Warwick Economics Summer School
Topics in Microeconometrics
Binary Response Models

Michele Aquaro

University of Warwick

This version: July 24, 2016

Reading material

- Textbook: *Introductory Econometrics: A Modern Approach*, 5th edition, J.M. Wooldridge, Chapter 17, Section 17.1.
- These slides. (When the symbol  appears on the top-right of the page, it means that the material presented is slightly more advanced.)
- Topics:
  1. Introduction
  2. Review of the Linear Probability Model
  3. Logit and Probit
  4. Reporting the Results

Section 1

Introduction

- Often we want to explain a qualitative outcome. For example, suppose we want to study the question of married women's labor force participation (in the labor force or out). Or, we want to know whether a young man is arrested for a crime during a certain period of time. Or, we want to know if an employee participates in a 401(k) pension plan.
- In these cases, the variable y we want to explain is a binary (zero-one) variable.
- Sometimes we say we have a "dummy dependent variable."

• The most we can know about y in terms of explanatory variables x_1, x_2, ..., x_k is

$$P(y = 1|x_1, x_2, ..., x_k) = p(x_1, x_2, ..., x_k) = p(\boldsymbol{x}),$$

which is called the **response probability**. ($\boldsymbol{x}$ is shorthand for all of the explanatory variables). It is the probability of a "success" (used loosely), that is, $y = 1$.
• Of course $P(y = 0|\boldsymbol{x}) = 1 - P(y = 1|\boldsymbol{x}) = 1 - p(\boldsymbol{x})$.

• As in regression, we are interested in the partial effects of the x_j on $p(\boldsymbol{x})$.
• For discrete x_j, look at changes in the response probability (usually holding other variables fixed). For example, if $x_K = train$ (job training indicator) and y is an employment indicator,

$$P(y = 1|x_1, ..., x_{K-1}, x_K = 1) - P(y = 1|x_1, ..., x_{K-1}, x_K = 0)$$

that is

$$p(x_1, ..., x_{K-1}, 1) - p(x_1, ..., x_{K-1}, 0)$$

is the effect of job training on the employment probability, at given values for the other covariates.
• For continuous x_j, these are usually the partial derivatives.

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j}.$$

• In nonlinear models generally, and binary response models specifically, it is often useful to have a single number to summarize the relationship between $P(y = 1|\boldsymbol{x})$ and x_j. In a linear model that is simply the coefficient.
• Generally, we might report an estimated **average partial effect (APE)** (also called the *average marginal effect*). The APE for a continuous x_j is

$$E_{\boldsymbol{x}}\left[\frac{\partial p(\boldsymbol{x})}{\partial x_j}\right],$$

which means we average the partial effect across the population distribution of $\boldsymbol{x}$. This is a weighted average of the partial effects at each outcome $\boldsymbol{x}$.

- Suppose x_K is a binary variable. Then its APE is

$$E_{\boldsymbol{x}_{(K)}}[p(\boldsymbol{x}_{(K)}, 1) - p(\boldsymbol{x}_{(K)}, 0)]$$

where $\boldsymbol{x}_{(K)}$ is the $1 \times (K-1)$ vector with x_K excluded.

• Some simple, useful facts about binary (zero-one) random variables are

$$E(y|\boldsymbol{x}) = P(y=1|\boldsymbol{x}) = p(\boldsymbol{x})$$
$$Var(y|\boldsymbol{x}) = p(\boldsymbol{x})[1-p(\boldsymbol{x})]$$

• So a binary variable has natural heteroskedasticity except in the special (and not very interesting) case where $p(\boldsymbol{x})$ does not depend on $\boldsymbol{x}$.

Section 2

Review of the Linear Probability Model

• The linear probability model is just a standard linear model where y happens to be binary. If we write down the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$$

when y is binary, how can we interpret the parameters
• y can only change from 0 to 1 or 1 to 0. Suppose $\beta_1 = .035$ and $x_1 = educ$. What does it mean for a one year increased in $educ$ to increase y by .035?

• We must rely on the expected value formulation of linear regression. Remember we can interpret the regression model as

$$E(y|\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

• We can interpret β_j as

$$\Delta E(y|\boldsymbol{x}) = \beta_j \Delta x_j \text{ holding other explanatory variables fixed}$$

Therefore,

$$\beta_j = \Delta E(y|\boldsymbol{x})$$

when $\Delta x_j = 1$ holding other explanatory variables fixed.

• Key relationship when y is binary:

$$E(y|\boldsymbol{x}) = P(y = 1|\boldsymbol{x})$$

• Thus, when we apply the linear model to binary y what we are really saying is

$$P(y = 1|\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k,$$

so that the response probability is linear (in the parameters). Therefore, we call a linear model for binary y a **linear probability model (LPM)**.

• The important point is that all partial effects are effects on the probability that $y = 1$ (sometimes called a "success"). Since $P(y = 0|\boldsymbol{x}) = 1 - P(y = 1|\boldsymbol{x})$ it is the only probability we need.

$$\Delta P(y = 1|\boldsymbol{x}) = \beta_j \Delta x_j \text{ holding other explanatory variables fixed}$$

and so

$$\beta_j = \Delta P(y = 1|\boldsymbol{x})$$

when $\Delta x_j = 1$ holding other explanatory variables fixed.

• The sample analog holds as well. When we have the OLS regression line

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k,$$

$\hat{y}$ is now the predicted probability $P(y = 1|\boldsymbol{x})$. So, the intercept is the predicted probability when each x_j is set to zero (which, as with other regression applications, may not make sense).

• $\hat{\beta}_j$ measures the change in the estimated probability of a "success" when $\Delta x_j = 1$, other factors held fixed.

• Just like in any regression application, we can have explanatory variables in logs, quadratics, and interactions as well as binary regressors.

EXAMPLE: Married Women's Labor Force Participation (MROZ.DTA)

- The variable $inlf$ is one if a woman worked for a wage during a certain year, and zero if not. We estimate a linear probability model to see the effects of variables on the probability of being in the labor force.
- The estimated equation is (standard errors in parentheses):

$$\begin{aligned} \widehat{inlf} = & \underset{(.154)}{.586} - \underset{(.0014)}{.0034} nwifeinc + \underset{(.007)}{.038} educ + \underset{(.006)}{.039} exper - \underset{(.00018)}{.00060} exper^2 \\ & - \underset{(.002)}{.016} age - \underset{(.034)}{.262} kidslt6 + \underset{(.013)}{.013} kidsge6 \\ n = & \ 753, \ R^2 = .264 \end{aligned}$$

(See slide 35 for the Stata output.)

- These are the usual OLS standard errors, even though they are not quite valid due to heteroskedasticity. So, we should use heteroskedasticity-robust standard errors in general.

• The intercept is not of interest here. The coefficient on $nwifeinc$ (other sources of income) shows a modest effect: if it increases by 20 ($\$20,000$), the associated probability of being in the labor force falls by .068, or 6.8 *percentage points*. The t statistic shows it is statistically significant at the 2% level.
• Each year of education is associated with a probability of being in the labor force .038 higher (or 3.8 percentage points higher).
• Past workforce experience has a positive but diminishing effect. The effect of the first year is about .039, and this diminishes to zero at $exper = .039/(2 \cdot .0006) = 32.5$. (Only 13 women have experience > 32.)

• Having young children has a very large negative effect: being in the labor force falls by .262 for each young child.

• It is unwise to extrapolate to extreme values when using any linear model, including this one.

• Using a linear model for a binary outcome is convenient because estimation is easy and so is interpretation.

• But the LPM does have some *peculiarities.*

1. The fitted values from an OLS regression are never guaranteed to be between zero and one, yet these fitted values are estimated probabilities.

This can be a problem if the aim is to make predictions, but it is rarely a big deal if we use the LPM to estimate partial effects. (In the MROZ example, only 33 of 753 fitted values are not in $[0, 1]$, see slide 35.)

2. The estimated partial effects are constant throughout the range of the explanatory variables, possibly leading to silly estimated effects for large changes. (This is related to predicted probabilities possibly being negative or greater than one.)
This is more of a problem because we know that, say, a variable with a positive effect on $P(y = 1|\boldsymbol{x})$ must eventually have a diminishing effect. But the linear model implies a constant effect (when the variable appears by itself).
• For example, take a woman who has no other source of income, 25 years of prior work experience, no children, who is 48 years old. As a function of *educ* the equation looks like

$$\widehat{inlf} = .417 + .038\ educ$$

At $educ = 12$, the predicted probability is .873, at $educ = 14$ it is .949, and at $educ = 16$, $\widehat{inlf} = 1.025$. For the estimated model to truly represent a probability, the effect of education should be diminishing – that is, the next year of education should increase the probability by less than the previous year so that the estimated probability never goes above one.

• Using logarithms does not bound the effect, and using quadratics often does not help. (In this example, a quadratic in *educ* gives an estimated *increasing* effect, not a diminishing effect.)
• But the LPM does a good job of approximating partial effects if we do not look at extreme values of the explanatory variables.

3. Because y is binary – and this really has nothing to do with the LPM per se – the LPM must exhibit heteroskedasticity except in the one case where no x_j affects $P(y = 1|\boldsymbol{x})$. This follows because for a binary variable,

$$\begin{aligned} Var(y|\boldsymbol{x}) &= p(\boldsymbol{x})[1 - p(\boldsymbol{x})] \\ &= Var(u|\boldsymbol{x}) \end{aligned}$$

where $p(\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$ is the linear response probability.

• This is a case where we *know* the homoskedasticity assumption (MLR.5, see textbook) must fail, and we know how. So, currently, we treat the usual t and F tests with suspicion, and the confidence intervals. (Turns out they are often pretty reliable in LPMs.)

Section 3

Logit and Probit

• A general model that ensures probabilities are between zero and one has the form

$$P(y = 1|\boldsymbol{x}) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k)$$

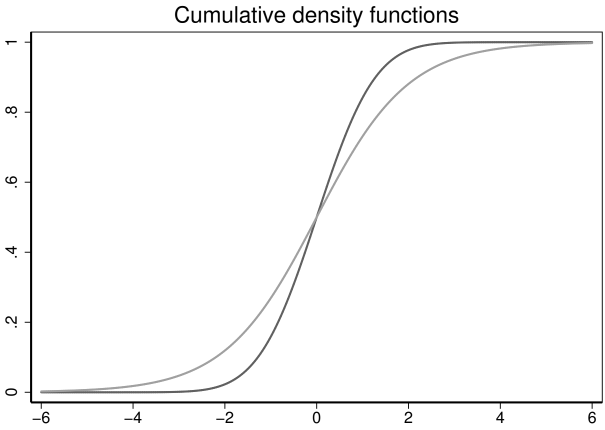
for some function G that takes values between zero and one.
• In most cases, $G(\cdot)$ is actually a cumulative distribution function (cdf) for a continuous random variable with probability density function $g(\cdot)$. Then, $G(\cdot)$ is strictly increasing, and the estimates are easier to interpret.

- The leading cases are

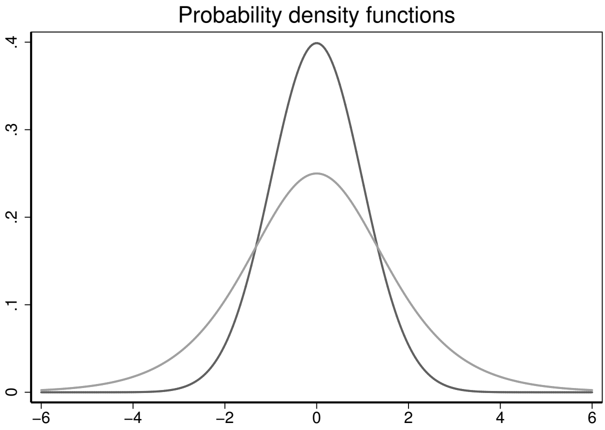
$$G(z) = \Lambda(z) = \frac{\exp(z)}{[1+\exp(z)]} \quad \text{(logit)}$$
$$G(z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) du \quad \text{(probit)}$$

- These functions have similar shapes but the logistic is more spread out.

Logit and Probit $G(z)$



Logit and Probit $g(z)$



• The estimation method for the β_j is maximum likelihood. Computationally straightforward, and runs quickly.

```
probit y x1 x2
logit y x1 x2
```

• Standard errors are reported, as are t statistics and confidence intervals.
• We can test multiple exclusion restrictions using a `test` command, just as with linear regression.

• What do we do with the estimates? Let x_j be continuous. Then the partial effect is

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = \beta_j g(\boldsymbol{x\beta})$$

where

$$\begin{aligned} g(z) &= \phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) && \text{for probit} \\ g(z) &= \exp(z)/[1+\exp(z)]^2 && \text{for logit} \end{aligned}$$

and, because $g(z) > 0$, β_j gives the direction of the partial effect. But its magnitude depends on $g(\boldsymbol{x\beta})$.

• For probit, the largest value of the scale factor is about $.4 = g(0)$. For logit, it is $g(0) = .25$.

• For two continous covariates, the ratio of the coefficients give the ratio of the partial effects, independent of $\boldsymbol{x}$.

$$\frac{\partial p(\boldsymbol{x})/\partial x_j}{\partial p(\boldsymbol{x})/\partial x_h} = \frac{\beta_j g(\boldsymbol{x}\boldsymbol{\beta})}{\beta_h g(\boldsymbol{x}\boldsymbol{\beta})} = \beta_j/\beta_h.$$

• No simple relationship exists for discrete variables or changes.
• In any case, we would like the magnitude of the effect.

- The APE consists in averaging partial effects for actual units:

$$\widehat{APE}_j = \hat{\beta}_j \left[\frac{1}{N} \sum_{i=1}^{N} g(\boldsymbol{x}_i \hat{\boldsymbol{\beta}}) \right]$$

- The scale factor multiplying $\hat{\beta}_j$ is below one, and sometimes well below one.
- It makes no sense to compare magnitudes of coefficients across probit, logit, LPM. Comparing APEs is preferred.
- In particular, if $\hat{\gamma}_j$ is the linear regression coefficient on x_j from estimating an LPM, it can be compared with $\widehat{APE}_j$ (provided no other function of x_j appears in the regressors).

- Suppose x_K is a binary variable. Then its APE is estimated as

$$\widehat{APE}_K = \frac{1}{N}\sum_{i=1}^{N}[G(\boldsymbol{x}_{i(K)}\hat{\boldsymbol{\beta}}_{(K)} + \hat{\beta}_K) - G(\boldsymbol{x}_{i(K)}\hat{\boldsymbol{\beta}}_{(K)})],$$

where $\boldsymbol{x}_{i(K)}$ is $\boldsymbol{x}_i$ but without x_{iK}.

Section 4

Reporting the Results

• Start with an LPM and report the coefficients along with robust standard errors.
• With logit and probit, one typically reports the estimated coefficients along with standard errors. The value of the log likelihood is sometimes reported, too (as it can be used to compute tests of exclusion restrictions).
• It is important to report APEs and their standard errors for logit and probit (maybe even in place of coefficients). Sometimes a key policy variable is of interest, and it is important to estimate the magnitude of the effect.

• If the variable of interest is continuous, or at least takes on lots of values, might compute the partial effects at different values (and average out the other explanatory variables).
• For example, in estimating the effects of young children on married women's labor force participation, we can estimate the partial effect in going from 0 to 1 and 1 to 2 children.
• Goodness-of-fit is not as important, but the percent correctly predicted for each outcome, and overall, is of some interest.

Example (MROZ.DTA)

```
clear all
capture log close
set more off
set linesize 80
qui log using lg_wooldridge2013_chapter17_mroz.txt, text replace

/* Example 17.1 */

use MROZ.DTA
des inlf nwifeinc educ exper expersq age kidslt6 kidsge6
sum inlf nwifeinc educ exper expersq age kidslt6 kidsge6

* LPM
regress inlf nwifeinc educ exper expersq age kidslt6 kidsge6, robust
predict lpm, xb
summ lpm if lpm < 0 | lpm > 1
gen lpm_c = (lpm >= .5)
tab inlf lpm_c
di (203 + 350) / 753

* Logit
logit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

* Probit
probit inlf nwifeinc educ exper expersq age kidslt6 kidsge6
margins, dydx(*)

/* Advanced Stata commands (please skip) */
scalar xbetaK = _b[_cons] + (_b[nwifeinc] * 20.13) + (_b[educ] * 12.3) ///
               + (_b[exper] * 10.6) + (_b[expersq] * (10.6^2)) ///
               + (_b[age] * 42.5) + (_b[kidslt6] * 0) + _b[kidsge6]
scalar cdf_value_0 = normal(xbetaK + (_b[kidslt6] * 0))
scalar cdf_value_1 = normal(xbetaK + (_b[kidslt6] * 1))
scalar cdf_value_2 = normal(xbetaK + (_b[kidslt6] * 2))
scalar list
di cdf_value_1 - cdf_value_0
di cdf_value_2 - cdf_value_1

qui log close
```

```
.
. /* Example 17.1 */
.
. use MROZ.DTA

. des inlf nwifeinc educ exper expersq age kidslt6 kidsge6

              storage  display     value
variable name   type   format      label      variable label
--------------------------------------------------------------------------------
inlf            byte   %9.0g                  =1 if in lab frce, 1975
nwifeinc        float  %9.0g                  (faminc - wage*hours)/1000
educ            byte   %9.0g                  years of schooling
exper           byte   %9.0g                  actual labor mkt exper
expersq         int    %9.0g                  exper^2
age             byte   %9.0g                  woman's age in yrs
kidslt6         byte   %9.0g                  # kids < 6 years
kidsge6         byte   %9.0g                  # kids 6-18

. sum inlf nwifeinc educ exper expersq age kidslt6 kidsge6

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        inlf |       753    .5683931    .4956295          0          1
    nwifeinc |       753    20.12896     11.6348  -.0290575         96
        educ |       753    12.28685    2.280246          5         17
       exper |       753    10.63081     8.06913          0         45
     expersq |       753    178.0385    249.6308          0       2025
-------------+--------------------------------------------------------
         age |       753    42.53785    8.072574         30         60
     kidslt6 |       753    .2377158     .523959          0          3
     kidsge6 |       753    1.353254    1.319874          0          8

.
. * LPM
. regress inlf nwifeinc educ exper expersq age kidslt6 kidsge6, robust

Linear regression                                      Number of obs =     753
                                                       F(  7,   745) =   62.48
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2642
                                                       Root MSE      =  .42713
```

```
------------------------------------------------------------------------------
             |               Robust
        inlf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0034052   .0015249    -2.23   0.026    -.0063988   -.0004115
        educ |   .0379953    .007266     5.23   0.000      .023731    .0522596
       exper |   .0394924     .00581     6.80   0.000     .0280864    .0508983
     expersq |  -.0005963     .00019    -3.14   0.002    -.0009693   -.0002233
         age |  -.0160908    .002399    -6.71   0.000    -.0208004   -.0113812
     kidslt6 |  -.2618105   .0317832    -8.24   0.000    -.3242058   -.1994152
     kidsge6 |   .0130122   .0135329     0.96   0.337     -.013555    .0395795
       _cons |   .5855192   .1522599     3.85   0.000     .2866098    .8844287
------------------------------------------------------------------------------

. predict lpm, xb

. summ lpm if lpm < 0 | lpm > 1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         lpm |        33    .4791733    .5917618  -.3451103   1.127151

. gen lpm_c = (lpm >= .5)

. tab inlf lpm_c

  =1 if in |
 lab frce, |         lpm_c
      1975 |         0          1 |     Total
-----------+----------------------+----------
         0 |       203        122 |       325
         1 |        78        350 |       428
-----------+----------------------+----------
     Total |       281        472 |       753


. di (203 + 350) / 753
.73439575

.
. * Logit
. logit inlf nwifeinc educ exper expersq age kidslt6 kidsge6
```

```
Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -402.38502
Iteration 2:   log likelihood = -401.76569
Iteration 3:   log likelihood = -401.76515
Iteration 4:   log likelihood = -401.76515

Logistic regression                               Number of obs   =        753
                                                  LR chi2(7)      =     226.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -401.76515                       Pseudo R2       =     0.2197

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0213452   .0084214    -2.53   0.011    -.0378509   -.0048394
        educ |   .2211704   .0434396     5.09   0.000     .1360303    .3063105
       exper |   .2058695   .0320569     6.42   0.000     .1430391    .2686999
     expersq |  -.0031541   .0010161    -3.10   0.002    -.0051456   -.0011626
         age |  -.0880244    .014573    -6.04   0.000     -.116587   -.0594618
     kidslt6 |  -1.443354   .2035849    -7.09   0.000    -1.842373   -1.044335
     kidsge6 |   .0601122   .0747897     0.80   0.422     -.086473    .2066974
       _cons |   .4254524   .8603697     0.49   0.621    -1.260841    2.111746
------------------------------------------------------------------------------

.
. * Probit
. probit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -402.06651
Iteration 2:   log likelihood = -401.30273
Iteration 3:   log likelihood = -401.30219
Iteration 4:   log likelihood = -401.30219

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

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
    nwifeinc |  -.0120237   .0048398    -2.48   0.013    -.0215096   -.0025378
        educ |   .1309047   .0252542     5.18   0.000     .0814074     .180402
       exper |   .1233476   .0187164     6.59   0.000     .0866641    .1600311
     expersq |  -.0018871      .0006    -3.15   0.002     -.003063   -.0007111
         age |  -.0528527   .0084772    -6.23   0.000    -.0694678   -.0362376
     kidslt6 |  -.8683285   .1185223    -7.33   0.000    -1.100628    -.636029
     kidsge6 |    .036005   .0434768     0.83   0.408     -.049208    .1212179
       _cons |   .2700768    .508593     0.53   0.595    -.7267473    1.266901
------------------------------------------------------------------------------

. margins, dydx(*)

Average marginal effects                        Number of obs     =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()
dy/dx w.r.t. : nwifeinc educ exper expersq age kidslt6 kidsge6

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0036162   .0014414    -2.51   0.012    -.0064413   -.0007911
        educ |   .0393703   .0072216     5.45   0.000     .0252161    .0535244
       exper |   .0370974   .0051522     7.20   0.000     .0269993    .0471956
     expersq |  -.0005675   .0001771    -3.20   0.001    -.0009146   -.0002204
         age |  -.0158957   .0023587    -6.74   0.000    -.0205186   -.0112728
     kidslt6 |  -.2611542   .0318597    -8.20   0.000    -.3235982   -.1987103
     kidsge6 |   .0108287   .0130584     0.83   0.407    -.0147654    .0364227
------------------------------------------------------------------------------

.
. /* Advanced Stata commands (please skip) */
. scalar xbetaK = _b[_cons] + (_b[nwifeinc] * 20.13) + (_b[educ] * 12.3) ///
>                 + (_b[exper] * 10.6) + (_b[expersq] * (10.6^2)) ///
>                 + (_b[age] * 42.5) + (_b[kidslt6] * 0) + _b[kidsge6]

. scalar cdf_value_0 = normal(xbetaK + (_b[kidslt6] * 0))

. scalar cdf_value_1 = normal(xbetaK + (_b[kidslt6] * 1))

. scalar cdf_value_2 = normal(xbetaK + (_b[kidslt6] * 2))
```

```
. scalar list
cdf_value_2 =  .11251305
cdf_value_1 =  .36506868
cdf_value_0 =  .69964706
     xbetaK =  .52338569

. di cdf_value_1 - cdf_value_0
-.33457838

. di cdf_value_2 - cdf_value_1
-.25255563

.
. qui log close
```

Example (MROZ.DTA)

- In the LPM, one more small child is estimated to reduce the probability of labor force participation by about 0.262 (26.2 percentage points), regardless of how many young children the woman already has (and regardless of the levels of the other explanatory variables).
- In the Probit model, let us consider a woman with $nwifeinc = 20.13$, $educ = 12.3$, $exper = 10.6$, $age = 42.5$, (these are roughly the sample averages), and $kidsge6 = 1$. For a woman with these characteristics, what is the estimated decrease in the probability of working in going from zero to one child? This will be about 0.334, that is the labor force participation probability is about 33.4 percentage points lower when a woman has one young child. If a woman goes from one to two children, the probability falls even more, but the marginal effect is not as large (-0.256 or about 25.5 percentage points).