly*.

  In model 1, all the explanatory variables *except* exp_i enter the probit index function linearly. For these explanatory variables, the marginal index effects equal their respective probit coefficients:

  1. marginal index effect of $\text{nwifeinc}_i = \dfrac{\partial x_i^T\beta}{\partial \text{nwifeinc}_i} = \beta_1$
  2. marginal index effect of $\text{ed}_i = \dfrac{\partial x_i^T\beta}{\partial \text{ed}_i} = \beta_2$
  3. marginal index effect of $\text{age}_i = \dfrac{\partial x_i^T\beta}{\partial \text{age}_i} = \beta_5$
  4. marginal index effect of $\text{kidslt6}_i = \dfrac{\partial x_i^T\beta}{\partial \text{kidslt6}_i} = \beta_6$
  5. marginal index effect of $\text{kidsge6}_i = \dfrac{\partial x_i^T\beta}{\partial \text{kidsge6}_i} = \beta_7$

  But the explanatory variable exp_i enters the index function $x_i^T\beta$ *nonlinearly*. Consequently, the marginal index effect of exp_i does not correspond to a single slope coefficient. In model 1, the marginal index effect of exp_i is a linear function of exp_i with coefficients β_3 and β_4:

  6. marginal index effect of $\text{exp}_i = \dfrac{\partial x_i^T\beta}{\partial \text{exp}_i} = \beta_3 + 2\beta_4 \text{exp}_i$

Computing and Testing Marginal Index Effects

- The **test** command can be used after a **probit** estimation command to compute ***large sample Wald tests*** of the marginal index effects of individual explanatory variables. The ***Wald statistic*** computed by the **test** command after **probit** has a ***chi-square distribution*** with degrees of freedom equal to q, the number of linear restrictions specified by the null hypothesis H_0:

  $$W \overset{a}{\sim} \chi^2[q] \text{ under } H_0, \text{ where q = number of restrictions specified by } H_0.$$

- The **lincom** command can be used after a **probit** estimation command to compute the estimated values of the marginal index effects of individual explanatory variables, and to perform ***two-tail asymptotic t-tests*** of the null hypothesis that these marginal index effects equal zero. The asymptotic null distribution of the t-statistic is N(0, 1), the standard normal distribution (which is what the t distribution converges to as degrees of freedom, or sample size, becomes indefinitely large):

  $$t \overset{a}{\sim} N(0,1) \text{ under } H_0.$$

  In *Stata 7*, the sample value of the ***t-statistics*** produced by the **lincom** command (and by the probit estimation commands) is labelled as ***z***, but is computed in exactly the same way as any t-statistic.

  Here are some examples of hypothesis tests on the estimated probit index function of model 1.

- ***Test 1 - Model 1:***

- ***Proposition:*** The non-wife income of the household has no effect on married women's labour force participation; the marginal index effect of $nwifeinc_i$ equals zero.

- ***Null and Alternative Hypotheses:***

  $H_0: \beta_1 = 0$
  $H_1: \beta_1 \neq 0$

- To calculate a **Wald test** of this hypothesis and the corresponding p-value for the calculated W-statistic, enter the following **test** and **return list** commands:

```
test nwifeinc     or    test nwifeinc = 0
return list
display sqrt(r(chi2))
```

- To calculate a **two-tail asymptotic t-test** of H_0 against H_1, enter the following **lincom** and **return list** commands:

```
lincom nwifeinc
return list
```

  Compare the results of this two-tail t-test with those of the previous Wald test. You should be able to explain why the two tests are equivalent.

- An alternative way to have *Stata* calculate a **two-tail asymptotic t-test** of this hypothesis is to enter the following **lincom** command:

```
lincom _b[nwifeinc]
```

  Note that this form of the lincom command uses **`_b[nwifeinc]`** rather than **`nwifeinc`** to refer to the ML probit coefficient estimate $\hat{\beta}_1$ of the regressor nwifeinc_i.

- ***Test 2 - Model 1:***

- ***Proposition:*** A married woman's education has no effect on the probability that she is a labour force participant; the marginal index effect of ed_i equals zero.

- ***Null and Alternative Hypotheses:***

  H_0: $\beta_2 = 0$
  H_1: $\beta_2 \neq 0$

- To calculate a **Wald test** of this hypothesis and the corresponding p-value for the calculated W-statistic, enter either of the following **test** commands:

```
test ed      or    test ed = 0
```

- To calculate a **two-tail asymptotic t-test** of H_0 against H_1, enter the following **lincom** command:

```
lincom ed
```

Note that this **lincom** command simply replicates the sample value of the z-statistic and its two-tail p-value in the output of the **probit** command for model 1.

- ***Test 3 - Model 1:***

- ***Proposition:*** Years of actual work experience have no effect on the probability of labour force participation for married women. In other words, the marginal index effect of exp_i equals zero, where

$$\text{marginal index effect of } exp_i = \frac{\partial x_i^T \beta}{\partial exp_i} = \beta_3 + 2\beta_4 exp_i .$$

- A sufficient condition for the marginal index effect of exp_i to equal zero for all married women is $\beta_3 = 0$ and $\beta_4 = 0$.

- ***Null and Alternative Hypotheses:***

H_0: $\beta_3 = 0$ *and* $\beta_4 = 0$
H_1: $\beta_3 \neq 0$ *and/or* $\beta_4 \neq 0$

- To calculate a **Wald test** of this hypothesis and the corresponding p-value for the calculated W-statistic, enter the following **test** commands:

```
test exp = 0
test expsq = 0, accum
```

- ***Test 4 - Model 1:***

- ***Proposition:*** The number and ages of dependent children in the family have no effect on the probability of married women's labour force participation. In other words, the marginal index effects of both ***kidslt6$_i$*** and ***kidsge6$_i$*** equal zero, where

$$\text{marginal index effect of kidslt6}_i = \frac{\partial\, x_i^T\beta}{\partial\, \text{kidslt6}_i} = \beta_6;$$

$$\text{marginal index effect of kidsge6}_i = \frac{\partial\, x_i^T\beta}{\partial\, \text{kidsge6}_i} = \beta_7.$$

- ***Null and Alternative Hypotheses:***

H_0: $\beta_6 = 0$ *and* $\beta_7 = 0$
H_1: $\beta_6 \neq 0$ *and/or* $\beta_7 \neq 0$

- To calculate a **Wald test** of this hypothesis and the corresponding p-value for the calculated W-statistic, enter the following **test** commands:

```
test kidslt6 = 0
test kidsge6 = 0, accum
```

❑ Marginal *probability* effects of *continuous* explanatory variables -- *dprobit*

Background

- The **marginal *probability* effects of *continuous* explanatory variables in probit models** are the partial derivatives of the standard normal c.d.f. $\Phi\left(x_i^T\beta\right)$ with respect to the individual explanatory variables:

$$\textbf{marginal } \textit{probability} \textbf{ effect of } \mathbf{X_j} = \frac{\partial\, \Phi\left(x_i^T\beta\right)}{\partial\, X_{ij}} = \frac{\partial\, \Phi\left(x_i^T\beta\right)}{\partial\, x_i^T\beta}\frac{\partial\, x_i^T\beta}{\partial\, X_{ij}} = \phi\left(x_i^T\beta\right)\frac{\partial\, x_i^T\beta}{\partial\, X_{ij}}$$

where

$\phi\left(x_i^T\beta\right) =$ the value of the standard normal p.d.f. at $x_i^T\beta$

$\frac{\partial x_i^T\beta}{\partial X_{ij}} =$ the marginal index effect of X_j

- For model 1, we have previously seen that the **marginal *index* effects** of the explanatory variables are:

1. marginal index effect of $\text{nwifeinc}_i = \frac{\partial x_i^T\beta}{\partial \text{nwifeinc}_i} = \beta_1$
2. marginal index effect of $\text{ed}_i = \frac{\partial x_i^T\beta}{\partial \text{ed}_i} = \beta_2$
3. marginal index effect of $\text{age}_i = \frac{\partial x_i^T\beta}{\partial \text{age}_i} = \beta_5$
4. marginal index effect of $\text{kidslt6}_i = \frac{\partial x_i^T\beta}{\partial \text{kidslt6}_i} = \beta_6$
5. marginal index effect of $\text{kidsge6}_i = \frac{\partial x_i^T\beta}{\partial \text{kidsge6}_i} = \beta_7$
6. marginal index effect of $\text{exp}_i = \frac{\partial x_i^T\beta}{\partial \text{exp}_i} = \beta_3 + 2\beta_4 \text{exp}_i$.

- The **marginal *probability* effects** of the explanatory variables in model 1 are:

1. marginal probability effect of $\text{nwifeinc}_i = \frac{\partial \Phi\left(x_i^T\beta\right)}{\partial \text{nwifeinc}_i} = \phi\left(x_i^T\beta\right)\beta_1$
2. marginal probability effect of $\text{ed}_i = \frac{\partial \Phi\left(x_i^T\beta\right)}{\partial \text{ed}_i} = \phi\left(x_i^T\beta\right)\beta_2$
3. marginal probability effect of $\text{age}_i = \frac{\partial \Phi\left(x_i^T\beta\right)}{\partial \text{age}_i} = \phi\left(x_i^T\beta\right)\beta_5$

4. marginal probability effect of $\text{kidslt6}_i = \dfrac{\partial \Phi(x_i^T\beta)}{\partial \text{kidslt6}_i} = \phi(x_i^T\beta)\beta_6$

5. marginal probability effect of $\text{kidsge6}_i = \dfrac{\partial \Phi(x_i^T\beta)}{\partial \text{kidsge6}_i} = \phi(x_i^T\beta)\beta_7$

6. marginal probability effect of $\text{exp}_i = \dfrac{\partial \Phi(x_i^T\beta)}{\partial \text{exp}_i} = \phi(x_i^T\beta)(\beta_3 + 2\beta_4 \text{exp}_i)$

- Note that the marginal probability effect of a continuous explanatory variable X_j is proportional to the marginal index effect of X_j, where the factor of proportionality is the standard normal p.d.f. at $x_i^T\beta$:

  marginal *probability* effect of X_j $= \phi(x_i^T\beta) \times$ **marginal *index* effect of X_j**

The *dprobit* command

- The **dprobit** command is an alternative to **probit** for computing ML estimates of probit models. However, **dprobit** differs from **probit** in that it reports the ***marginal probability effects*** of the individual regressors (and their standard errors) rather than the probit coefficient estimates $\hat{\beta}_j$ (and their standard errors).

- To obtain ML estimates of the marginal probability effects for model 1, evaluated at the sample mean values of the regressors

  $\bar{x}^T = (\overline{\text{nwifeinc}}\ \overline{\text{ed}}\ \overline{\text{exp}}\ \overline{\text{expsq}}\ \overline{\text{age}}\ \overline{\text{kidslt6}}\ \overline{\text{kidsge6}}\ 1)$,

  enter the commands:

```
dprobit inlf nwifeinc ed exp expsq age kidslt6 kidsge6
estimate list
```

  Carefully inspect the results of these commands.

- To compare the output of the above **dprobit** command with that of the corresponding **probit** command, enter the command:

```
probit
```

Note that the values of the t-statistics (labelled as z) and their two-tail p-values (labelled $P > |z|$) are identical in the output of the **probit** and **dprobit** commands. That's because the reported t-statistics are testing the same hypothesis, namely the null hypothesis that the probit coefficient $\beta_j = 0$ against the two-sided alternative hypothesis that $\beta_j \neq 0$.

The *dprobit* command with the *at(matname)* option

- An alternative way to obtain ML estimates of marginal probability effects in probit models is to use the **at(*matname*)** option on the **dprobit** command, where ***matname*** is the name of a user-defined vector that contains the values of the regressors at which the estimates of the marginal probability effects are evaluated. Here is how to use this option.

- First, use **summarize** and **scalar** commands to compute and save the sample means of the regressors. Enter the commands:

```
summarize nwifeinc, meanonly
scalar nwincbar = r(mean)
summarize ed, meanonly
scalar edbar = r(mean)
summarize exp, meanonly
scalar expbar = r(mean)
summarize expsq, meanonly
scalar expsqbar = r(mean)
summarize age, meanonly
scalar agebar = r(mean)
summarize kidslt6, meanonly
scalar klt6bar = r(mean)
summarize kidsge6, meanonly
scalar kge6bar = r(mean)
scalar list nwincbar edbar expbar expsqbar agebar klt6bar
kge6bar
```

- Next, use **matrix** commands to create and display a row vector **xbar** that contains the sample mean values of the regressors and a 1 for the intercept coefficient. Enter the commands:

```
matrix xbar = (nwincbar, edbar, expbar, expsqbar, agebar,
klt6bar, kge6bar, 1)
matrix list xbar
```

- Now use the **dprobit** command with the **at(*xbar*)** option to compute ML estimates of the marginal probability effects for model 1 and their standard errors at the sample mean values of the regressors. Enter the command:

```
dprobit inlf nwifeinc ed exp expsq age kidslt6 kidsge6, at(xbar)
```

  The output of this command is identical to that for the **dprobit** command without the **at(*xbar*)** option; the default option for **dprobit** is to compute marginal probability effects at the sample mean values of the regressors.

Marginal probability effects for explanatory variables that enter *nonlinearly* in the probit index function

- The only explanatory variable for which the above **dprobit** command does not produce an estimate of the true marginal probability effect is exp_i, years of actual work experience. Recall that the **correct marginal probability effect of exp_i** is:

$$\text{marginal probability effect of } \text{exp}_i = \frac{\partial \Phi\left(\mathbf{x}_i^T \beta\right)}{\partial \text{exp}_i} = \phi\left(\mathbf{x}_i^T \beta\right)(\beta_3 + 2\beta_4 \text{exp}_i).$$

- Evaluated at the sample mean values of the regressors in the vector

$$\bar{\mathbf{x}}^T = (1 \ \overline{\text{nwifeinc}} \ \overline{\text{ed}} \ \overline{\text{exp}} \ \overline{\text{expsq}} \ \overline{\text{age}} \ \overline{\text{kidslt6}} \ \overline{\text{kidsge6}}),$$

  the **marginal probability effect of exp_i** is:

$$\text{marginal probability effect of } \text{exp}_i \text{ at } \bar{\mathbf{x}}^T = \phi\left(\bar{\mathbf{x}}^T \beta\right)(\beta_3 + 2\beta_4 \overline{\text{exp}}).$$

- The **ML estimate of the marginal probability effect of $\mathbf{exp_i}$** at $\bar{x}^T$ is:

  estimate of marginal probability effect of exp_i at $\bar{x}^T = \phi(\bar{x}^T\hat{\beta})(\hat{\beta}_3 + 2\hat{\beta}_4\overline{exp})$.

The following steps demonstrate how to compute $\phi(\bar{x}^T\hat{\beta})(\hat{\beta}_3 + 2\hat{\beta}_4\overline{exp})$.

- First, re-estimate the probit coefficient estimates of model 1. Enter the **probit** command:

```
probit inlf nwifeinc ed exp expsq age kidslt6 kidsge6
```

- Second, compute the estimated value of the index function $\bar{x}^T\hat{\beta}$. Enter the following commands:

```
matrix xbarbhat = xbar*bhat
matrix list xbarbhat
scalar xbarb = el(xbarbhat, 1, 1)
scalar list xbarb
```

- Third, compute the value of the standard normal p.d.f. at $\bar{x}^T\hat{\beta}$, i.e., the value of $\phi(\bar{x}^T\hat{\beta})$. Enter the following **scalar** commands:

```
scalar phixbarb = normden(xbarb)
scalar list phixbarb
```

- Fourth, compute and display the value of $\phi(\bar{x}^T\hat{\beta})(\hat{\beta}_3 + 2\hat{\beta}_4\overline{exp})$. Enter the following **display** command:

```
display phixbarb*(_b[exp] + 2*_b[expsq]*expbar)
```

- Alternatively, use the **lincom** command to compute and display the value of $\phi(\bar{x}^T\hat{\beta})(\hat{\beta}_3 + 2\hat{\beta}_4\overline{exp})$. Enter the following commands:

```
lincom phixbarb*(_b[exp] + 2*_b[expsq]*expbar)
lincom phixbarb*(exp + 2*expsq*expbar)
```

Note that these two **lincom** commands produce exactly the same output.

Another example of the *dprobit* command with the *at(matname)* option

- The **dprobit** command with the **at(*matname*)** option can be used to compute ML estimates of the marginal probability effects for model 1 at ***any user-selected*** **values of the regressors**. This section demonstrates how to do this.

- First, use a **matrix** command to compute and save the selected values of the regressors. The selected regressor values are nwifeinc = 20, ed = 14, exp = 10, expsq = 100, age = 40, kidslt6 = 0, and kidsge6 = 1. In other words, we are going to estimate the marginal probability effects in model 1 for married women whose non-wife family income is $20,000 per year, who have 14 years of formal education and 10 years of actual work experience, who are 40 years of age, who have no dependent children under 6 years of age, and who have one dependent child 6 years of age or older. Enter the command:

```
matrix xvec = (20, 14, 10, 100, 40, 0, 1, 1)
matrix list xvec
```

  Note that the value 1 is included as the last element of the regressor vector named ***xvec*** to allow for the probit intercept coefficient β_0.

- Now use the **dprobit** command with the **at(*xvec*)** option to compute ML estimates of the marginal probability effects in model 1 for married women with the observed characteristics described by the regressor values in the vector ***xvec***. Enter the command:

```
dprobit inlf nwifeinc ed exp expsq age kidslt6 kidsge6, at(xvec)
```

  Compare the output of this **dprobit** command with that of the earlier **dprobit** command. The estimates of the marginal probability effects are different because they are evaluated at different values of the regressor vector x_i^T.

❑ Marginal *probability* effects of continuous explanatory variables -- *mfx compute*

- There is another *Stata 7* command that can be used to compute the marginal probability effects of individual regressors following probit estimation. This is the **mfx compute** command.

The **mfx compute** command is a general *Stata 7* command for computing marginal effects following estimation commands.

When used after a **probit** command, the **mfx compute** command computes the marginal probability effects of individual regressors at specified values of the regressors. It also computes correct standard errors for these estimated marginal probability effects, meaning standard errors that account for the fact that $\phi\left(x_0^T\hat{\beta}\right)$ is an estimate of $\phi\left(x_0^T\beta\right)$, where $\hat{\beta}$ is an ML estimate of the probit coefficient vector β.

- To begin, re-estimate probit model 1 by ML. Enter the **probit** command:

```
probit inlf nwifeinc ed exp expsq age kidslt6 kidsge6
```

- Compute the marginal probability effects of the individual regressors in model 1 at the **sample *mean* values of the regressors** $\bar{x}^T$. Enter either of the following commands:

```
mfx compute
mfx compute, at(mean)
```

These two commands are identical, because the default option for the **mfx compute** command is to evaluate the marginal probability effects at the sample means of the regressors.

- Now compute the marginal probability effects of the regressors in model 1 at the **sample *medians* of the regressors**. Enter the following command:

```
mfx compute, at(median)
```

- The **mfx compute** command can also compute marginal probability effects at ***any user-selected values* of the regressors**. To illustrate, we use the same regressor values selected in the previous section on the **dprobit** command, namely: nwifeinc = 20, ed = 14, exp = 10, expsq = 100, age = 40, kidslt6 = 0, and kidsge6 = 1. In other words, we are again going to estimate the marginal probability effects in model 1 for married women whose non-wife family income is $20,000 per year, who have 14 years of formal education and 10 years of actual work experience, who are 40 years of age, and who have no dependent children under 6 years of age and one dependent child 6 years of age or older. Enter the following two commands:

```
mfx compute, at(20 14 10 100 40 0 1 1)
mfx compute, at(nwifeinc=20, ed=14, exp=10, expsq=100, age=40,
kidslt6=0, kidsge6=1)
```

These two commands are identical. They illustrate the two alternative ways in which the **mfx compute** command can be used to compute marginal probability effects in probit models at any set of user-selected values of the regressors.

- **Marginal *probability* effects of binary explanatory variables -- *dprobit* and *mfx compute***

Model 2

All the explanatory variables in model 1 are continuous, or at least take more than two discrete values. But the use of ***dummy variable*** **regressors** -- binary variables that take only the two values 0 and 1 -- is as common in binary dependent variables models as it is in conventional linear regression models. This section explains how *Stata 7* computes estimates of the **marginal probability effects of dummy variable regressors** in probit models.

For this purpose, we replace the explanatory variables ***kidslt6$_i$*** and ***kidsge6$_i$*** in model 1 with a single dummy variable ***dkidslt6$_i$*** defined as follows:

dkidslt6$_i$ = 1 if ***kidslt6$_i$*** **> 0** for the i-th married woman
= 0 if if ***kidslt6$_i$*** **= 0** for the i-th married woman

The **probit index function**, or regression function, **for model 2** is:

$$x_i^T\beta = \beta_0 + \beta_1 \text{nwifeinc}_i + \beta_2 \text{ed}_2 + \beta_3 \text{exp}_i + \beta_4 \text{exp}_i^2 + \beta_5 \text{age}_i + \beta_6 \text{dkidslt6}_i$$

where all other explanatory variables apart from ***dkidslt6$_i$*** are as previously defined for model 1; that is, all other explanatory variables in model 2 are continuous variables.

This section addresses the question of how to compute the marginal probability effect of the dummy explanatory variable ***dkidslt6$_i$***.

Background

Marginal probability effects of continuous explanatory variables

- As we have previously seen, the **marginal *probability* effects of *continuous* explanatory variables in probit models** are the partial derivatives of the standard normal c.d.f. $\Phi\left(x_i^T\beta\right)$ with respect to the individual explanatory variables:

  marginal *probability* effect of X_j $= \dfrac{\partial \Phi\left(x_i^T\beta\right)}{\partial X_{ij}} = \dfrac{\partial \Phi\left(x_i^T\beta\right)}{\partial x_i^T\beta}\dfrac{\partial x_i^T\beta}{\partial X_{ij}} = \phi\left(x_i^T\beta\right)\dfrac{\partial x_i^T\beta}{\partial X_{ij}}$

  where

  $\phi\left(x_i^T\beta\right) =$ the value of the standard normal p.d.f. at $x_i^T\beta$

  $\dfrac{\partial x_i^T\beta}{\partial X_{ij}} =$ the marginal index effect of X_j, where X_j is a continuous variable.

- The **marginal *probability* effects** of the ***continuous*** **explanatory variables in model 2** are therefore:

  1. marginal probability effect of $\text{nwifeinc}_i = \dfrac{\partial \Phi\left(x_i^T\beta\right)}{\partial \text{nwifeinc}_i} = \phi\left(x_i^T\beta\right)\beta_1$
  2. marginal probability effect of $\text{ed}_i = \dfrac{\partial \Phi\left(x_i^T\beta\right)}{\partial \text{ed}_i} = \phi\left(x_i^T\beta\right)\beta_2$
  3. marginal probability effect of $\text{exp}_i = \dfrac{\partial \Phi\left(x_i^T\beta\right)}{\partial \text{exp}_i} = \phi\left(x_i^T\beta\right)(\beta_3 + 2\beta_4\,\text{exp}_i)$
  4. marginal probability effect of $\text{age}_i = \dfrac{\partial \Phi\left(x_i^T\beta\right)}{\partial \text{age}_i} = \phi\left(x_i^T\beta\right)\beta_5$

Marginal probability effects of dummy explanatory variables

- The **marginal *probability* effect of a *dummy*, or *binary*, explanatory variable in probit models** is computed by evaluating the standard normal c.d.f. $\Phi\left(x_i^T\beta\right)$ at the two different values of the dummy variable D_j, and then taking the difference.

- Define two different vectors of regressor values:

  x_{1i}^T = any vector of regressor values with $D_j = 1$;
  x_{0i}^T = the same vector of regressor values but with $D_j = 0$.

- The **marginal probability effect of the dummy variable D_j** is defined as:

  marginal *probability* effect of D_j $= \Phi\left(x_{1i}^T\beta\right) - \Phi\left(x_{0i}^T\beta\right)$.

- ***Example:*** Compute the **marginal probability effect of the dummy variable *dkidslt6$_i$* in model 2.**

  First, define x_{1i}^T with the continuous variables set to any arbitrary values and the dummy variable $\text{dkidslt6}_i = 1$:

  $$x_{1i}^T = \left(1 \;\; \text{nwifeinc}_i \;\; \text{ed}_i \;\; \text{exp}_i \;\; \text{exp}_i^2 \;\; \text{age}_i \;\; 1\right)$$

  Second, define x_{0i}^T with the continuous variables set to the *same* arbitrary values but with the dummy variable $\text{dkidslt6}_i = 0$:

  $$x_{0i}^T = \left(1 \;\; \text{nwifeinc}_i \;\; \text{ed}_i \;\; \text{exp}_i \;\; \text{exp}_i^2 \;\; \text{age}_i \;\; 0\right)$$

  The **conditional probability that *inlf$_i$* = 1 when *dkidslt6$_i$* = 1** is:

  $$\begin{aligned}\Pr\left(\text{inlf}_i = 1 \middle| \text{dkidslt6}_i = 1\right) &= \Phi\left(x_{1i}^T\beta\right) \\ &= \Phi\left(\beta_0 + \beta_1\text{nwifeinc}_i + \beta_2\text{ed}_2 + \beta_3\,\text{exp}_i + \beta_4\,\text{exp}_i^2 + \beta_5\text{age}_i + \beta_6\right)\end{aligned}$$

The **conditional probability that $inlf_i = 1$ when $dkidslt6_i = 0$** is:

$$\begin{aligned}\Pr\left(\text{inlf}_i = 1 \middle| \text{dkidslt6}_i = 0\right) &= \Phi\left(\text{x}_{0i}^{\text{T}}\beta\right) \\ &= \Phi\left(\beta_0 + \beta_1 \text{nwifeinc}_i + \beta_2 \text{ed}_2 + \beta_3 \text{exp}_i + \beta_4 \text{exp}_i^2 + \beta_5 \text{age}_i\right)\end{aligned}$$

The **marginal *probability* effect of the dummy variable $dkidslt6_i$ in model 2 is:**

$$\begin{aligned}&\Pr\left(\text{inlf}_i = 1 \middle| \text{dkidslt6}_i = 1\right) - \Pr\left(\text{inlf}_i = 1 \middle| \text{dkidslt6}_i = 0\right) \\ &= \Phi\left(\text{x}_{1i}^{\text{T}}\beta\right) - \Phi\left(\text{x}_{0i}^{\text{T}}\beta\right) \\ &= \Phi\left(\beta_0 + \beta_1 \text{nwifeinc}_i + \beta_2 \text{ed}_2 + \beta_3 \text{exp}_i + \beta_4 \text{exp}_i^2 + \beta_5 \text{age}_i + \beta_6\right) \\ &\quad - \Phi\left(\beta_0 + \beta_1 \text{nwifeinc}_i + \beta_2 \text{ed}_2 + \beta_3 \text{exp}_i + \beta_4 \text{exp}_i^2 + \beta_5 \text{age}_i\right)\end{aligned}$$

ML Estimation of Model 2 -- *probit*

- Before estimating model 2, it is necessary to define the new dummy variable ***$dkidslt6_i$***. Enter the commands:

```
generate dkidslt6 = kidslt6 > 0
summarize dkidslt6 kidslt6
tab2 dkidslt6 kidslt6
```

- Compute ML estimates of model 2. Enter the **probit** command:

```
probit inlf nwifeinc ed exp expsq age dkidslt6
```

Marginal Probability Effects for Model 2 -- *mfx compute*

- Compute the marginal probability effects in model 2 at the **sample *mean* values of the regressors** $\bar{\text{x}}^{\text{T}}$. Enter either of the following commands:

```
mfx compute
mfx compute, at(mean)
```

Note again that these two commands are identical. Note in particular the value of the marginal probability effect for the dummy explanatory variable ***$dkidslt6_i$***.

- Compute the marginal probability effects of the regressors in model 2 at the **sample *medians* of the regressors**. Enter the following command:

```
mfx compute, at(median)
```

- Now use the **mfx compute** command to compute the marginal probability effects in model 2 for married women whose non-wife family income is $20,000 per year, who have 14 years of formal education and 10 years of actual work experience, who are 40 years of age, and who have ***one or more*** **dependent children under 6 years of age**. Enter the following two commands:

```
mfx compute, at(20 14 10 100 40 1 1)
mfx compute, at(nwifeinc=20, ed=14, exp=10, expsq=100, age=40,
dkidslt6=1)
```

  These two **mfx compute** commands produce identical output.

- Finally, use the **mfx compute** command to estimate the marginal probability effects in model 2 for married women whose non-wife family income is $20,000 per year, who have 14 years of formal education and 10 years of actual work experience, who are 40 years of age, and who have ***no*** **dependent children under 6 years of age**. Enter the following two commands:

```
mfx compute, at(20 14 10 100 40 0 1)
mfx compute, at(nwifeinc=20, ed=14, exp=10, expsq=100, age=40,
dkidslt6=0)
```

  Carefully compare the results of these two commands for married women with ***no*** **dependent children under 6 years of age** with the results of the preceding two **mfx compute** commands for married women with ***one or more*** **dependent children under 6 years of age**. Which marginal probability effects are the same, and which are different?

Marginal Probability Effects for Model 2 -- *dprobit*

- Use the **dprobit** command *without* the **at(*matname*)** option to compute the marginal probability effects in model 2 at the **sample *mean* values of the regressors** $\bar{x}^T$. Enter the following command:

```
dprobit inlf nwifeinc ed exp expsq age dkidslt6
```

  Compare the output for this **dprobit** command with the output for the following **mfx compute** command (which you have previously entered):

```
mfx compute
```

- Use the **dprobit** command *with* the **at(*matname*)** option to compute the marginal probability effects in model 2 for married women whose non-wife family income is $20,000 per year (nwifeinc = 20), who have 14 years of formal education (ed = 14) and 10 years of actual work experience (exp = 10, expsq = 100), who are 40 years of age age = 40), and who have *no* dependent children under 6 years of age (dkidslt6 = 0). First, you will have to create a vector containing the specified values of the regressors for model 2; the **dprobit** command does not permit number lists in the **at( )** option. Enter the following commands:

```
matrix xvec2 = (20, 14, 10, 100, 40, 0, 1)
matrix list xvec2
dprobit inlf nwifeinc ed exp expsq age dkidslt6, at(xvec2)
```

  Carefully compare the output for this **dprobit** command with the output for the following **mfx compute** command (which you have previously entered):

```
mfx compute, at(nwifeinc=20, ed=14, exp=10, expsq=100, age=40,
dkidslt6=0)
```

❑ Preparing to End Your *Stata* Session

Before you end your *Stata* session, you should do two things.

- First, you may want to **save the current data set**. Enter the following **save** command with the **replace** option to save the current data set as *Stata*-format data set **mroz.dta**:

  save mroz, replace

- Second, **close the log file** you have been recording. Enter the command:

  log close

❑ End Your *Stata* Session -- exit

- **To end your *Stata* session**, use the **exit** command. Enter the command:

  exit *or* **exit, clear**

❑ Cleaning Up and Clearing Out

After returning to Windows, you should copy all the files you have used and created during your *Stata* session to your own diskette. These files will be found in the ***Stata working directory***, which is usually **C:\courses** or **C:\Data** on the computers in Dunning 350, and **D:** on the computers in MC B405. There is one file you will want to be sure you have: the *Stata* log file **452tutr6.log**. If you saved the *Stata*-format data set **mroz.dta**, you will want to take it with you as well. Use the Windows 95/98/2000/ME or MS-DOS **copy** command to copy any files you want to keep to your own diskette in the A:-drive.

Finally, **as a courtesy to other users** of the computing classroom, please delete all the files you have used or created from the *Stata* working directory.