# THE STATA JOURNAL

# Estimation of marginal effects using margeff

Tamás Bartus
Department of Sociology and Social Policy
Corvinus University, Budapest
tamas.bartus@uni-corvinus.hu

**Abstract.** This article describes the user-written program `margeff`, which enables the fast estimation of (average) marginal effects. Besides describing the program, this article offers a new discussion of some problems that are related to computation of marginal effects. I will argue that (1) marginal effects computed at means are not good approximations of average marginal effects, computed as means of marginal effects evaluated at each observations, if some of the parameter estimates are large; (2) both average marginal effects and marginal effects computed at means might produce wrong estimates for dummies that are part of a set of indicator variables indicating different categories of a single underlying variable; and (3) the use of marginal effects computed at means is preferred if some of the regressors are mathematical transformations of other regressors.

**Keywords:** st0086, margeff, mfx, average marginal effects, marginal effects at the mean, dummy variables, squared variables

## 1 Introduction

This article describes the user-written program `margeff`, which enables the fast estimation of (average) marginal effects. There are several reasons for using `margeff` instead of `mfx` or other user-written postestimation commands, such as the `spost` package. First, to my best knowledge, `margeff` is the first command that can estimate *average* marginal effects after a broad range of regression models. `margeff` can be viewed as a significant extension of Jonah B. Gelbach's `margfx`, which estimates average marginal effects after probit and logit models. Nevertheless, `margeff` can also estimate marginal effects at means. Second, `margeff` is fast and reliable because it calculates marginal effects analytically rather than numerically. Third, users might find `margeff` easier to use and its output easier to read. `margeff` presents the results as if marginal effects were real estimation results. For example, after multiple-outcome regression models, such as multinomial and ordered regression models, all possible outcomes are displayed.

Besides describing the program, this article offers a new discussion of some problems that are related to computation of marginal effects. There are two methods of estimating marginal effects. One method is the computation of the average of discrete or partial changes over all observations, yielding average marginal effects (hereafter abbreviated as AME). The other method is the computation of marginal effects at fixed values of the independent variables. The most-often used values are sample means. This method yields marginal effects at the mean (hereafter abbreviated as MEM). In the literature, there is not much discussion of which of these two methods should be used, and the discussion does not seem to be conclusive. The main argument in favor of AME is based

on a demand for realism: the sample means used during the calculation of MEM might refer to either nonexistent or inherently nonsensical observations, a problem typically encountered when there are dummies among the regressors (see, for example, Long 1997, 74). Greene also notes that current practice favors the use of AME. Ironically, the supply side does not seem to respond to this demand; most statistical software, such as Stata (and Limdep), estimates MEM. This adjustment problem can be rationalized on the grounds that MEM is a good (asymptotically valid) approximation of AME (Greene 1997, 876).

In my opinion, AME is more appropriate for providing a realistic interpretation of estimation results. Nevertheless, I feel that there is room for more discussion of how to estimate marginal effects. In this article, I will argue that MEM is not a good approximation of AME if some of the parameter estimates are large. Then I will turn to the issue of how to calculate marginal effects for dummy variables. The point I wish to make is that the main issue is not whether MEMs for dummy variables are unrealistic or nonsensical; rather, the main issue is that routine estimation of either AMEs or MEMs might lead to incorrect estimates for dummies that are part of a set of indicator variables indicating different categories of a single underlying variable. Finally, I will argue that the use of MEM is preferred if some of the regressors are mathematical transformations of other regressors.

The paper is organized as follows. Section 2 is devoted to the discussion of obtaining correct marginal effects. The section begins with an examination of the conditions under which AME should be used instead of MEM. I begin this section by showing that under some conditions MEM cannot be considered as a good approximation of AME. Then I discuss two practical issues of obtaining correct marginal effects for (i) dummy variables that indicate different categories of a single variable and (ii) continuous variables that are mathematical transformations of each other. I will argue that `mfx` does not offer an easy solution to both of these practical problems. Section 3 describes the `margeff` command. Section 4 contains examples mainly illustrating the problems examined in section 2. Section 5 gives some concluding notes.

## 2    Average marginal effect or marginal effect at the mean?

### 2.1    Definition of AME and MEM

Consider the single-equation regression model

$$E(y) = F(\beta x)$$

where $\beta x$ denotes the linear combination of parameters and variables and $F(\cdot)$ is the cumulative distribution function that maps the values of $\beta x$ to the $[\,0,1\,]$ interval.

Following the standard interpretation of linear statistical models, marginal effects should measure the change in the expected value of $y$ as one independent variable increases by unity while all other variables are kept constant. Then the AME of the $i$th explanatory variable is

$$\text{AME}_i = \frac{1}{n}\sum_{k=1}^{n}\big\{F(\beta x^k + \beta_i) - F(\beta x^k)\big\} \tag{1}$$

and the MEM is

$$\text{MEM}_i = F(\beta\overline{x} + \beta_i) - F(\beta\overline{x}) \tag{2}$$

For continuous variables, (1) and (2) are rarely applied. Instead, researchers estimate the effect of an infinitely small change. Let $f(\cdot)$ be the derivative of $F(\cdot)$ with respect to $\beta x$. Consider a continuous variable $x_i$. The AME of the $i$th continuous variable, $\text{AME}_i$ is given by

$$\text{AME}_i = \beta_i \frac{1}{n}\sum_{k=1}^{n} f(\beta x^k) \tag{3}$$

where $\beta x^k$ denotes the value of the linear combination of parameters and variables for the $k$th observation. Let $\overline{x}$ be a vector containing the means of the explanatory variables. The MEM for $x_i$ is defined as

$$\text{MEM}_i = \beta_i f(\beta\overline{x}) \tag{4}$$

There are also separate formulas to estimate marginal effects for dummy variables. The well-known formula for AME is

$$\text{AME}_i = \frac{1}{n}\sum_{k=1}^{n}\big\{F(\beta x^k \mid x_i^k = 1) - F(\beta x^k \mid x_i^k = 0)\big\} \tag{5}$$

In a similar fashion, one can define the MEM for dummy variables as

$$\text{MEM}_i = F(\beta\overline{x} \mid \overline{x} = 1) - F(\beta\overline{x} \mid \overline{x} = 0) \tag{6}$$

This definition is meaningful since it avoids the problem of setting dummy variables to means.

The remainder of this section provides a new discussion of the issue of obtaining correct marginal effects. In subsection 2.2, I show the conditions under which MEM cannot be considered as a good approximation of AME. Then I discuss two practical issues of obtaining correct marginal effects. In subsection 2.3, I show how to obtain correct marginal effects for dummy variables that indicate different categories of a single variable. Finally, subsection 2.4 argues that MEM should be used instead of AME if the model includes continuous variables that are mathematical transformations of each other. I will also argue that `mfx` does not offer an easy solution to the latter two problems.

## 2.2   MEM and AME: similarities and differences

In this subsection, I show the conditions under which MEM cannot be considered as a good approximation of AME. Although it was argued that MEM is a good approximation of AME (Greene 1997, 876), there is some evidence that MEMs and AMEs can be different, even in considerably large samples (Bockarjova and Hazans 2000).

Our exposition will focus on marginal effects for continuous variables. Using (3) and (4), the difference between the AME and the MEM can be written as

$$\text{AME}_i - \text{MEM}_i = \beta_i \frac{1}{n} \sum_{k=1}^{n} \left\{ f(\beta x^k) - f(\beta \overline{x}) \right\} \tag{7}$$

Using a second-order Taylor series expansion around $\overline{x}$, $f(\beta x^k) - f(\beta \overline{x})$ can be approximated as

$$f(\beta x^k) - f(\beta \overline{x}) \ f'(\beta \overline{x})(\beta x^k - \beta \overline{x}) + \frac{1}{2} f''(\beta \overline{x})(\beta x^k - \beta \overline{x})^2 \tag{8}$$

where $f'(\cdot)$ and $f''(\cdot)$ are the first and second derivatives of $f(\cdot)$ with respect to $\beta x$. Substituting (8) into (7) yields

$$\text{AME}_i - \text{MEM}_i \approx \beta_i \frac{1}{n} \sum_{k=1}^{n} \{ f'(\beta \overline{x})(\beta x^k - \beta \overline{x}) + \frac{1}{2} f''(\beta \overline{x})(\beta x^k - \beta \overline{x})^2 \} \tag{9}$$

Note that the difference between AME and MEM does not depend on the first term of the Taylor expansion because the sample sum of $(x_i^k - \overline{x}_i)$ is zero. Thus the *absolute* difference between AME and MEM reduces to

$$\text{AME}_i - \text{MEM}_i \approx \frac{1}{2} \beta_i f''(\beta \overline{x}) \text{Var}(\beta x) \tag{10}$$

where $\text{Var}(\beta x)$ denotes the sample variance of the linear prediction. The *relative* difference between AME and MEM is

$$\frac{\text{AME}_i - \text{MEM}_i}{\text{MEM}_i} \approx \frac{1}{2} \frac{f''(\beta \overline{x})}{f(\beta \overline{x})} \text{Var}(\beta x) \tag{11}$$

In other words, the relative difference between AME and MEM is proportional to the value of the second derivative of the density function evaluated at the sample means and the sample variance of the linear prediction.

A similar result holds for dummy variables. For notational simplicity, write the above regression model as

$$E(y) = F(\delta D + \gamma z)$$

where $D$ is dummy variable and $z$ is a vector of other regressors. Using two second-order Taylor series expansions around $\gamma \overline{z} + \delta$ and $\gamma \overline{z}$, the difference between (5) and (6) can be written as

$$\text{AME}_D - \text{MEM}_D \approx \frac{1}{2} \triangle F''(\gamma \overline{z}, \delta) \text{Var}(\gamma z) \qquad (12)$$

where

$$\triangle F''(\gamma \overline{z}, \delta) = F''(\gamma \overline{z} + \delta) - F''(\gamma \overline{z})$$

Equations (10)–(12) have several important implications. First, we already noted that the difference between AME and MEM does not depend on the first term of the Taylor expansion because the sample sum of $(x_i^k - \overline{x}_i)$ is zero. However, if the marginal effect were evaluated at another vector $\xi$, which would contain any statistics of the corresponding explanatory variables or a vector containing any constants, the sample sum of $(x_i^k - \xi_i)$ would not be zero. Therefore, the marginal effects computed at means are the best approximation of average marginal effects among those approximations that evaluate marginal effects at fixed values.

Second, MEM might both underestimate and overestimate AME, depending solely on the sign of the second derivative of the density function. Consider, for example, the probit model. Let $\Phi(x)$ and $\varphi(x)$ denote the cumulative distribution function and the density function of the standard normal distribution. The second derivative of $\varphi(x)$ is $(x^2 - 1)\varphi(x)$. The latter expression is zero if $x = -1$ or $x = 1$. Since $\Phi(-1) \approx 0.16$ and $\Phi(1) \approx 0.84$, the second derivative of $\varphi(x)$ is negative if the predicted probability evaluated at the sample means lies between 0.16 and 0.84. Hence, MEM overestimates AME if the predicted probability evaluated at the sample means lies between 0.16 and 0.84. If, however, the predicted probability evaluated at the sample means is larger than 0.84 or smaller than 0.16, MEM underestimates AME. Note also that the difference between MEM and AME depends on the value of the second derivative. Consider again the probit model. If the predicted probability evaluated at the sample means is around 0.5, the second derivative of the density function is around 0.4, and thus the difference between AME and MEM might be quite large. The difference between AME and MEM can be neglected if the predicted probability evaluated at the sample means is close to either 0.16 or 0.84.

Third, the difference between AME and MEM increases with the variance of the linear prediction. The variance of the linear prediction reflects the heterogeneity of individual marginal effects. The variance of the linear prediction is large if the parameter estimates are large. This implies that, keeping all other factors constant, the difference between AME and MEM is large when the parameter estimates are large, as well.

To illustrate the argument, consider the Program Effectiveness data listed in table 19.1 in Greene (1997). The data can be downloaded from the URL *http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm*. The model to be fitted is a probit model of `grade` on `gpa`, `tuce`, and `psi`. The parameter estimates are not shown here because they are the same as those reported by Greene in table 19.2. After fitting the probit model, I estimated both MEM and AME. The respective estimation results are as follows:

```
. mfx compute
Marginal effects after probit
      y  = Pr(grade) (predict)
         =   .26580809
```

| variable | dy/dx | Std. Err. | z | P>\|z\| | [ | 95% C.I. | ] | X |
|---|---|---|---|---|---|---|---|---|
| gpa | .5333471 | .23246 | 2.29 | 0.022 | .077726 | .988968 | | 3.11719 |
| tuce | .0169697 | .02712 | 0.63 | 0.531 | -.036184 | .070123 | | 21.9375 |
| psi* | .464426 | .17028 | 2.73 | 0.006 | .130682 | .79817 | | .4375 |

```
(*) dy/dx is for discrete change of dummy variable from 0 to 1
. margeff
Average marginal effects on Prob(grade==1) after probit
```

| grade | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| gpa | .3607863 | .1129461 | 3.19 | 0.001 | .1394161 | .5821565 |
| tuce | .0114793 | .0184085 | 0.62 | 0.533 | -.0246007 | .0475592 |
| psi | .3737518 | .1399912 | 2.67 | 0.008 | .0993741 | .6481295 |

There is a substantial difference in MEM and AME for two variables, `gpa` and `psi`. The relative difference between AME and MEM, as measured by the left-hand side of (11), is about $-32$ percent for `gpa` and about $-20$ percent for `psi`. Researchers guessing the AMEs on the basis of MEMs would considerably overestimate the AMEs.

Why do MEMs overestimate AMEs? Equation (11) implies that MEM is larger than AME if the value of the third derivative of the cumulative probability function evaluated at the means of independent variables is negative. Let $z$ be the sample mean of the linear prediction after the probit model. The second derivative of the standard normal density function $\varphi(z)$ is $(z^2 - 1)\varphi(z)$. Thus the right-hand side of (11) simplifies to $0.5(z^2 - 1)\text{Var}(z)$. This expression can be evaluated as follows:

```
. predict xb, xb
. summarize xb
. display (r(mean)^2-1)*r(Var)/2
```

The result is $-0.4$. On the basis of (11), the relative error is expected to be $-0.4$ for all variables. The observed relative errors are about $-0.32$ for `gpa` and `tuce` and $-0.20$ for `psi`. The observed relative errors are not identical to the expected relative errors. However, the expected value is relatively close to the observed value if we consider the continuous variables `gpa` and `tuce`. The large discrepancy between observed and expected values for the dummy variable `psi` is probably due to the fact that (11) is only approximately valid for dummy variables.

## 2.3   Dummy variables

In this subsection, I discuss the issue of how to obtain correct marginal effects for dummy variables. The point I wish to make is that the main issue is not whether MEMs for dummy variables are unrealistic or nonsensical; rather, the main issue is that routine estimation of either AMEs or MEMs might produce wrong estimates for dummy variables that indicate different categories of a single underlying variable. Sets of dummies are typically included if the explanatory variables include categorical variables having more than two categories. Stata users typically create sets of dummies using the `xi` command or the `xi:` prefix command.

It is generally viewed to be problematic to evaluate marginal effects at means of dummy variables since means of dummies refer to nonexisting observations. Note that the problem of nonsensical observations is not related to the estimation of marginal effects for dummy variables. In the previous section, we have seen that, even for dummy variables, MEMs are good approximations of AMEs, provided that the sample variance of the linear prediction and the second derivative of the cumulative distribution function evaluated at the linear prediction are both small. The argument that MEMs assume nonexisting data points can be rejected on the grounds that MEMs is a good approximation of the realistic AMEs, as long as these two conditions hold. Rather, the problem of nonsensical observations arises in the context of estimating marginal effects for continuous variables. The argument is that (4) cannot be used in the presence of dummy variables because the evaluation of $f(\beta \overline{x})$ would assume nonexistent and even nonsensical observations.

Assume for the moment that the demand for realism is compelling. Fortunately, realistic marginal effects can easily be computed. The demand for realism dictates that one should evaluate the density function so that the dummy in question is set either to zero or one. But this implies that one computes two separate quantities. This method is not attractive since if there are $K$ dummies, one should calculate $2K$ different marginal effects. This difficulty can be avoided if one computes a weighted average of the "realistic" marginal effects. In our simple model, the MEM of the continuous variable $z$ is

$$\text{MEM}_z = p\beta_z f(\gamma \overline{z} + \delta) + (1 - p)\beta_z f(\gamma \overline{z}) \tag{13}$$

where $p$ denotes the sample mean of $D$. To summarize, the claim that MEM is inappropriate in the presence of dummies boils down to the estimation of (13).

However, one can still insist on the computation of MEM in the presence of dummy variables on the basis that evaluating marginal effects at means of dummy variables might be a good approximation of meaningful average marginal effects. To see this, consider the simple regression model

$$E(y) = F(\delta D + \gamma z)$$

where $D$ is a dummy variable and $z$ is a continuous variable. First, note that (4), the formula for computing the MEM of the continuous variables, implies that the MEM of the continuous variable is $\gamma f(\gamma \overline{z} + \delta p)$. Since $f(\gamma \overline{z} + \delta p)$ can be approximated as $f(\gamma \overline{z}) + \delta p f'(\gamma \overline{z})$, (4) can be rewritten as

$$\text{MEM}_z \approx \beta_z f(\gamma \overline{z}) + \beta_z \delta p f'(\gamma \overline{z}) \tag{14}$$

The same formula can be obtained if one approximates $f(\gamma \overline{z} + \delta)$ as $f(\gamma \overline{z}) + \delta f'(\gamma \overline{z})$, and substitutes this approximation into (13)! This means that (13) is equivalent to (4), meaning that the usual formula of MEM for continuous variables can be used, even if there are dummies among the regressors. This result holds if the error terms of the first-order Taylor series expansions are approximately the same. Again the argument that MEMs assume nonexisting data points can be rejected on the grounds that MEMs are a good approximation of the realistic AMEs, as long as the neglected terms in the Taylor expansion are approximately the same.

Dummy variables raise a more fundamental problem if the regression model includes several dummies that indicate different categories of a single underlying variable. One implicit assumption behind the computation of marginal effects is that (5) or (6) should be applied to the full estimation sample. This implicit assumption might be labeled the "assumption of unrestricted computations". This assumption, however, leads to incorrect marginal effects if the computations involve a set of dummies indicating different categories of a single underlying variable.

To illustrate the problem, consider a $3 \times 2$ table, where the probabilities of success in the successive rows are 0.2, 0.5, and 0.8. This table can be generated using the following command:

```
. tabi 80 20  50 50  20 80, replace
```

Let the first category of the row variable be the reference row. Obviously, the effect of the second category of the row variable is 0.3 since the difference in probabilities of successes between the second and the first rows is 0.3. By the same token, the effect of the third category of the row variable is 0.6.

Surprisingly, marginal effects different from the expected ones are obtained if one uses the mfx compute command:

```
. quietly replace col = col-1

. quietly xi: logit col i.row [fw=pop]

. mfx compute

Marginal effects after logit
     y  = Pr(col) (predict)
        =          .5
```

| variable | dy/dx | Std. Err. | z | P>\|z\| | [ | 95% C.I. | ] | X |
|---|---|---|---|---|---|---|---|---|
| _Irow_2* | .3294081 | .06965 | 4.73 | 0.000 | .192896 | .465921 | | .333333 |
| _Irow_3* | .5798346 | .05297 | 10.95 | 0.000 | .476022 | .683647 | | .333333 |

```
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

What went wrong? Let $x_2, x_3, \ldots, x_K$ denote the dummy variables that indicate the second, third, ..., $K$th categories of the underlying variable $x$. The first category of $x$ is thus the reference category. Let $x_{-k}$ denote the vector of these dummies except $x_k$. The assumption of unrestricted computations is unrealistic for the following reason. Imagine that you wish to compute the average marginal effects of $x_2, x_3, \ldots, x_K$ dummies using (5). Intuitively, the marginal effect of the $k$th dummy should reflect the difference in the expected value of the outcome under study between the $k$th category of $x$ and the reference category. Thus the computations must be restricted to observations belonging to the first and the $k$th categories of $x$. This restriction is met if the computations are restricted to the observations where the equality $x_{-k} = 0$ holds. Since (5) does not incorporate this restriction, the marginal effects obtained take into account observations that are irrelevant in obtaining the true marginal effects. As a consequence, unrestricted marginal effects obtained using (5) might be incorrect.

Restricted calculations can be carried out with `mfx`, but the workaround is tedious. The appropriate restrictions can be added using either the `if` condition or the `at()` option. The use of the `if` condition seems more convenient. In our example, one should first issue the following two commands:

```
. mfx compute if row==2

Marginal effects after logit
     y  = Pr(col) (predict)
        =          .5
```

| variable | dy/dx | Std. Err. | z | P>\|z\| | [ | 95% C.I. | ] | X |
|---|---|---|---|---|---|---|---|---|
| _Irow_2* | .3 | .06403 | 4.69 | 0.000 | .174501 | .425499 | | 1 |
| _Irow_3* | .4411765 | .04357 | 10.13 | 0.000 | .355778 | .526575 | | 0 |

```
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

```
. mfx compute if row==3

Marginal effects after logit
      y  = Pr(col) (predict)
         =          .8
```

| variable | dy/dx | Std. Err. | z | P>|z| | [ | 95% C.I. | ] | X |
|---|---|---|---|---|---|---|---|---|
| _Irow_2* | .1411765 | .0316 | 4.47 | 0.000 | .079245 | .203108 | | 0 |
| _Irow_3* | .6 | .05657 | 10.61 | 0.000 | .489129 | .710871 | | 1 |

```
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

However, there is only one correct marginal effect in each of the above mfx outputs. Thus there is an additional task of combining the correct results in a separate table. As the number of dummies increases, this solution becomes more and more inefficient.

margeff implements a more-efficient procedure with the help of the dummies() option. One should issue only one command that produces the correct results:

```
. margeff, dummies(_I*)

Average marginal effects on Prob(col==1) after logit
```

| col | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| _Irow_2 | .3 | .0689653 | 4.35 | 0.000 | .1648305 | .4351695 |
| _Irow_3 | .6 | .0565681 | 10.61 | 0.000 | .4891286 | .7108714 |

What the dummies() option does is that it automatically assigns appropriate if conditions to each member of a list of dummies and then uses these conditions during the computations. For example, if margeff is told that $x_1$, $x_2$, and $x_3$ indicate different categories of a single variable, the conditions

```
if x2==0 & x3==0
if x1==0 & x3==0
if x1==0 & x2==0
```

are assigned in this order to the variables. Later, when margeff computes the marginal effect of each dummy, these conditions are used when (5) or (6) is evaluated.

## 2.4    Mathematical transformations

In this subsection, I argue that MEM rather than AME offers a more convenient solution to the problem of obtaining marginal effects if the model includes continuous variables that are mathematically transformations of each other. The typical example is the inclusion of age squared next to age. In such situations, researchers wish to know the "total" marginal effect of age, that is, the sum of the direct effect of age and its indirect effect mediated by age squared. Note that neither the marginal effect of age nor the marginal effect of age squared is the desired total effect because marginal effects assume the manipulation of only one variable and keeping all other variables constant. Clearly,

the idea of total marginal effect implies the idea of manipulating all mathematically related variables.

In principle, obtaining the total marginal effects is straightforward since direct and indirect effects are easy to obtain, and the rule of combining them is clear. Let $x$ be a continuous variable and $g(x)$ its mathematical transformation. The direct effect of $x$ is the effect as if $x$ influenced only the outcome but not $g(x)$. This direct effect is just the marginal effect of $x$. The indirect effect involves the effect of $g(x)$ on the outcome and the effect of $x$ on $g(x)$, while ignoring the fact that an increase in $x$ affects the outcome directly. Using the chain rule, the indirect effect is the marginal effect of $g(x)$ times the derivative of $g(x)$. Thus the total marginal effect of $x$ is the sum of marginal effects of variables, weighted by the derivatives of the variables with respect to $x$.

To state this result formally, let $z_j$ $(j = 1, \ldots, J)$ denote the $j$th mathematical transformation of $x_i$. The total marginal effects of $x_i$ are defined by

$$\text{AME}(x)^{\text{TOTAL}} = \sum_{(j)} \text{AME}\big\{g_j(x)\big\} \frac{dg_j(x)}{dx} \tag{15}$$

and

$$\text{MEM}(x)^{\text{TOTAL}} = \sum_{(j)} \text{MEM}\big\{g_j(\overline{x})\big\} \frac{dg_j(x)}{dx}\big| \ x = \overline{x} \tag{16}$$

Equations (15) and (16) immediately suggest that the total marginal effect of $x$ can be obtained in two steps. The first step is the (separate) estimation of marginal effects of each variable. The second step is the combination of the marginal effects using (15) or (16).

The comparison of (15) and (16) reveals that the combination of the marginal effects is much easier if one uses MEMs instead of AMEs. In (16), the derivative of any $g_j(x)$ is a *scalar*. For example, the derivative of age squared with respect to age is $2\overline{x}$. If the average age in the sample is, say, 40 years, then the total marginal effect of age is the marginal effect of age plus 80 times the marginal effect of age squared. Estimating MEMs makes the computation very easy. In general, if the user knows the sample means, the computation of the total marginal effect can make use of the `lincom` or `nlcom` commands. In this way, also the standard error for the total marginal effect can be obtained.

Contrary to this, the computation of AMEs is a bit more complicated. In (15), the derivative of any $g_j(x)$ is a *variable*. However, the total marginal effect must be a *scalar*. To arrive at a scalar, one could either take the mean of each derivative or evaluate the derivatives at means. The latter strategy shows again that computing MEMs is easier in this situation.

Finally, note that the official `mfx` command does not allow for a quick combination of marginal effects. The reason is that `mfx` is not an estimation command; thus Stata's

postestimation commands will not find the matrix of estimated marginal effects. Users of `mfx` can implement this two-step procedure provided they write a small program that begins with saving `mfx` results as proper estimation results and then proceeds with the appropriate `lincom` command.

To enable quick access to estimated marginal effects and their standard errors, `margeff` has a `replace` option that causes `margeff` to overwrite the estimation results. The `replace` option turns `margeff` into an estimation command. This has the additional advantage that researchers can quickly generate publishable tables of marginal effects using either the official `estimates table` command or the user-written commands `estout` and `outreg`.

# 3 The margeff command

## 3.1 Syntax

margeff $\left[\underline{\text{c}}\text{ompute}\right]$ $\left[\mathit{if}\right]$ $\left[\mathit{in}\right]$ $\left[\text{, at}(\mathit{atlist})\ \underline{\text{c}}\text{ount}\ \underline{\text{nodi}}\text{screte}\right.$
    $\underline{\text{d}}\text{ummies}(\mathit{varlist}_1\ \left[\backslash\ \mathit{varlist}_2\ \ldots\right])\ \underline{\text{e}}\text{form}\ \underline{\text{model}}(\mathit{stata\_cmd})\ \underline{\text{nooff}}\text{set}$
    $\underline{\text{out}}\text{come}(\mathit{numlist})\ \underline{\text{percent}}\ \underline{\text{replace}}\left.\right]$

margeff $\underline{\text{replay}}$

where *atlist* is

    $\{\text{mean}\,|\,\text{median}\,|\,\text{zero}\ \left[\mathit{varname}\ =\ \#\ \left[,\ \mathit{varname}\ =\ \#\right]\ \left[\ldots\right]\right]\}$
    $\{\left[\text{mean}\right]\,|\,\left[\text{median}\right]\,|\,\left[\text{zero}\right]\ \mathit{varname}\ =\ \#\ \left[,\ \mathit{varname}\ =\ \#\right]\ \left[\ldots\right]\}$

*varlist*$_1$ $\left[\backslash\ \mathit{varlist}_2\ \ldots\right]$ are lists of dummy variables, where all dummies of a list indicate different categories of the same underlying categorical variable

and *stata_cmd* is one of

| [group 1:] | probit | logit | logistic | cloglog | heckprob | xtprobit |
|---|---|---|---|---|---|---|
| [group 2:] | oprobit | ologit | gologit | mlogit | biprobit | |
| [group 3:] | poisson | nbreg | zip | zinb | | |

## 3.2 Description

`margeff` analytically estimates marginal effects and standard errors for marginal effects using the delta method. The default behavior of `margeff` is the following:

1. `margeff` calculates average marginal effects, that is, changes in the quantities of interest evaluated for each observations, and the reported marginal effects are sample

averages of these changes. Note that `margeff` can also compute marginal effects evaluated at sample means or at other values if the user specifies the `at(`*atlist*`)` option. Calculations are restricted to the estimation sample.

2. For continuous variables, marginal effects are partial changes in the quantities of interest. For dummy variables, marginal effects are discrete changes in the quantities of interest as the dummy variable changes from 0 to 1. Dummies are automatically detected. Before using `margeff`, please make sure that you coded your dummy variables as 0/1 variables. Otherwise, your dummies will be considered as being continuous variables, and you run the risk of obtaining misleading results. Users can change this behavior by specifying the `count` or `nodiscrete` options.

3. Quantities of interest are defined as follows:

   [group 1:] the probability of positive outcome

   [group 2:] the probabilities of all possible outcomes defined by the dependent variables

   [group 3:] the expected number of counts or the incidence rate

4. `margeff` supports only the above listed models. The `model(`*stata_cmd*`)` option, however, adds some flexibility. This is useful if users wish to obtain marginal effects after the survey version of *stata_cmd* or after their own program.

5. `margeff` behaves as a postestimation command (see `help postest`). However, the option `replace` forces `margeff` to behave as an estimation command (see `help estcom`). This enables the use of postestimation commands, such as `lincom` or `test` after `margeff`.

Typed without arguments, `margeff` replays the results of previous `margeff` computations, provided the last use of `margeff` was combined with the `replace` option.

`margeff replay` replays the results of the previous `margeff` computation.

## 3.3   Options

`at(`*atlist*`)` forces `margeff` to estimate marginal effects at points specified in *atlist* instead of estimating average marginal effects.

   `at(mean`|`median`|`zero` $\big[$ *varname* = $\#$ $\big[$ , *varname* = $\#$$\big]$ $\big[$ . . .$\big]$ $\big]$`)` specifies that the marginal effects be evaluated at means, at medians of the independent variables, or at zeros. It also allows users to specify particular values for one or more independent variables, assuming that the rest are means, medians, or zeros.

   `at(`$\big[$`mean`$\big]$|$\big[$`median`$\big]$|$\big[$`zero`$\big]$ *varname* = $\#$ $\big[$ , *varname* = $\#$$\big]$ $\big[$ . . .$\big]$`)` specifies that the marginal effects be evaluated at particular values for one or more independent variables, assuming that the rest are means.

`count` modifies the calculation of marginal effects for count variables, i.e., variables that take more than two values and all of the values are integers. By default, `margeff` treats count variables as continuous variables, thus marginal effects correspond to small changes in the independent variables. If the `count` option is specified, the marginal effects are changes in probabilities when the count variables increases by unity.

`nodiscrete` forces `margeff` to treat dummy variables as if they were continuous. Recall that if `nodiscrete` is not specified, the marginal effect of a dummy variable is calculated as the discrete change in the expected value of the dependent variable as the dummy variable changes from 0 to 1.

`dummies(`*varlist*$_1$ $\left[ \backslash \ varlist_2 \ \left[ \ldots \right] \right]$`)` modifies the calculation of marginal effects for dummy variables. Let $x$ be a categorical variables with $K + 1 (K > 1)$ categories. In this case, not $x$ but $K$ dummies—say, `D1`, ..., `DK`—are included in the regression model. The estimated marginal effects for these $K$ dummies may be misleading. The correct result is obtained if one specifies the `dummies(D*)` option.

`eform` forces `margeff` to define the quantity of interest as $\exp(\mathbf{x}\beta)$, where $\mathbf{x}\beta$ is the linear prediction.

`model(`*stata_cmd*`)` forces `margeff` to estimate marginal effects as if the preceding estimation command were *stata_cmd*. *stata_cmd* must be one of the supported commands that are listed in the *Syntax* subsection. The `model(`*stata_cmd*`)` is likely to be helpful if you wish to obtain marginal effects after the survey version of a supported estimation command or you wish to fit a supported model using `glm`. The estimates, however, will be meaningful only if the estimation command issued by the user and *stata_cmd* have the same link functions. It is the user's responsibility to ensure that this condition holds.

`nooffset` causes `margeff` to ignore the offset variable during the calculations.

`outcome(`*numlist*`)` causes `margeff` to display only the probability outcomes corresponding to the numbers specified in *numlist* (see [U] **11.1.8 numlist**). This option is useful after estimation commands listed as "[group 2]" models. The correspondence rule is the following:

   1. After `biprobit`, numbers 1, 2, 3, and 4 refer to outcomes p00, p01, p10, and p11.

   2. After `oprobit`, `ologit`, `gologit`, and `mlogit`, number $\#$ refers to the $\#$th category of the dependent variable. Thus number 1 always indicates the lowest category, and if the dependent variable has, say, 5 categories, number 5 indicates the highest category.

`percent` causes `margeff` to display the results as a percentage.

`replace` causes `margeff` to overwrite the estimation results left behind. This option is useful if

1. you wish to include marginal effects in publication-quality tables using either the official `estimates table` command or the user-written commands `outreg` or `estout`; or

2. your model contains a variable (such as age-squared) that is a mathematical transformation of another independent variable and you wish to obtain the total effect of that variable (age) using the `lincom` or the `nlcom` command.

## 3.4   Methods and Formulas

### Marginal effects after single-equation models

Estimation of marginal effects after single-equation models uses (1)–(6), depending on the type of variable and the contents of the `at(`*atlist*`)`, `count`, and `nodiscrete` options. If the `at(`*atlist*`)` option is not specified, (1), (3), and (5) are applied to count, continuous, and dummy variables, respectively. If the `at(`*atlist*`)` option is specified, (2), (4), and (6) are applied to count, continuous, and dummy variables. Note that the `count` option enables the perception of count variables, and the `nodiscrete` option forbids the perception of dummy variables.

### Marginal effects after multiple-equation models

If multiple-equation models are encountered, the computations are a bit different. The handling of multiple-equation models is shown here only in the context of estimating average marginal effects. Let $\beta_i^h$ denote the parameter estimate of $x_i$ in equation $h$ $(h = 1, \ldots, H)$. The symbol $\beta^h x$ refers to the linear combination of the parameter estimates and the variables in the $h$th equation. Similarly, $\beta^h x^k$ denotes the value of $\beta^h x$ for the $k$th observation in the estimation sample of size $n$ $(k = 1, \ldots, n)$. Given these notations, the multiple-equation models are defined as

$$E(y) = F(\beta^1 x, \ldots, \beta^H x)$$

Since partial changes are expressed by partial derivatives, the average marginal effect of the $i$th *continuous* variable is given by

$$\text{AME}_i = \frac{1}{n} \sum_{k=1}^{n} \sum_{h=1}^{H} \frac{\partial F_m(\beta^1 x, \ldots, \beta^H x)}{\partial \beta_h^1} \beta_h^1 \tag{17}$$

For *dummy* variables, the average marginal effect for the $i$th variable is defined as

$$\text{AME}_i = \frac{1}{n} \sum_{k=1}^{n} \left\{ F(\beta^1 x, \ldots, \beta^H x | \ x_i^k = 1) - F(\beta^1 x, \ldots, \beta^H x | \ x_i^k = 0) \right\} \tag{18}$$

Finally, the formula for *count* variables is

$$\text{AME}_i = \frac{1}{n} \sum_{k=1}^{n} \left\{ F(\beta^1 x + \beta_i^1, \ldots, \beta^H x + \beta_i^H) - F(\beta^1 x, \ldots, \beta^H x) \right\} \tag{19}$$

### Standard errors for marginal effects

`margeff` also provides estimates for the standard errors of marginal effects. Using a first-order Taylor-series expansion around the true values of parameters, one can show that the covariance matrix of marginal effects, denoted by $\mathbf{V}(\text{AME})$, is approximately $\mathbf{G}\mathbf{V}(\beta)\mathbf{G}^T$, where $\mathbf{G}$ contains the derivatives of marginal effects with respect to the parameter estimates and $\mathbf{V}(\beta)$ is the covariance matrix of parameter estimates. After estimation, only $\mathbf{V}(\beta)$ is available. $\mathbf{G}$ can be constructed as follows. The partial derivative of the $i$th marginal effect with respect to the $j$th parameter estimate is

$$\frac{\partial \text{AME}_i}{\partial \beta_j^h} = \frac{1}{n} \sum_{k=1}^{n} x_j^k \left\{ \triangle_{ij} f(\beta^1 x, \ldots, \beta^H x) + \beta_i^h f'(\beta^1 x, \ldots, \beta^H x) \right\} \tag{20}$$

if the $i$th variable is *continuous*;

$$\frac{\partial \text{AME}_i}{\partial \beta_j^h} = \frac{1}{n} \sum_{k=1}^{n} x_j^k \left\{ f(\beta^1 x, \ldots, \beta^H x |\, x_i^k = 1) - (1 - \triangle_{ij}) f(\beta^1 x, \ldots, \beta^H x |\, x_i^k = 0) \right\} \tag{21}$$

if the $i$th variable is *dummy*; and

$$\begin{aligned}
\frac{\partial \text{AME}_i}{\partial \beta_j^h} = {} & \triangle_{ij} \frac{1}{n} \sum_{k=1}^{n} f(\beta^1 x + \beta_i^1, \ldots, \beta^H x + \beta_i^H) \\
& + \frac{1}{n} \sum_{k=1}^{n} x_j \left\{ f(\beta^1 x + \beta_i^1, \ldots, \beta^H x + \beta_i^H) - f(\beta^1 x, \ldots, \beta^H x) \right\}
\end{aligned} \tag{22}$$

if the $i$th variable is a *count*. In (20)–(22), $\triangle_{ij} = 1$ if $i = j$, and $\triangle_{ij} = 0$, otherwise.

The derivative of the density function, $f'(\cdot)$ is calculated numerically. The computations use the simple formula

$$\frac{\partial f(\beta x)}{\partial \beta x} = \frac{1}{n\delta} \sum_{k=1}^{n} f(\beta x^k + \delta) - f(\beta x^k) \tag{23}$$

where $\delta$ is automatically set to $10^{-6}/(p95_i - p5_i)$, where $p95_i$ and $p5_i$ are the 95th and the 5th percentiles of the variable for which the marginal effects are estimated. However,

if the 95th and the 5th percentiles are equal, the maximum and minimum values are used instead.

## 3.5  Saved results

If `margeff` is used without the `replace` option, `margeff` adds the following macros and matrices to the existing estimation results:

Macros
| | |
|---|---|
| `e(margeff_cmd)` | *stata_cmd* |
| `e(margeff_display)` | display form; either `asis` or `percent` |
| `e(margeff_def)` | quantities of interest |
| `e(margeff_method)` | method of calculation; `average`, `mean`, `median`, or `zero` |
| `e(margeff_at)` | list of variables specified in `at(`*atlist*`)` |
| `e(margeff_discrete)` | `off` if the `nodiscrete` option was specified; otherwise, `on` |
| `e(margeff_count)` | `on` if the `count` option was specified; otherwise, `off` |
| `e(margeff_dvars)` | the list of dummy variables, independently of the treatment of dummies during calculation |
| `e(margeff_cvars)` | the list of count variables, independently of the treatment of count variables during calculation |

Matrices
| | |
|---|---|
| `e(margeff_b)` | vector of estimated marginal effects |
| `e(margeff_V)` | covariance matrix of estimated marginal effects |

If `margeff` is used as an estimation command—that is, the user does specify the `replace` option—`margeff` erases all locals and matrices, and, besides the above listed macros and matrices, returns the following macros and matrices:

Macros
| | |
|---|---|
| `e(cmd)` | `margeff` |
| `e(depvar)` | name of the dependent variable |

Matrices
| | |
|---|---|
| `e(b)` | vector of estimated marginal effects |
| `e(V)` | covariance matrix of estimated marginal effects |

`margeff` leaves the `e(sample)` function and all scalars unchanged.

# 4  Examples

## 4.1  Data and variables

The use of `margeff` will be illustrated with the help of the Scientific Productivity of Biochemists Data ($N = 915$). The dataset is described in Long (1997) and can be accessed through the URL *http://www.indiana.edu/~jslsoc/stata/socdata/couart2.dta*. Scientific productivity is measured as the number of published articles (variable `art`). During the subsequent analyses, we will use a dummy variable indicating the existence of any publication (`artbin`). Additionally, the square of the variable measuring the publications of the mentor was created (`ment2`). Our example session starts with the following Stata commands:

```
. use http://www.indiana.edu/~jslsoc/stata/socdata/couart2.dta
. generate byte artbin =  art > 0
. summarize ment
. generate ment2 = (ment-9)^2
```

The last two commands prepare the estimation of the total marginal effect of `ment`. Note that the `summarize` command shows that mean productivity of mentors is about 8.767 articles. Instead of this number, the square term is centered around 9. Centering is necessary if one wishes to obtain marginal effects after variables that are mathematical transformations of each other.

## 4.2    Marginal effects after sets of dummy variables

We begin by illustrating the estimation of marginal effects after dummy variables that indicate different categories of a common categorical variable. The dataset contains a variable measuring the number of small children (variable `kid5`). One might argue that the relationship between scientific productivity and the number of children is not linear but piecewise. That is, productivity does not drop substantially after having the first child, but it does drop substantially after having the second and especially the third child. To test this idea, the following logit model is fitted.

```
. xi: logit artbin i.kid5 ment phd fem mar
```
(*output omitted*)

| artbin | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| _Ikid5_1 | -.3161219 | .2151719 | -1.47 | 0.142 | -.7378511 | .1056072 |
| _Ikid5_2 | -.5904391 | .2602111 | -2.27 | 0.023 | -1.100443 | -.0804348 |
| _Ikid5_3 | -.769156 | .5325752 | -1.44 | 0.149 | -1.812984 | .2746721 |
| ment | .0802673 | .0130344 | 6.16 | 0.000 | .0547203 | .1058142 |
| phd | .0221113 | .0795652 | 0.28 | 0.781 | -.1338336 | .1780562 |
| fem | -.2520657 | .1592848 | -1.58 | 0.114 | -.5642582 | .0601267 |
| mar | .3369237 | .188536 | 1.79 | 0.074 | -.0326001 | .7064476 |
| _cons | .2367794 | .2955546 | 0.80 | 0.423 | -.342497 | .8160557 |

The parameter estimates suggest that our hypothesis is partially correct. On the one hand, a significant difference in productivity can be found between researchers having no small children and researchers having two small children, but having only one child does not affect productivity. On the other hand, evidence does not support the idea that researchers with three children are less productive than researchers having no small children. To improve on interpretation, average marginal effects are estimated using `margeff`. Since `_Ikid5_1`, `_Ikid5_2`, and `_Ikid5_3` indicate different categories of a single underlying variable, `kid5`, the `dummies()` option must also be specified. The command and the output are as follows:

```
. margeff, dummies(_I*)
Average marginal effects on Prob(artbin==1) after logit
```

| artbin | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| _Ikid5_1 | -.0629149 | .0437762 | -1.44 | 0.151 | -.1487147 | .0228849 |
| _Ikid5_2 | -.1224407 | .0562078 | -2.18 | 0.029 | -.2326059 | -.0122754 |
| _Ikid5_3 | -.164933 | .1206803 | -1.37 | 0.172 | -.4014621 | .071596 |
| ment | .0157561 | .0024035 | 6.56 | 0.000 | .0110453 | .0204669 |
| phd | .0043404 | .0156151 | 0.28 | 0.781 | -.0262648 | .0349455 |
| fem | -.0496628 | .0313712 | -1.58 | 0.113 | -.1111493 | .0118237 |
| mar | .067134 | .0378592 | 1.77 | 0.076 | -.0070687 | .1413367 |

Thus keeping other things equal, the difference in productivity between researchers who have no small children and researchers who have two small children is about 12 percent.

## 4.3 Obtaining correct marginal effects when the model includes polynomial terms

The next example illustrates how to compute correct marginal effects after models that include an independent variable, which is a mathematical transformation of another independent variable. The typical example is the inclusion of the square of another variable. Our aim is to access the hypothesis that the productivity of students is a nonlinearly increasing or inversely U-shaped function of the productivity of mentors. The logic behind this reasoning is that the productivity of the mentor motivates students, but only if the mentor does not publish too much, since increase in mentor's productivity decreases the time spent on supervising and encouraging students. The natural way of testing such nonlinear effects is the inclusion of the square of the variable that measures the productivity of the mentor. To assess this hypothesis, a logit model of `artbin` on `ment`, `ment2`, and other variables was fitted. The results are as follows:

```
. logit artbin ment ment2 phd fem kid5 mar
  (output omitted)
```

| artbin | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| ment | .0894668 | .0141288 | 6.33 | 0.000 | .0617748 | .1171587 |
| ment2 | -.0010174 | .000576 | -1.77 | 0.077 | -.0021464 | .0001115 |
| phd | .0111894 | .0800323 | 0.14 | 0.889 | -.145671 | .1680499 |
| fem | -.2524847 | .159493 | -1.58 | 0.113 | -.5650852 | .0601157 |
| kid5 | -.2849059 | .111405 | -2.56 | 0.011 | -.5032557 | -.0665562 |
| mar | .3352097 | .1815673 | 1.85 | 0.065 | -.0206557 | .691075 |
| _cons | .2528066 | .2959217 | 0.85 | 0.393 | -.3271892 | .8328024 |

The signs of the coefficients of `ment` and `ment2` support the hypothesis that the productivity of students depends nonlinearly on the productivity of mentors. The coefficient of the squared term is not significant at the conventional 5-percent level.

Before estimating marginal effects, it is useful to find out whether the effect is increasing or inversely U-shaped. The effect is inversely U-shaped if the value of `ment` that maximizes the linear prediction falls within the range of `ment`. The value of `ment` that maximizes the linear prediction can be calculated as follows.

```
. di -_b[ment]/_b[ment2]
87.933327
```

This value is outside the observed range of `ment` (the maximum of `ment` is 77). Thus the effect is not inversely U-shaped but is increasing at a decreasing rate, so the effect of `ment` is always positive.

The total marginal effect of `ment` will be evaluated at the point where the mentor has 9 publications. The estimation of marginal effects proceeds in first steps. First, we compute marginal effects at fixed values:

```
. margeff, at(mean ment=9 ment2=0) replace
Marginal effects at means on Prob(artbin==1) after logit
Variables set to specific values:       ment ment2
```

| artbin | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| ment | .0174711 | .0025626 | 6.82 | 0.000 | .0124485 | .0224937 |
| ment2 | -.0001987 | .0001116 | -1.78 | 0.075 | -.0004174 | .00002 |
| phd | .0021851 | .0156281 | 0.14 | 0.889 | -.0284455 | .0328157 |
| fem | -.0494358 | .0312446 | -1.58 | 0.114 | -.1106742 | .0118025 |
| kid5 | -.0556365 | .0214759 | -2.59 | 0.010 | -.0977285 | -.0135444 |
| mar | .0665377 | .0364026 | 1.83 | 0.068 | -.00481 | .1378855 |

The second step is the computation of the weighted sum of the marginal effects of `ment` and `ment2`. The weights for `ment` and `ment2` are 1 and 18, respectively. The `replace` option of `margeff` enables easy computations, since one can use the `lincom` command.

```
. lincom ment+2*9*ment2
 ( 1)  ment + 18 ment2 = 0
```

| artbin | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| (1) | .0138947 | .0021854 | 6.36 | 0.000 | .0096114 | .0181781 |

Thus a small increase in the number of articles by the mentor increases the probability of having published articles by about 1.4 percent. Without forming the linear combination, one would conclude that the effect size is a bit larger, 1.7 percent.

# 5   Concluding notes

This article describes the user-written program `margeff`, which enables the fast estimation of (average) marginal effects. I hope I have shown that `margeff` is an attractive

alternative to the official Stata command `mfx`. `margeff` is recommended to users who (1) insist that average marginal effects are more realistic than marginal effects at the mean; (2) often include sets of dummy variables or polynomial terms in their regression models; (3) wish to obtain marginal effects after models having several outcomes; or (4) wish to produce nice tables containing the marginal effects using either the official `estimates table` or the user-written `outreg` or `estout` commands. Additionally, due to the analytic computations, `margeff` is much faster and more reliable than `mfx`, which relies on numerical derivations.

However, `margeff` also has limitations. Ironically, the limitations are related to the fact that it calculates marginal effects analytically using the cumulative distribution functions and the density functions. Unfortunately, the programming of these functions might be challenging (at least it was for the author). The author encountered such a difficulty with `clogit` and `nlogit`. Thus there is a trade-off between computational speed and the number of supported estimation commands. `margeff` aims at quick estimation of marginal effects, but only after a limited number of regression models. Contrary to this, `mfx` aims at estimating marginal effects after a broad range of regression models, but at the price of relatively slow (numerical) calculations. `margeff` and `mfx` are thus the opposite ends of a scale.

# 6   Acknowledgments

The author wishes to thank the anonymous reviewer for helpful comments and suggestions. Some parts of the code and the help file are taken from the official `mfx.ado` and `mfx.hlp` files and from `margfx.ado`, written by Jonah B. Gelbach.

# 7   References

Bockarjova, M. and M. Hazans. 2000. *Marginal effects distributions in logit models of labour markets.* Paper presented at the seminar "Labor Markets, Work, and Welfare During the Transition and Integration Processes", Vilnius, Lithuania.

Greene, W. H. 1997. *Econometric Analysis.* 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage.

**About the Author**

Tamás Bartus is associate professor at the Department of Sociology and Social Policy, Corvinus University, Budapest. His research interests include the study of label market consequences of social networks and commuting distances.