# Using `margins` to estimate partial effects


David M. Drukker

Director of Econometrics
Stata



2010 Italian Stata Users Group meeting
Bologna
November 2010

# Overview

- This talk shows how to use the `margins` command to estimate the mean of the partial effects and the partial effects at the mean
- This talk highlights some important points about estimating partial effects
    - In nonlinear models, the partial effect at the mean can differ significantly from the mean of the partial effect
    - Standard parameter estimators; such maximum-likelihood, least squares, and generalized method of moments; only require a missing-at-random assumption, but estimating the mean of the partial effects requires a missing-completely-at-random assumption
- This talk will also illustrate some basic uses of Stata's factor variables

# Factor variable syntax

- Stata supports operators for factor variables

| | |
|---|---|
| `i.` | unary operator to specify indicators |
| `c.` | unary operator to treat as continuous |
| `#` | binary operator to specify interactions |
| `##` | binary operator to specify factorial interactions |

# Earnings data

```
. use earn2b
. summarize age
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         age |      7373     40.1968    13.22641         15         80
. tabulate educ3 hourly
                      |        hourly
                educ3 | nonhourly     hourly |     Total
----------------------+----------------------+----------
No high school diplom |       192        766 |       958
  HIGH SCHOOL DIPLOMA |       616      1,641 |     2,257
SOME COLLEGE NO DEGRE |       472        945 |     1,417
ASSOCIATE OCCUPATIONA |       122        244 |       366
   ASSOCIATE ACADEMIC |       110        133 |       243
    BACHELOR´S DEGREE |       987        369 |     1,356
      MASTER´S DEGREE |       447         89 |       536
  PROFESSIONAL DEGREE |       110         18 |       128
     DOCTORATE DEGREE |       104          8 |       112
----------------------+----------------------+----------
                Total |     3,160      4,213 |     7,373
```

# regress with factor variables

```
. regress lnearn age c.age#c.age i.educ3 i.hourly
      Source |       SS       df       MS              Number of obs =    7352
-------------+------------------------------           F( 11,  7340) =  353.39
       Model |  1866.97842    11  169.725311           Prob > F      =  0.0000
    Residual |  3525.27413  7340  .480282579           R-squared     =  0.3462
-------------+------------------------------           Adj R-squared =  0.3453
       Total |  5392.25255  7351  .733540001           Root MSE      =  .69302

------------------------------------------------------------------------------
      lnearn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .1284447   .0034719    37.00   0.000     .1216388    .1352507
             |
 c.age#c.age |  -.0013821   .0000405   -34.09   0.000    -.0014615   -.0013026
             |
       educ3 |
          3  |   .3663099   .0272751    13.43   0.000     .3128428     .419777
          4  |   .3965967   .0293683    13.50   0.000     .3390264     .454167
          5  |   .5247704   .0432303    12.14   0.000     .4400267    .6095141
          6  |   .5574536   .0505165    11.04   0.000     .4584268    .6564805
          7  |   .7062318   .0314011    22.49   0.000     .6446767     .767787
          8  |   .7281191   .0398533    18.27   0.000     .6499951    .8062431
          9  |   .9653706   .0666575    14.48   0.000     .8347028    1.096038
         10  |   .8957075   .0708855    12.64   0.000     .7567514    1.034663
             |
    1.hourly |  -.2135234   .0186841   -11.43   0.000    -.2501496   -.1768972
       _cons |   3.373212   .0719173    46.90   0.000     3.232233    3.514191
------------------------------------------------------------------------------
```

# Use coeflegend option to see parameter names

```
. regress lnearn age c.age#c.age i.educ3 i.hourly, coeflegend
      Source |       SS       df       MS              Number of obs =    7352
-------------+------------------------------           F( 11,  7340) =  353.39
       Model |  1866.97842    11  169.725311           Prob > F      =  0.0000
    Residual |  3525.27413  7340  .480282579           R-squared     =  0.3462
-------------+------------------------------           Adj R-squared =  0.3453
       Total |  5392.25255  7351  .733540001           Root MSE      =  .69302

------------------------------------------------------------------------------
      lnearn |      Coef.  Legend
-------------+----------------------------------------------------------------
         age |   .1284447  _b[age]
             |
 c.age#c.age |  -.0013821  _b[c.age#c.age]
             |
       educ3 |
          3  |   .3663099  _b[3.educ3]
          4  |   .3965967  _b[4.educ3]
          5  |   .5247704  _b[5.educ3]
          6  |   .5574536  _b[6.educ3]
          7  |   .7062318  _b[7.educ3]
          8  |   .7281191  _b[8.educ3]
          9  |   .9653706  _b[9.educ3]
         10  |   .8957075  _b[10.educ3]
             |
    1.hourly |  -.2135234  _b[1.hourly]
       _cons |   3.373212  _b[_cons]
------------------------------------------------------------------------------
```

# interaction syntax

```
. regress lnearn i.educ3 c.age#c.age c.age##i.hourly, vsquish
      Source |       SS           df       MS      Number of obs   =      7352
-------------+----------------------------------   F( 12,   7339)  =    325.59
       Model |  1873.37108        12  156.114257   Prob > F        =    0.0000
    Residual |  3518.88146      7339  .479476967   R-squared       =    0.3474
-------------+----------------------------------   Adj R-squared   =    0.3464
       Total |  5392.25255      7351  .733540001   Root MSE        =    .69244

------------------------------------------------------------------------------
      lnearn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       educ3 |
          3  |   .3672511   .0272535    13.48   0.000     .3138264    .4206757
          4  |   .3954825   .0293452    13.48   0.000     .3379575    .4530076
          5  |    .525017    .043194    12.15   0.000     .4403442    .6096897
          6  |   .5596144   .0504776    11.09   0.000     .4606638    .6585649
          7  |   .7089366   .0313835    22.59   0.000     .6474159    .7704572
          8  |   .7212365   .0398645    18.09   0.000     .6430907    .7993824
          9  |   .9621752   .0666073    14.45   0.000     .8316057    1.092745
         10  |   .8775882   .0709997    12.36   0.000     .7384085    1.016768
 c.age#c.age |  -.0014171   .0000416   -34.04   0.000    -.0014987   -.0013355
         age |   .1345898   .0038557    34.91   0.000     .1270315    .1421481
    1.hourly |  -.0082572   .0592347    -0.14   0.889    -.1243743    .1078599
hourly#c.age |
          1  |  -.0049327   .0013509    -3.65   0.000    -.0075809   -.0022845
       _cons |   3.178811   .0894314    35.54   0.000       3.0035    3.354122
------------------------------------------------------------------------------
```

# A model for binary data

- The probit model for binary data is one of the most widely used nonlinear models
- The dependent variable $y_i$ that we observe takes on values 0 and 1.
- One way to model this process is assume that there is a latent continuous variable $y_i^*$ such that

$$y_i = \begin{cases} 1 & \text{if } y_i* = \mathbf{x}_i\beta + \epsilon_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

- Specifying $Pr(y = 1|\mathbf{x}) = F(\mathbf{x}\beta)$ to be the cumulative distribution for $\epsilon_i$ conditional on $\mathbf{x}$ yields

$$\begin{aligned} Pr(y^* > 0|\mathbf{x}) &= Pr(\epsilon > -\mathbf{x}\beta|\mathbf{x}) \\ &= Pr(\epsilon < \mathbf{x}\beta|\mathbf{x}) \quad (\text{if } \epsilon \text{ has a symmetric distribution}) \\ &= F(\mathbf{x}\beta) \end{aligned}$$

# Estimation and inference in the probit model

- After choosing a distribution function, we have a fully specified parametric model
- Maximum-likelihood is the estimation framework most often applied
- Using the standard normal distribution for $F(\mathbf{x}\beta)$ yields the probit model

# Accident data

- We have some (fictional) data on individuals and whether or not they have had an accident in the last year
  - `crash` is 1 if person has been the driver in an accident in the last year
  - `cvalue` is the value of the person's car
  - `kids` is the number of children (under 18) for which the person is a guardian
  - `tickets` is the number of tickets the individual has received in the last three years
  - `male` is 1 if the person is male

# Probit example

```
. use accidents2

. probit crash tickets traffic i.male, nolog

Probit regression                               Number of obs     =        948
                                                LR chi2(3)        =     720.22
                                                Prob > chi2       =     0.0000
Log likelihood = -60.522949                     Pseudo R2         =     0.8561

------------------------------------------------------------------------------
       crash |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     tickets |   2.464657   .2768335     8.90   0.000     1.922073     3.00724
     traffic |    .159089   .0604682     2.63   0.009     .0405735    .2776045
      1.male |   5.892127   .7758214     7.59   0.000     4.371545    7.412709
       _cons |  -12.63666   1.529302    -8.26   0.000    -15.63403   -9.639279
------------------------------------------------------------------------------
Note: 516 failures and 13 successes completely determined.

. estimates store probit1
```

# Interpreting the estimated parameters

- The sign of the coefficient gives the direction of the effect, but not the marginal effect
- The estimated coefficients estimate $\frac{\beta}{\sigma}$, so their magnitudes are in units of the standard-deviation of the errors
- Marginal effect at a point $\tilde{\mathbf{x}}$ is $\frac{\partial E[y|\mathbf{x}]}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\tilde{\mathbf{x}}} = \frac{\partial F(\mathbf{x}\boldsymbol{\beta})}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\tilde{\mathbf{x}}} = f(\tilde{\mathbf{x}}\boldsymbol{\beta})\boldsymbol{\beta}$
- The relative marginal effects do not depend $\mathbf{x}$

$$\frac{\frac{\partial F(\mathbf{x}\boldsymbol{\beta})}{\partial x_j}}{\frac{\partial F(\mathbf{x}\boldsymbol{\beta})}{\partial x_k}} = \frac{f(\mathbf{x}\boldsymbol{\beta})\beta_j}{f(\mathbf{x}\boldsymbol{\beta})\beta_k} = \frac{\beta_j}{\beta_k}$$

- Use `testnl` to test hypotheses about the relative effects

```
. testnl _b[1.male]/_b[tickets] = 2
  (1)   _b[1.male]/_b[tickets] = 2
               chi2(1) =        8.86
           Prob > chi2 =        0.0029
```

# Marginal effects

- The good thing about marginal effects at point $\tilde{\mathbf{x}}$ is that all the information we need for estimation and inference about the marginal effect is contained in the ML point estimates and estimated VCE
- The bad thing about marginal effects at point $\tilde{\mathbf{x}}$ is that we must choose $\tilde{\mathbf{x}}$
- Use `margins` to estimate marginal effects at a point $\tilde{\mathbf{x}}$
- Conventionally, $\tilde{\mathbf{x}} = \bar{\mathbf{x}}$ when the variables in $\mathbf{x}$ are continuous
- See [Long and Freese(2006)] for more about interpreting the parameter estimates from cross-sectional binary-model regressions

# Marginal effects at means via margins

```
. margins , dydx(tickets traffic) atmeans
Conditional marginal effects                      Number of obs   =        948
Model VCE    : OIM

Expression   : Pr(crash), predict()
dy/dx w.r.t. : tickets traffic
at           : tickets         =    1.436709 (mean)
               traffic         =    5.201121 (mean)
               0.male          =    .5327004 (mean)
               1.male          =    .4672996 (mean)
```

| | dy/dx | Delta-method Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| tickets | 2.45e-07 | 8.06e-07 | 0.30 | 0.762 | -1.34e-06 | 1.82e-06 |
| traffic | 1.58e-08 | 5.14e-08 | 0.31 | 0.759 | -8.49e-08 | 1.17e-07 |

- Note the small effect of tickets and traffic

# Marginal effects at means by hand

```
. estat summarize

  Estimation sample probit                Number of obs =      948

  -------------------------------------------------------------
      Variable |       Mean   Std. Dev.         Min        Max
  -------------+-----------------------------------------------
         crash |   .1624473    .3690553           0          1
       tickets |   1.436709    1.849456           0          7
       traffic |   5.201121    2.924058     .005189    9.99823
        1.male |   .4672996    .4991929           0          1
  -------------------------------------------------------------

. matrix list r(stats)

r(stats)[4,4]
              mean          sd         min         max
  crash  .16244726   .36905531           0           1
tickets  1.4367089   1.8494562           0           7
traffic  5.2011207   2.9240582   .00518857   9.9982338
 1.male  .46729958   .49919289           0           1

. matrix r = r(stats)

. scalar f1 = normalden(_b[tickets]*r[2,1]+_b[traffic]*r[3,1]             ///
>         +_b[1.male]*r[4,1] + _b[_cons])

. display f1*_b[tickets]
2.446e-07

. display f1*_b[traffic]
1.579e-08
```

# Discrete effects at means via margins

```
. margins , dydx(male) atmeans

Conditional marginal effects                      Number of obs     =        948
Model VCE    : OIM

Expression   : Pr(crash), predict()
dy/dx w.r.t. : 1.male
at           : tickets         =    1.436709 (mean)
               traffic         =    5.201121 (mean)
               0.male          =    .5327004 (mean)
               1.male          =    .4672996 (mean)
```

| | dy/dx | Delta-method Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| 1.male | .0087485 | .007247 | 1.21 | 0.227 | -.0054553 | .0229523 |

Note: dy/dx for factor levels is the discrete change from the base level.

# Discrete effects at means by hand

```
. estat summarize

  Estimation sample probit                    Number of obs =      948

      Variable |       Mean   Std. Dev.         Min          Max
               |
         crash |   .1624473    .3690553           0            1
       tickets |   1.436709    1.849456           0            7
       traffic |   5.201121    2.924058     .005189      9.99823
        1.male |   .4672996    .4991929           0            1

. matrix list r(stats)

r(stats)[4,4]
              mean          sd         min         max
  crash  .16244726   .36905531           0           1
tickets  1.4367089   1.8494562           0           7
traffic  5.2011207   2.9240582   .00518857   9.9982338
 1.male  .46729958   .49919289           0           1

. matrix r = r(stats)

. local xb0 = _b[tickets]*r[2,1]+_b[traffic]*r[3,1] + _b[_cons]

. display normal(`xb0´+_b[1.male]) - normal(`xb0´)
.00874852
```

# Average partial effects

- Average partial effect of $x_k$ is

$$\frac{\beta_k}{N}\sum_{i=1}^{N} f(\mathbf{x}_i\boldsymbol{\beta})$$

  if $x_k$ is continuous

- If $x_k$ is discrete, the average partial effect is the average of the discrete differences in the predicted probabilities

# Marginal effects at a point versus Average marginal effects

- A marginal effect at a point is an estimate of the marginal effect at chosen covariate values
  - The marginal effect for a given person
- An average marginal effect is an estimate of a population-averaged marginal effect
  - The mean marginal effect for a population
  - The distribution of the covariates must be representative to consistently estimate the population-averaged marginal effect
- Mean partial effects and marginal effects at the mean are different quantities and can produce different estimates
  - Let $g(\mathbf{x}) = \frac{\partial F(\mathbf{x})}{\partial x}$
  - $g()$ is nonlinear implies that $g(\bar{\mathbf{x}}) \xrightarrow{p} g(E[\mathbf{x}]) \neq E[g(\mathbf{x})] \xleftarrow{p} N^{-1}\sum_{i=1}^{N} g(\mathbf{x}_i)$

# Average marginal effects via margins

```
. margins , dydx(tickets traffic)

Average marginal effects                          Number of obs   =        948
Model VCE    : OIM

Expression   : Pr(crash), predict()
dy/dx w.r.t. : tickets traffic
```

| | dy/dx | Delta-method Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| tickets | .0857818 | .0031049 | 27.63 | 0.000 | .0796963 | .0918672 |
| traffic | .0055371 | .0020469 | 2.71 | 0.007 | .0015251 | .009549 |

- Note that these values are much larger than marginal effects at means
- Note that these estimates are statistically significant

# Average marginal effects by hand

```
. predict  double xb, xb
. generate double me_tickets = normalden(xb)*_b[tickets]
. generate double me_traffic = normalden(xb)*_b[traffic]
. summarize me_tickets me_traffic if e(sample)
```

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| me_tickets | 948 | .0857818 | .2090093 | 4.59e-35 | .9818822 |
| me_traffic | 948 | .0055371 | .0134912 | 2.96e-36 | .0633787 |

# Average discrete effects via margins

```
. margins , dydx(male)

Average marginal effects                          Number of obs   =        948
Model VCE     : OIM

Expression    : Pr(crash), predict()
dy/dx w.r.t. : 1.male

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      1.male |   .2092058   .0105149    19.90   0.000      .188597    .2298145
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

# Average discrete effects by hand

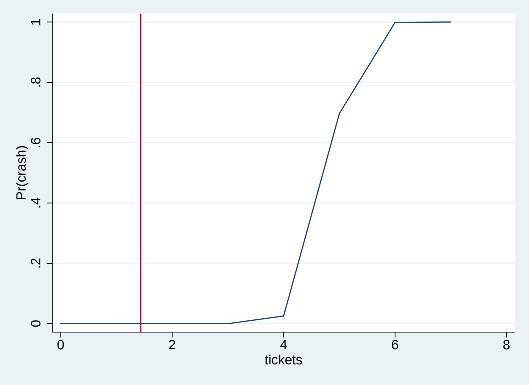
```
. generate double xb0 = _b[tickets]*tickets + _b[traffic]*traffic + _b[_cons]
. generate double de = normal(xb0 + _b[1.male]) - normal(xb0)
. summarize de
```

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| de | 948 | .2092058 | .3605846 | 7.79e-12 | .996267 |

# Treating tickets as discrete I

```
. estimates restore probit1
(results probit1 are active now)

. preserve

.         replace tickets   = _n-1 in 1/8
(7 real changes made)

.         replace male      = .4672996 in 1/8
(8 real changes made)

.         replace traffic   = 5.2011 in 1/8
(8 real changes made)

.         predict Fhat in 1/8
(option pr assumed; Pr(crash))
(940 missing values generated)

.         graph twoway line Fhat tickets in 1/8, xline(1.4367)

. restore
```

# Treating tickets as discrete II




- The mean of tickets is about 1.43, and the slope of the probabilty function is about zero when tickets is less than 3
- When tickets is greater than or equal to 3, the slope of the probabilty function is greater than 0

# Treating tickets as discrete III

```
. margins , at(tickets = (0 1 2 3)) post coeflegend
Predictive margins                              Number of obs     =        948
Model VCE    : OIM
Expression   : Pr(crash), predict()
1._at        : tickets         =           0
2._at        : tickets         =           1
3._at        : tickets         =           2
4._at        : tickets         =           3

------------------------------------------------------------------------------
             |     Margin  Legend
-------------+----------------------------------------------------------------
         _at |
          1  |   1.66e-09  _b[1bn._at]
          2  |   .0001208  _b[2._at]
          3  |   .0549183  _b[3._at]
          4  |   .4052946  _b[4._at]
------------------------------------------------------------------------------
```

# Treating tickets as discrete IV

```
. lincom _b[2._at] - _b[1bn._at]
 ( 1)  - 1bn._at + 2._at = 0
```

| | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| (1) | .0001208 | .0001671 | 0.72 | 0.470 | -.0002067 | .0004484 |

```
. lincom _b[3._at] - _b[2._at]
 ( 1)  - 2._at + 3._at = 0
```

| | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| (1) | .0547975 | .0177313 | 3.09 | 0.002 | .0200448 | .0895502 |

```
. lincom _b[4._at] - _b[3._at]
 ( 1)  - 3._at + 4._at = 0
```

| | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| (1) | .3503763 | .0225727 | 15.52 | 0.000 | .3061346 | .3946179 |

```
. estimates restore probit1
(results probit1 are active now)
```

# Treating tickets as discrete V

```
. generate double xb_b = _b[_cons] + _b[traffic]*traffic + _b[1.male]*male
. generate double pr0 = normal(xb_b + 0*_b[tickets])   // prob when tickets=0
. generate double pr1 = normal(xb_b + 1*_b[tickets])   // prob when tickets=1
. generate double pr2 = normal(xb_b + 2*_b[tickets])   // prob when tickets=2
. generate double pr3 = normal(xb_b + 3*_b[tickets])   // prob when tickets=3
. generate pe_d01 = pr1-pr0
. sum pe_d01

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      pe_d01 |        948    .0001208    .0003387   2.52e-24   .0031395
. generate pe_d12 = pr2-pr1
. sum pe_d12

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      pe_d12 |        948    .0547975    .0794281   1.05e-14   .3911403
. generate pe_d23 = pr3-pr2
. sum pe_d23

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      pe_d23 |        948    .3503763    .3749537   1.11e-07   .7821735
```

# Missing data and partial effects I

- ML estimators are consistent if some of the data is missing at random
  - Missing at random allows the mechanism that causes data to be missing to depend on the covariates $\mathbf{x}$ and a disturbance that is independent of everything else in the model
  - This is sometimes called selection on observables
  - See [Cameron and Trivedi(2005)] and [Wooldridge(2002)] for discussions and proofs
  - The sample distribution of the covariates need not be representative of the population distribution

# Missing data and partial effects II

- Estimating population averaged partial effects requires the much stronger assumption that the sample distribution of the covariates is representative
  - Missing completely at random guarantees that the sample distribution of the covariates is representative
  - Missing completely at random requires the mechanism that causes data to be independent of everything else in the model
  - In some cases, we can use weights to make the weighted sample covariate distribution representative
  - We need a representative sample of covariates for $N^{-1}\sum_{i=1}^{N} w_i g(\mathbf{x}_i) \xrightarrow{p} E[g(\mathbf{x})]$

# Missing data and partial effects III

- We also need a representative sample covariate distribution to estimate $E[x]$
- If we choose $\tilde{\mathbf{x}}$ in way that does not depend on our sample, we can perform estimation and inference for the partial effect at $\tilde{\mathbf{x}}$ because all the information we need is contained in the ML point estimates and estimated VCE, which only require missing at random

# Bibilography


- Cameron, A. Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and applications*, Cambridge: Cambridge University Press.
- Long, J. Scott and Jeremy Freese. 2006. *Regression models for categorical dependent variables using Stata*, College Station, Texas: Stata Press.
- Wooldridge, Jeffrey. 2002. *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Massachusetts: MIT Press.