Stata 10/11 **Tutorial 6**

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

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

TASKS: *Stata 10/11 Tutorial 6* demonstrates how to use *Stata* to compute **maximum likelihood (ML) estimates of probit models** for binary dependent variables. It introduces you to **hypothesis testing** in probit models, in particular to testing that the **marginal *index* effects** (and hence the **marginal *probability* effects**) of individual explanatory variables are zero. Finally, it demonstrates how to compute the **marginal probability 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 estimates of *probit* **coefficients** in probit models of binary dependent variables.

  dprobit   Used to compute ML estimates of the **marginal *probability* effects** of explanatory variables 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:

  normalden(*z*)   Computes *value of the standard normal density function (p.d.f.)* for a given value *z* of a standard normal random variable.

  normal(*z*)   Computes *value of the standard normal distribution function (c.d.f.)* for a given value *z* of a standard normal random variable.

  invnormal(*p*)   Computes *the inverse of the standard normal distribution function (c.d.f.);* if **normal(*z*) = *p*, then invnormal(*p*) = *z*.**

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 *Stata 10/11 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 **C:\data**.

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

❑ **Start Your *Stata* Session**

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

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

❑ **Record Your *Stata* Session and *Stata* Commands -- log using, cmdlog 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 **452tutorial6.log**. To open (begin) the log file **452tutorial6.log**, enter in the Command window:

```
log using 452tutorial6.log
```

This command opens a text-format (ASCII) file called **452tutorial6.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 Stata can read. Once you have opened the **452tutorial6.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 **452tutorial6.log** file.

To record only the *Stata* commands you type during your *Stata* session, use the *Stata* **cmdlog using** command. To start (open) the command log file **452tutorial6.txt**, enter in the Command window:

```
cmdlog using 452tutorial6
```

This command opens a plain text-format (ASCII) file called **452tutorial6.txt** in the current *Stata* working directory. All commands you enter during your *Stata* session are recorded in this 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 ECON452 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 *Stata 10/11 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, replace
  ```

  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* **and** *summarize*

- <u>**To summarize the contents of the current dataset**</u>, use the **describe** command. Recall from *Stata 10/11 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
  ```

- <u>**To compute summary statistics for the variables in the current dataset**</u>, use the **summarize** command. Recall from *Stata 10/11 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 tutorial, 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 tutorial.

- **To obtain more detailed summary statistics** on the variables used in this tutorial, enter the following **summarize** command with the **detail** option:

  summarize inlf nwifeinc ed exp age kidslt6 kidsge6, detail

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

A Generic Probit Model

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

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:

$$Y_i = 1 \ \text{ if } \ Y_i^* > 0 \tag{2.1}$$
$$Y_i = 0 \ \text{ if } \ Y_i^* \leq 0 \tag{2.2}$$

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

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

$$\Pr(Y_i = 0) = \Pr(Y_i^* \leq 0) = \Pr(x_i^T \beta + u_i \leq 0) \tag{3.2}$$

♦ 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:

$$\Pr(Y_i = 1) = \Pr(Y_i^* > 0) = \Phi(x_i^T \beta) \tag{4.1}$$

$$\Pr(Y_i = 0) = \Pr(Y_i^* \leq 0) = 1 - \Phi(x_i^T \beta) \tag{4.2}$$

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:

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

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 (2009): Sec. 7.5, pp. 246-250; Example 8.8, p. 291; and Example 17.1, pp. 584-586.

♦ 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:

$$x_i^T\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 nwifeinc_i + \hat{\beta}_2 ed_i + \hat{\beta}_3 exp_i + \hat{\beta}_4 exp_i^2 + \hat{\beta}_5 age_i + \hat{\beta}_6 kidslt6_i + \hat{\beta}_7 kidsge6_i$$

♦ The **ML probit coefficient estimates** for Model 1 are:

$$\hat{\beta}_1 = \frac{\partial x_i^T\hat{\beta}}{\partial nwifeinc_i}; \quad \hat{\beta}_2 = \frac{\partial x_i^T\hat{\beta}}{\partial ed_i}; \quad \hat{\beta}_3 = \frac{\partial x_i^T\hat{\beta}}{\partial exp_i}; \quad \hat{\beta}_4 = \frac{\partial x_i^T\hat{\beta}}{\partial exp_i^2}$$

$$\hat{\beta}_5 = \frac{\partial x_i^T\hat{\beta}}{\partial age_i}; \quad \hat{\beta}_6 = \frac{\partial x_i^T\hat{\beta}}{\partial kidslt6_i}; \quad \hat{\beta}_7 = \frac{\partial x_i^T\hat{\beta}}{\partial kidsge6_i}.$$

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.

Marginal *Index* Effects of *Continuous* Explanatory Variables – Model 1

♦ 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 $nwifeinc_i$ = $\dfrac{\partial x_i^T\beta}{\partial nwifeinc_i}$ = β_1

2. marginal index effect of ed_i = $\dfrac{\partial x_i^T\beta}{\partial ed_i}$ = β_2

3. marginal index effect of age_i = $\dfrac{\partial x_i^T\beta}{\partial age_i}$ = β_5

4. marginal index effect of $kidslt6_i$ = $\dfrac{\partial x_i^T\beta}{\partial kidslt6_i}$ = β_6

5. marginal index effect of $kidsge6_i$ = $\dfrac{\partial x_i^T\beta}{\partial kidsge6_i}$ = β_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 exp_i = $\dfrac{\partial x_i^T\beta}{\partial exp_i}$ = $\beta_3 + 2\beta_4 exp_i$

Marginal *Probability* Effects of *Continuous* Explanatory Variables – Model 1

♦ 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

In other words, the marginal probability effect of a continuous explanatory variable X_j is simply the marginal index effect of X_j multiplied by $\phi\left(x_i^T\beta\right)$, the value of the standard normal p.d.f. evaluated at $x_i^T\beta$.

♦ The **marginal *probability* effects** of the explanatory variables in Model 1 are:

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 $\text{age}_i = \dfrac{\partial\,\Phi\left(x_i^T\beta\right)}{\partial\,\text{age}_i} = \phi\left(x_i^T\beta\right)\beta_5$

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

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

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

♦ 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. $\phi\left(x_i^T \beta\right)$ evaluated at $x_i^T \beta$:

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

♦ Note too that estimation of the marginal probability effects of a continuous explanatory variable X_j requires one to choose a specific vector of regressor values x_i^T. Common choices for x_i^T are the **sample *mean*** and **sample *median*** values of the regressors.

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

• 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
ereturn 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
```

Note that the vector of ML probit coefficient estimates $\hat{\beta}$ is displayed in *Stata* format with the intercept coefficient estimate as the last, rather than the first, element. For Model 1, $\hat{\beta}$ is displayed in *Stata* format as

$$\hat{\beta} = \left(\hat{\beta}_1 \quad \hat{\beta}_2 \quad \hat{\beta}_3 \quad \hat{\beta}_4 \quad \hat{\beta}_5 \quad \hat{\beta}_6 \quad \hat{\beta}_7 \quad \hat{\beta}_0 \right)^T$$

❑ **Testing marginal effects of *continuous* explanatory variables -- *test* and *lincom***

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 a **probit** command has the following form:

$$W_{ML} = \left(R\hat{\beta}_{ML} - r\right)^T \left(R\hat{V}_{ML}R^T\right)^{-1}\left(R\hat{\beta}_{ML} - r\right)$$

where

$\hat{\beta}_{ML}$ = the ML estimator of the probit coefficient vector β;

$\hat{V}_{ML} = \hat{V}_{ML}\left(\hat{\beta}_{ML}\,|\,X\right)$ = the ML estimator of $V_{ML}\left(\hat{\beta}_{ML}\,|\,X\right)$, the variance-covariance matrix of $\hat{\beta}_{ML}$.

The **null distribution of the W statistic** is a *chi-square distribution* with degrees of freedom equal to q, the number of linear restrictions specified by H_0:

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

♦   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):

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

In *Stata*, the sample value of the *t-statistic* 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.

This section presents 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 family has no effect on married women's labour force participation; the marginal index effect and marginal probability effect of nwifeinc$_i$ equal zero.

- ### *Marginal Index and Probability Effects:*

  marginal index effect of nwifeinc$_i$ $= \dfrac{\partial\, x_i^T \beta}{\partial\, \text{nwifeinc}_i} = \beta_1$

  marginal probability effect of nwifeinc$_i$ $= \dfrac{\partial\, \Phi(x_i^T \beta)}{\partial\, \text{nwifeinc}_i} = \phi(x_i^T \beta)\beta_1$

  A sufficient condition for both the marginal index and marginal probability effects of nwifeinc$_i$ to equal zero for all married women is $\beta_1 = 0$.

- ### *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 _b[nwifeinc]     or    lincom nwifeinc
  return list
  ```

  Note that the **lincom** command uses either **_b[nwifeinc]** or **nwifeinc** to refer to the ML probit coefficient estimate $\hat{\beta}_1$ of the regressor nwifeinc$_i$. 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.

♦ *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 and probability effects of ed_i equal zero.

♦ *Marginal Index and Probability Effects:*

$$\text{marginal index effect of } ed_i = \frac{\partial\, x_i^T \beta}{\partial\, ed_i} = \beta_2$$

$$\text{marginal probability effect of } ed_i = \frac{\partial\, \Phi\left(x_i^T \beta\right)}{\partial\, ed_i} = \phi\left(x_i^T \beta\right)\beta_2$$

A sufficient condition for both the marginal index and marginal probability effects of ed_i to equal zero for all married women is $\beta_2 = 0$.

♦ *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 _b[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 probability (and index) effect of exp_i equals zero.

♦ *Marginal Index and Probability Effects:*

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

$$\text{marginal probability effect of } exp_i = \frac{\partial \Phi(x_i^T \beta)}{\partial exp_i} = \phi(x_i^T \beta)(\beta_3 + 2\beta_4 \, exp_i)$$

A sufficient condition for both the marginal index and marginal probability effects 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:* For married women with five years of actual work experience, years of actual work experience have no effect on the probability of labour force participation. In other words, the marginal probability (and index) effects of exp_i equals zero for married women for whom $exp_i = 5$, where

$$\text{marginal } \textit{index} \text{ effect of } exp_i \text{ for } exp_i = 5 = \beta_3 + 2\beta_4 \, exp_i = \beta_3 + 2\beta_4 5$$

marginal *probability* effect of $\exp_i$ for $\exp_i = 5$ $= \phi\left(x_i^T\beta\right)(\beta_3 + 2\beta_4 \exp_i)$
$$= \phi\left(x_i^T\beta\right)(\beta_3 + 2\beta_4 5)$$

A minimally sufficient condition for both the marginal index and marginal probability effects of $\exp_i$ to equal zero for married women with five years of actual work experience (for whom $\exp_i = 5$) is $\beta_3 + 2\beta_4 5 = 0$.

♦ *Null and Alternative Hypotheses:*

H_0: $\beta_3 + 2\beta_4 5 = \beta_3 + 10\beta_4 = 0$
H_1: $\beta_3 + 2\beta_4 5 = \beta_3 + 10\beta_4 \neq 0$

• To calculate a **Wald t-test** of this hypothesis and the corresponding two-tail p-value for the calculated t-statistic, enter the following **lincom** command:

```
lincom _b[exp] + 2*_b[expsq]*5
```

♦ *Test 5 - Model 1:*

♦ *Proposition:* For the median married women in the sample, years of actual work experience have no effect on the probability of labour force participation. In other words, the marginal probability and index effects of $\exp_i$ equal zero for married women for whom $\exp_i = \exp50p$, where $\exp50p$ is the sample median value of $\exp_i$,

marginal index effect of $\exp_i$ for $\exp_i = \exp50p$ is $\beta_3 + 2\beta_4 \exp_i = \beta_3 + 2\beta_4 \exp50p$

marginal probability effect of $\exp_i$ for $\exp_i = \exp50p$ is $\phi\left(x_i^T\beta\right)(\beta_3 + 2\beta_4 \exp_i)$
$$= \phi\left(x_i^T\beta\right)(\beta_3 + 2\beta_4 \exp50p)$$

A minimally sufficient condition for both the marginal index and marginal probability effects of $\exp_i$ to equal zero for married women with median years of actual work experience (for whom $\exp_i = \exp50p$) is $\beta_3 + 2\beta_4 \exp50p = 0$, meaning that the marginal *index* effect of $\exp_i$ equals zero when evaluated at $\exp50p$, the sample median value of $\exp_i$.

• First, use the **summarize** command with the **detail** option to compute the sample median value of $\exp_i$. Enter the commands:

```
summarize exp, detail
return list
```

Note that the sample median value of exp_i is saved in the scalar **r(p50)** and equals 9 years.

- Use a **scalar** command to save the sample median value of exp_i. Enter the **scalar** command:

```
scalar exp50p = r(p50)
```

♦ *Null and Alternative Hypotheses:*

$$H_0: \beta_3 + 2\beta_4 \, exp50p \; = \; \beta_3 + 2\beta_4 9 \; = \; 0$$
$$H_1: \beta_3 + 2\beta_4 \, exp50p \; = \; \beta_3 + 2\beta_4 9 \; \neq \; 0$$

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

```
lincom _b[exp] + 2*_b[expsq]*exp50p
lincom _b[exp] + 2*_b[expsq]*9
```

Note that these two **lincom** commands produce identical results.

♦ <u>*Test 6 - Model 1*</u>*:*

♦ *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 probability (and index) effects of both *kidslt6ᵢ* and *kidsge6ᵢ* equal zero.

♦ *Marginal Index and Probability Effects:*

$$\text{marginal index effect of kidslt6}_i \; = \frac{\partial \, x_i^T \beta}{\partial \, kidslt6_i} \; = \; \beta_6$$

$$\text{marginal probability effect of kidslt6}_i \; = \frac{\partial \, \Phi\!\left(x_i^T \beta\right)}{\partial \, kidslt6_i} \; = \; \phi\!\left(x_i^T \beta\right)\beta_6$$

marginal index effect of kidsge6$_i$ = $\dfrac{\partial x_i^T \beta}{\partial \text{kidsge6}_i}$ = β_7

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

A sufficient condition for the marginal index and probability effects of both kidslt6$_i$ and kidsge6$_i$ to equal zero for all married women is $\beta_6 = 0$ and $\beta_7 = 0$.

♦ *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
```

❑ **Computing marginal *probability* effects of *continuous* explanatory variables -- *dprobit***

Computing Marginal Index Effects: Background

♦ Recall that the **marginal *probability* effect** of a *continuous* **explanatory variable X$_j$** is simply the **marginal *index* effect of X$_j$** multiplied by $\phi(x_i^T \beta)$, the value of the standard normal p.d.f. evaluated at $x_i^T \beta$:

$$\textbf{marginal } \textit{probability } \textbf{effect of X}_j = \frac{\partial \Phi(x_i^T \beta)}{\partial X_{ij}} = \frac{\partial \Phi(x_i^T \beta)}{\partial x_i^T \beta} \frac{\partial x_i^T \beta}{\partial X_{ij}} = \phi(x_i^T \beta)\frac{\partial x_i^T \beta}{\partial X_{ij}}$$

$$= \phi(x_i^T \beta) \times \textbf{marginal } \textit{index } \textbf{effect of X}_j$$

where

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

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

♦ In order to compute estimates of the marginal probability effects of continuous explanatory variables, it is necessary to compute the value of $\phi\!\left(x_i^T \hat{\beta}\right)$, where $\hat{\beta}$ is the vector of ML probit coefficient estimates and x_i^T is a vector of regressor values.

♦ It is also necessary to choose a specific vector of regressor values x_i^T. Common choices for x_i^T are the **sample *mean*** and **sample *median*** values of the regressors, although others may certainly be appropriate in particular circumstances.

Introduction to 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).

♦ The **dprobit** command estimates the ***marginal probability effects*** of the individual regressors **at some user-selected vector of regressor values** x_i^T.

♦ This means that when using the **dprobit** command, you must select a vector of regressor values x_i^T at which to evaluate the marginal probability effects of the explanatory variables. This is done using the **at(*vecname*)** option on the **dprobit** command, where ***vecname*** 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.

♦ The default version of the **dprobit** command, which does not use the **at(*vecname*)** option, estimates the marginal probability effects at the sample mean values of the regressors; i.e., the **dprobit** command *without* the **at(*vecname*)** option sets $x_i^T = \overline{x}^T$,

where the vector $\bar{x}^T$ contains the sample means of all the regressors.

The *dprobit* command *without* the at(*vecname*) option

♦ 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
ereturn list
```

Carefully inspect the following results of these two commands (which are reproduced on the following page).

```
. dprobit inlf nwifeinc ed exp expsq age kidslt6 kidsge6

Iteration 0:   log likelihood =  -514.8732
(output omitted)
Iteration 4:   log likelihood = -401.30219

Probit regression, reporting marginal effects          Number of obs =    753
                                                        LR chi2(7)    = 227.14
                                                        Prob > chi2   = 0.0000
Log likelihood = -401.30219                             Pseudo R2     = 0.2206

------------------------------------------------------------------------------
    inlf |     dF/dx   Std. Err.      z    P>|z|     x-bar  [   95% C.I.   ]
---------+--------------------------------------------------------------------
nwifeinc | -.0046962   .0018903    -2.48   0.013    20.129  -.008401 -.000991
      ed |  .0511287   .0098592     5.18   0.000   12.2869   .031805  .070452
     exp |  .0481771   .0073278     6.59   0.000   10.6308   .033815  .062539
   expsq | -.0007371   .0002347    -3.15   0.002   178.039  -.001197 -.000277
     age | -.0206432   .0033079    -6.23   0.000   42.5378  -.027127  -.01416
 kidslt6 | -.3391514   .0463581    -7.33   0.000   .237716  -.430012 -.248291
 kidsge6 |  .0140628   .0169852     0.83   0.408   1.35325  -.019228  .047353
---------+--------------------------------------------------------------------
  obs. P |  .5683931
 pred. P |   .581542  (at x-bar)
------------------------------------------------------------------------------
    z and P>|z| correspond to the test of the underlying coefficient being 0

. ereturn list

scalars:
               e(N) =  753
            e(ll_0) = -514.8732045671461
              e(ll) = -401.3021932657578
            e(df_m) =  7
            e(chi2) =  227.1420226027767
            e(r2_p) =  .2205805435085081
            e(pbar) =  .5683930942895087

           e(xbar) =  .2058398916263727
          e(offbar) =  0
```

$$= \bar{x}^{T}\hat{\beta}$$

```
macros:
             e(cmd) : "dprobit"
           e(dummy) : " 0 0 0 0 0 0 0 0"
           e(title) : "Probit regression"
          e(depvar) : "inlf"
        e(crittype) : "log likelihood"
         e(predict) : "probit_p"
      e(properties) : "b V"
       e(estat_cmd) : "probit_estat"
        e(chi2type) : "LR"

matrices:
               e(b) :  1 x 8
               e(V) :  8 x 8
         e(se_dfdx) :  1 x 7
            e(dfdx) :  1 x 7

functions:
           e(sample)
```

- One of the saved scalars from the above **dprobit** command is $\overline{x}^T\hat{\beta}$, the estimated value of the probit index function $x_i^T\beta$ evaluated at the sample means of the regressors $\overline{x}^T$ and the ML probit coefficient estimates $\hat{\beta}$. To display the value of $\overline{x}^T\hat{\beta}$, enter the command:

      display e(xbar)

- The estimated value of $\Phi\!\left(\overline{x}^T\hat{\beta}\right)$, the standard normal c.d.f. $\Phi\!\left(x_i^T\beta\right)$ evaluated at $\overline{x}^T\hat{\beta}$, can be computed using the *Stata* statistical function **normal(z)**. Enter the command:

      display normal(e(xbar))

   Note that the value returned by this **display** command equals the value of "**pred. P**" in the output of the **dprobit** command.

- The value of $\phi\!\left(\overline{x}^T\hat{\beta}\right)$, the estimated value of the standard normal p.d.f. $\phi\!\left(x_i^T\beta\right)$ evaluated at $\overline{x}^T\hat{\beta}$, can be computed using the *Stata* statistical function **normalden(z)**. Enter the command:

      display normalden(e(xbar))

- 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(vecname)* option

♦ An alternative way to obtain ML estimates of marginal probability effects in probit models is to use the **at(*vecname*)** option on the **dprobit** command, where *vecname* 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.

To illustrate how the **at(*vecname*)** option works, we use it to compute estimates of the marginal probability effects of the regressors in Model 1 at the vector of sample means, i.e., at $x_i^T = \overline{x}^T$.

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

```
summarize nwifeinc
scalar nwincbar = r(mean)
summarize ed
scalar edbar = r(mean)
summarize exp
scalar expbar = r(mean)
summarize expsq
scalar expsqbar = r(mean)
summarize age
scalar agebar = r(mean)
summarize kidslt6
scalar klt6bar = r(mean)
summarize kidsge6
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. To see this, enter the above **dprobit** command without the **at(*xbar*)** option:

  dprobit inlf nwifeinc ed exp expsq age kidslt6 kidsge6

  Carefully compare the results of these two **dprobit** commands. They are identical.

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:

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

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

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

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

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

- The **ML estimate of the marginal probability effect of exp$_i$** at $\overline{x}^T$ is:

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

The following steps demonstrate how to compute $\phi(\overline{x}^T\hat{\beta})(\hat{\beta}_3 + 2\hat{\beta}_4\overline{\exp})$ for Model 1, and how to test the null hypothesis $\phi(\overline{x}^T\hat{\beta})(\hat{\beta}_3 + 2\hat{\beta}_4\overline{\exp}) = 0$.

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

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

- Note from the above **ereturn list** command that the temporarily-saved scalar **e(xbar)** contains the value of $\overline{x}^T\hat{\beta}$, the estimated value of the probit index function $x_i^T\beta$ evaluated at the sample means of the regressors $\overline{x}^T$ and the ML probit coefficient estimates $\hat{\beta}$. To display the value of $\overline{x}^T\hat{\beta}$, enter the **display** command:

    ```
    display e(xbar)
    ```

- Now use the **normalden(z)** statistical function to compute the value of $\phi(\overline{x}^T\hat{\beta})$ and save it as a scalar. Enter the following **display** and **scalar** commands:

    ```
    display normalden(e(xbar))
    scalar phixbarb = normalden(e(xbar))
    scalar list phixbarb
    ```

    The scalar **phixbarb** now equals the value of $\phi(\overline{x}^T\hat{\beta})$ for Model 1.

- Finally, use the **lincom** command both to compute the value of $\phi(\overline{x}^T\hat{\beta})(\hat{\beta}_3 + 2\hat{\beta}_4\overline{\exp})$ and to perform a t-test of the null hypothesis that $\phi(\overline{x}^T\hat{\beta})(\hat{\beta}_3 + 2\hat{\beta}_4\overline{\exp}) = 0$. Enter either of the following **lincom** 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(vecname)* option

♦ The **dprobit** command with the **at(*vecname*)** 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 commands:

```
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 commands:

```
dprobit inlf nwifeinc ed exp expsq age kidslt6 kidsge6, at(xvec)
ereturn list
display e(at)
```

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.

Note that the temporarily-saved scalar **e(at)** contains the value of $\Phi(x_i^T\hat{\beta})$, the estimated value of the standard normal *distribution* function (or c.d.f.) evaluated at the ML probit coefficient estimates $\hat{\beta}$ and the specified vector *xvec*, where *xvec* is

$$x_i^T = (20\ 14\ 10\ 100\ 40\ 0\ 1\ 1).$$

- Compute an estimate of the **marginal *probability* effect of *exp*$_i$** for married women with the observed characteristics specified by the regressor vector ***xvec***, and test the null hypothesis that this marginal probability effect equals zero. First, we have to compute the estimated value of the standard normal p.d.f. $\phi(x_i^T\beta)$ evaluated at $\hat{\beta}$ and the specified vector ***xvec***; this is done using a **scalar** command together with the *Stata* statistical functions **invnormal(*p*)** and **normalden(*z*)**. Enter the commands:

```
display invnormal(e(at))
scalar phixvec = normalden(invnormal(e(at)))
scalar list phixvec
lincom phixvec*(_b[exp] + 2*_b[expsq]*10)
```

Since the temporarily-saved scalar **e(at)** contains the value of $\Phi(x_i^T\hat{\beta})$, the *Stata* statistical function **invnormal(e(at))** returns the value of $x_i^T\hat{\beta}$. The saved scalar **phixvec** thus equals the value of $\phi(x_i^T\hat{\beta})$.

- Compute estimates of the **marginal *probability* effects** of nwifeinc$_i$, ed$_i$, age$_i$, kidslt6$_i$, and kidsge6$_i$ for married women with the observed characteristics specified by the regressor vector ***xvec***, and test the null hypothesis that each of these marginal probability effects equals zero. Enter the **lincom** commands:

```
lincom phixvec*_b[nwifeinc]
lincom phixvec*_b[ed]
lincom phixvec*_b[age]
lincom phixvec*_b[kidslt6]
lincom phixvec*_b[kidsge6]
```

Compare the estimated marginal probability effects of nwifeinc$_i$, ed$_i$, age$_i$, kidslt6$_i$, and kidsge6$_i$ produced by these **lincom** commands with those produced by the previous **dprobit** command. They should be identical because each of the explanatory variables nwifeinc$_i$, ed$_i$, age$_i$, kidslt6$_i$, and kidsge6$_i$ enter linearly the probit index function for Model 1.

❑ **Marginal *probability* effects of *binary* explanatory variables – Model 2**

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* 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 ***kidslt6$_i$* = 0** for the i-th married woman

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

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

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. $\Phi(x_i^T\beta)$ with respect to the individual explanatory variables:

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

where

$\phi(x_i^T\beta) = $ 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 $\text{nwifeinc}_i = \dfrac{\partial \Phi(x_i^T\beta)}{\partial \text{nwifeinc}_i} = \phi(x_i^T\beta)\beta_1$

2. marginal probability effect of $\text{ed}_i = \dfrac{\partial \Phi(x_i^T\beta)}{\partial \text{ed}_i} = \phi(x_i^T\beta)\beta_2$

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

4. marginal probability effect of $\text{age}_i = \dfrac{\partial \Phi(x_i^T\beta)}{\partial \text{age}_i} = \phi(x_i^T\beta)\beta_5$

Marginal probability effects of binary (or dummy) explanatory variables

♦ The **marginal *probability* effect of a *binary*, or *dummy*, explanatory variable D_j in probit models** is computed by evaluating the standard normal c.d.f. $\Phi(x_i^T\beta)$ 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 $= \Phi(x_{1i}^T\beta) - \Phi(x_{0i}^T\beta)$.

♦ *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 dkidslt6$_i$ = 1:

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

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

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

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

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

The **conditional probability that** *inlf*$_i$ **= 1 when** *dkidslt6*$_i$ **= 0** is:

$$\Pr\left(\text{inlf}_i = 1 \middle| \text{dkidslt6}_i = 0\right) = \Phi\left(x_{0i}^T \beta\right)$$
$$= \Phi\left(\beta_0 + \beta_1 \text{nwifeinc}_i + \beta_2 \text{ed}_i + \beta_3 \exp_i + \beta_4 \exp_i^2 + \beta_5 \text{age}_i\right)$$

The **marginal** *probability* **effect of the dummy variable** *dkidslt6*$_i$ **in Model 2** is the difference between the **conditional probability that** *inlf*$_i$ **= 1 when** *dkidslt6*$_i$ **= 1** and the **conditional probability that** *inlf*$_i$ **= 1 when** *dkidslt6*$_i$ **= 0**:

$$\Pr\left(\text{inlf}_i = 1 \middle| \text{dkidslt6}_i = 1\right) - \Pr\left(\text{inlf}_i = 1 \middle| \text{dkidslt6}_i = 0\right)$$

$$= \Phi\left(x_{1i}^T \beta\right) - \Phi\left(x_{0i}^T \beta\right)$$

$$= \Phi\left(\beta_0 + \beta_1 \text{nwifeinc}_i + \beta_2 \text{ed}_i + \beta_3 \exp_i + \beta_4 \exp_i^2 + \beta_5 \text{age}_i + \beta_6\right)$$
$$- \Phi\left(\beta_0 + \beta_1 \text{nwifeinc}_i + \beta_2 \text{ed}_i + \beta_3 \exp_i + \beta_4 \exp_i^2 + \beta_5 \text{age}_i\right)$$

❑ **Marginal *probability* effects of *binary* explanatory variable *dkidslt6* -- *dprobit***

- Before estimating Model 2, it is necessary to create the new dummy variable *dkidslt6$_i$*. Enter the commands:

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

Marginal Probability Effects for Model 2 -- *dprobit*

- Use the **dprobit** command *without* the **at(*vecname*)** option to compute the marginal probability effects in Model 2 at the **sample *mean* values of the regressors**, i.e., at $x_i^T = \bar{x}^T$. Enter the following command:

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

♦ The next series of *Stata* commands will demonstrate how to use the **dprobit** command with the **at(*vecname*)** option to compute the **marginal *probability* effect of the dummy variable *dkidslt6$_i$* 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), and who are 40 years of age (age = 40):

$$\Pr\left(inlf_i = 1 \mid dkidslt6_i = 1\right) - \Pr\left(inlf_i = 1 \mid dkidslt6_i = 0\right)$$

$$= \Phi\left(x_{1i}^T\beta\right) - \Phi\left(x_{0i}^T\beta\right)$$

$$= \Phi\left(\beta_0 + \beta_1 nwifeinc_i + \beta_2 ed_i + \beta_3 \exp_i + \beta_4 \exp_i^2 + \beta_5 age_i + \beta_6\right)$$
$$\quad - \Phi\left(\beta_0 + \beta_1 nwifeinc_i + \beta_2 ed_2 + \beta_3 \exp_i + \beta_4 \exp_i^2 + \beta_5 age_i\right)$$

$$= \Phi\left(\beta_0 + \beta_1 20 + \beta_2 14 + \beta_3 10 + \beta_4 100 + \beta_5 40 + \beta_6\right)$$
$$\quad - \Phi\left(\beta_0 + \beta_1 20 + \beta_2 14 + \beta_3 10 + \beta_4 100 + \beta_5 40\right)$$

The procedure for this computation consists of three steps.

Step 1: Compute an estimate of the probability of labour force participation for married women with the specified characteristics **who have *one or more* dependent children under 6 years of age, for whom dkidslt6$_i$ = 1**: i.e., compute an estimate of

$$\Phi\left(x_{1i}^T\beta\right) = \Phi\left(\beta_0 + \beta_1 nwifeinc_i + \beta_2 ed_i + \beta_3 \exp_i + \beta_4 \exp_i^2 + \beta_5 age_i + \beta_6\right)$$
$$= \Phi\left(\beta_0 + \beta_1 20 + \beta_2 14 + \beta_3 10 + \beta_4 100 + \beta_5 40 + \beta_6\right).$$

Step 2: Compute an estimate of the probability of labour force participation for married women with the specified characteristics **who have *no* dependent children under 6 years of age, for whom dkidslt6$_i$ = 0**: i.e., compute an estimate of

$$\Phi\left(x_{0i}^T\beta\right) = \Phi\left(\beta_0 + \beta_1 nwifeinc_i + \beta_2 ed_i + \beta_3 \exp_i + \beta_4 \exp_i^2 + \beta_5 age_i\right)$$
$$= \Phi\left(\beta_0 + \beta_1 20 + \beta_2 14 + \beta_3 10 + \beta_4 100 + \beta_5 40\right).$$

Step 3: Compute an estimate of the difference $\Phi\left(x_{1i}^T\beta\right) - \Phi\left(x_{0i}^T\beta\right)$, which is the marginal probability effect of having one or more dependent children under 6 years of age for married women who have the specified characteristics.

- **Step 1:** Use the **dprobit** command *with* the **at(*vecname*)** 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 *one or more* dependent children under 6 years of age (dkidslt6 = 1)**. You will first have to create a vector containing the specified values of the regressors for Model 2, since the **dprobit** command does not permit number lists in the **at( )** option. Note that in *Stata* format, the vector x_{1i}^T with the dummy variable *dkidslt6$_i$* = **1** is written as:

$$x_{1i}^T = \left(nwifeinc_i \ \ ed_i \ \ \exp_i \ \ \exp_i^2 \ \ age_i \ \ 1 \ \ 1\right) = \left(20 \ \ 14 \ \ 10 \ \ 100 \ \ 40 \ \ 1 \ \ 1\right).$$

Enter the following commands:

```
matrix x1vec = (20, 14, 10, 100, 40, 1, 1)
matrix list x1vec
dprobit inlf nwifeinc ed exp expsq age dkidslt6, at(x1vec)
ereturn list
```

Display and save the value of $\Phi\left(x_{1i}^T\hat\beta\right)$ generated by the above **dprobit** command, where $\Phi\left(x_{1i}^T\hat\beta\right)$ is an estimate of $\Pr\left(\mathrm{inlf}_i = 1 \mid \mathrm{dkidslt6}_i = 1\right)$. Enter the commands:

```
display e(at)
scalar PHIx1vec = e(at)
scalar list PHIx1vec
```

- **Step 2:** Now use the **dprobit** command *with* the **at(*vecname*)** option to compute the marginal probability effects in Model 2 for married women whose non-wife family income is $20,000 per year (nwifeinc = 20), who have 14 years of formal education (ed = 14) and 10 years of actual work experience (exp = 10, expsq = 100), who are 40 years of age (age = 40), and **who have *no* dependent children under 6 years of age (dkidslt6 = 0)**. First, you will have to create a vector containing the specified values of the regressors for Model 2; the **dprobit** command does not permit number lists in the **at( )** option. Note that in *Stata* format, the vector x_{0i}^T with the dummy variable ***dkidslt6$_i$ = 0*** is written as:

$$x_{0i}^T = \left(\mathrm{nwifeinc}_i \ \mathrm{ed}_i \ \exp_i \ \exp_i^2 \ \mathrm{age}_i \ 0 \ 1\right) = \left(20 \ 14 \ 10 \ 100 \ 40 \ 0 \ 1\right)$$

Enter the following commands:

```
matrix x0vec = (20, 14, 10, 100, 40, 0, 1)
matrix list x0vec
dprobit inlf nwifeinc ed exp expsq age dkidslt6, at(x0vec)
ereturn list
```

Display and save the value of $\Phi\left(x_{0i}^T\hat\beta\right)$ generated by the above **dprobit** command, where $\Phi\left(x_{0i}^T\hat\beta\right)$ is an estimate of $\Pr\left(\mathrm{inlf}_i = 1 \mid \mathrm{dkidslt6}_i = 0\right)$. Enter the commands:

```
display e(at)
scalar PHIx0vec = e(at)
scalar list PHIx0vec
```

- **Step 3:** Finally, compute the estimate of the difference $\Phi\left(x_{1i}^{T}\beta\right) - \Phi\left(x_{0i}^{T}\beta\right)$, which is the marginal probability effect having one or more dependent children under 6 years of age for married women who have the specified characteristics. Enter the commands:

  ```
  scalar diffPHI = PHIx1vec - PHIx0vec
  scalar list PHIx1vec PHIx0vec diffPHI
  ```

- Carefully compare the results of this three-step procedure with the output of the two **dprobit** commands you have estimated. Enter the following commands:

  ```
  * Model 2 at x0vec: dprobit
  dprobit inlf nwifeinc ed exp expsq age dkidslt6, at (x0vec)
  * Model 2 at x1vec: dprobit
  dprobit inlf nwifeinc ed exp expsq age dkidslt6, at (x1vec)
  ```

  The *Stata* output listing produced by these commands is reproduced on the following page. Note in particular the highlighted results in the output listing for these two **dprobit** commands.

```
. * Model 2 at x0vec: dprobit
. dprobit inlf nwifeinc ed exp expsq age dkidslt6, at (x0vec)

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -410.52123
Iteration 2:   log likelihood = -407.00272
Iteration 3:   log likelihood = -406.98832

Probit regression, reporting marginal effects          Number of obs =     753
                                                        LR chi2(6)    = 215.77
                                                        Prob > chi2   = 0.0000
Log likelihood = -406.98832                             Pseudo R2     = 0.2095


------------------------------------------------------------------------------
    inlf |      dF/dx   Std. Err.      z    P>|z|      x [    95% C.I.    ]
---------+--------------------------------------------------------------------
nwifeinc |  -.0032397   .0013341    -2.39   0.017     20  -.005854 -.000625
      ed |   .0347428   .0061286     4.92   0.000     14   .022731  .046755
     exp |   .0334919   .0050403     6.32   0.000     10   .023613  .043371
   expsq |  -.0005032   .0001622    -2.94   0.003    100  -.000821 -.000185
     age |  -.0152501   .0021914    -6.73   0.000     40  -.019545 -.010955
dkidslt6*|  -.3741459   .0527655    -7.04   0.000      0  -.477564 -.270728
---------+--------------------------------------------------------------------
  obs. P |   .5683931
 pred. P |    .583103   (at x-bar)
 pred. P |   .7935022  (at x)
------------------------------------------------------------------------------
(*) dF/dx is for discrete change of dummy variable from 0 to 1
    z and P>|z| correspond to the test of the underlying coefficient being 0

. * Model 2 at x1vec: dprobit
. dprobit inlf nwifeinc ed exp expsq age dkidslt6, at (x1vec)

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -410.52123
Iteration 2:   log likelihood = -407.00272
Iteration 3:   log likelihood = -406.98832

Probit regression, reporting marginal effects          Number of obs =     753
                                                        LR chi2(6)    = 215.77
                                                        Prob > chi2   = 0.0000
Log likelihood = -406.98832                             Pseudo R2     = 0.2095


------------------------------------------------------------------------------
    inlf |      dF/dx   Std. Err.      z    P>|z|      x [    95% C.I.    ]
---------+--------------------------------------------------------------------
nwifeinc |  -.0044364   .0018708    -2.39   0.017     20  -.008103  -.00077
      ed |   .0475765   .0100112     4.92   0.000     14   .027955  .067198
     exp |   .0458635   .0076519     6.32   0.000     10   .030866  .060861
   expsq |  -.0006891   .0002397    -2.94   0.003    100  -.001159 -.000219
     age |  -.0208833   .0029759    -6.73   0.000     40  -.026716 -.015051
dkidslt6*|  -.3741459   .0527655    -7.04   0.000      1  -.477564 -.270728
---------+--------------------------------------------------------------------
  obs. P |   .5683931
 pred. P |    .583103   (at x-bar)
 pred. P |   .4193563  (at x)
------------------------------------------------------------------------------
(*) dF/dx is for discrete change of dummy variable from 0 to 1
    z and P>|z| correspond to the test of the underlying coefficient being 0
```

❑ **Marginal *probability* effects of *continuous* explanatory variables in Model 2 --
   *dprobit***

For any given set of values of the continuous explanatory variables $nwifeinc_i$, ed_i, exp_i
and age_i in Model 2, there are two distinct sets of marginal probability effects to
compute: (1) those for married women with the specified values of $nwifeinc_i$, ed_i, exp_i
and age_i **who have *no* dependent children under 6 years of age (for whom *dkidslt6_i* =
0)**; and (2) those for married women with the specified values of $nwifeinc_i$, ed_i, exp_i and
age_i **who have *one or more* dependent children under 6 years of age (for whom
dkidslt6_i = 1)**.

Marginal Probability Effects for Women for Whom *dkidslt6* = 0

♦ In this section, we use the **dprobit** command with the **at(*vecname*)** option to
   compute the **marginal *probability* effects of the *continuous* explanatory variables
   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 (for whom *dkidslt6_i* =
   0)**. These are given by the following expressions:

1. marginal probability effect of $nwifeinc_i$ $\quad = \quad \phi\left(x_{0i}^{T}\beta\right)\beta_{1}$

2. marginal probability effect of ed_i $\qquad\qquad = \quad \phi\left(x_{0i}^{T}\beta\right)\beta_{2}$

3. marginal probability effect of exp_i $\qquad\quad = \quad \phi\left(x_{0i}^{T}\beta\right)\left(\beta_{3} + 2\beta_{4}\,exp_{i}\right)$
   $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad = \quad \phi\left(x_{0i}^{T}\beta\right)\left(\beta_{3} + 2\beta_{4}10\right)$

4. marginal probability effect of age_i $\qquad\quad = \quad \phi\left(x_{0i}^{T}\beta\right)\beta_{5}$

Note that in all these four expressions, the standard normal p.d.f. $\phi\left(x_{i}^{T}\beta\right)$ is evaluated
at the vector x_{0i}^{T} with the dummy variable *dkidslt6_i* = 0, where x_{0i}^{T} in *Stata* format is

$$x_{0i}^{T} = \left(nwifeinc_{i}\ ed_{i}\ exp_{i}\ exp_{i}^{2}\ age_{i}\ 0\ 1\right) = \left(20\ 14\ 10\ 100\ 40\ 0\ 1\right).$$

You have previously created the vector x_{0i}^{T} and named it *x0vec*.

- First, use the **dprobit** command with the **at(*vecname*)** option to re-estimate Model 2 and evaluate the marginal probability effects at x_{0i}^T. Enter the commands:

```
dprobit inlf nwifeinc ed exp expsq age dkidslt6, at(x0vec)
ereturn list
display e(at)
```

Note that the temporarily saved scalar **e(at)** created by this **dprobit** command equals the value of $\Phi\left(x_{0i}^T\hat{\beta}\right)$, which is an estimate of $\Pr\left(inlf_i = 1 | dkidslt6_i = 0\right)$.

- Second, use the *Stata* **invnormal(*p*)** statistical function to save the value of $x_{0i}^T\hat{\beta}$ as the scalar *x0vecb*. Enter the **scalar** commands:

```
scalar x0vecb = invnormal(e(at))
scalar list x0vecb
```

- Third, use the *Stata* **normalden(*z*)** statistical function to save the value of $\phi\left(x_{0i}^T\hat{\beta}\right)$ as the scalar *phix0vec*. Enter the **scalar** commands:

```
scalar phix0vec = normalden(x0vecb)
scalar list x0vecb phix0vec
```

- Finally, use **lincom** commands to compute the values of the **marginal *probability* effects** of the four continuous explanatory variables in Model 2 for women with the specified characteristics who have no pre-school aged children, and to test the null hypothesis that each marginal probability effect equals zero. Enter the **lincom** commands:

```
lincom phix0vec*_b[nwifeinc]
lincom phix0vec*_b[ed]
lincom phix0vec*(_b[exp] + 2*_b[expsq]*10)
lincom phix0vec*_b[age]
```

Compare the estimated marginal probability effects of the explanatory variables nwifeinc$_i$, ed$_i$ and age$_i$ produced by these **lincom** commands with those computed by the preceding **dprobit** command with the **at(*x0vec*)** option; they are identical because the variables nwifeinc$_i$, ed$_i$ and age$_i$ enter the probit index function for Model 2 linearly.

Marginal Probability Effects for Women for Whom *dkidslt6* = 1

♦ In this section, we use the **dprobit** command with the **at(*vecname*)** option to compute the **marginal *probability* effects of the *continuous* explanatory variables in Model 2** for married women whose non-wife family income is $20,000 per year (nwifeinc = 20), who have 14 years of formal education (ed = 14) and 10 years of actual work experience (exp = 10, expsq = 100), who are 40 years of age (age = 40), and **who have *one or more* dependent children under 6 years of age (for whom *dkidslt6$_i$* = 1)**. These are given by the following expressions:

1. marginal probability effect of nwifeinc$_i$ $\quad = \quad \phi\left(x_{1i}^T\beta\right)\beta_1$

2. marginal probability effect of ed$_i$ $\qquad = \quad \phi\left(x_{1i}^T\beta\right)\beta_2$

3. marginal probability effect of exp$_i$ $\qquad = \quad \phi\left(x_{1i}^T\beta\right)(\beta_3 + 2\beta_4 \exp_i)$
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad = \quad \phi\left(x_{1i}^T\beta\right)(\beta_3 + 2\beta_4 10)$

4. marginal probability effect of age$_i$ $\qquad = \quad \phi\left(x_{1i}^T\beta\right)\beta_5$

Note that in all four of these expressions, the standard normal p.d.f. $\phi\left(x_i^T\beta\right)$ is evaluated at the vector x_{1i}^T with the dummy variable *dkidslt6$_i$* = 1, where x_{1i}^T in *Stata* format is

$$x_{1i}^T = \left(\text{nwifeinc}_i \ \ \text{ed}_i \ \ \exp_i \ \ \exp_i^2 \ \ \text{age}_i \ \ 1 \ 1\right) = \left(20 \ \ 14 \ \ 10 \ \ 100 \ \ 40 \ \ 1 \ \ 1\right).$$

You have previously created the vector x_{1i}^T and named it ***x1vec***.

• First, use the **dprobit** command with the **at(*vecname*)** option to re-estimate Model 2 and evaluate the marginal probability effects at x_{1i}^T. Enter the commands:

```
dprobit inlf nwifeinc ed exp expsq age dkidslt6, at(x1vec)
ereturn list
display e(at)
```

Note that the temporarily saved scalar **e(at)** created by this **dprobit** command equals the value of $\Phi\left(x_{1i}^T\hat{\beta}\right)$, which is an estimate of $\Pr\left(\text{inlf}_i = 1 \mid \text{dkidslt6}_i = 1\right)$.

- Second, use the *Stata* **invnormal(*p*)** statistical function to save the value of $x_{1i}^T\hat{\beta}$ as the scalar ***x1vecb***. Enter the **scalar** commands:

  ```
  scalar x1vecb = invnormal(e(at))
  scalar list x1vecb
  ```

- Third, use the *Stata* **normalden(*z*)** statistical function to save the value of $\phi\left(x_{1i}^T\hat{\beta}\right)$ as the scalar ***phix1vec***. Enter the **scalar** commands:

  ```
  scalar phix1vec = normalden(x1vecb)
  scalar list x1vecb phix1vec
  ```

- Finally, use **lincom** commands to compute the values of the **marginal *probability* effects** of the four continuous explanatory variables in Model 2 for women with the specified characteristics who have one or more pre-school aged children, and to test the null hypothesis that each marginal probability effect equals zero. Enter the **lincom** commands:

  ```
  lincom phix1vec*_b[nwifeinc]
  lincom phix1vec*_b[ed]
  lincom phix1vec*(_b[exp] + 2*_b[expsq]*10)
  lincom phix1vec*_b[age]
  ```

  Compare the estimated marginal probability effects of the explanatory variables $nwifeinc_i$, ed_i and age_i produced by these **lincom** commands with those computed by the preceding **dprobit** command with the **at(*x1vec*)** option; they are identical because the variables $nwifeinc_i$, ed_i and age_i enter the probit index function for Model 2 linearly.

❑ **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 command log file** you have been recording. Enter the command:

```
cmdlog close
```

• Third, **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:\data** on the computers in Dunning 350, and **D:\courses** on the computers in MC B111. There are two files you will want to be sure you have: the *Stata* log file **452tutorial6.log**; and the *Stata* command log file **452tutorial6.txt**. You will also want to take with you the *Stata*-format data set **mroz.dta**, as you will need it for *Stata 10/11 Tutorial 7*. Use the Windows **copy** command to copy any files you want to keep to your own portable electronic storage device (e.g., a flash memory stick).

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.