The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

The *Stata Journal* is indexed and abstracted in the following:

- CompuMath Citation Index®
- Current Contents/Social and Behavioral Sciences®
- RePEc: Research Papers in Economics
- Science Citation Index Expanded (also known as SciSearch®)
- Scopus™
- Social Sciences Citation Index®

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press. Stata, Stata, Stata Press, Mata, Mata, and NetCourse are registered trademarks of StataCorp LP.

The Stata Journal (2012)
12, Number 2, pp. 308–331

Using the margins command to estimate and interpret adjusted predictions and marginal effects

Richard Williams
Department of Sociology
University of Notre Dame
Notre Dame, IN
Richard.A.Williams.5@nd.edu

Abstract. Many researchers and journals place a strong emphasis on the sign and statistical significance of effects—but often there is very little emphasis on the substantive and practical significance of the findings. As Long and Freese (2006, *Regression Models for Categorical Dependent Variables Using Stata* [Stata Press]) show, results can often be made more tangible by computing predicted or expected values for hypothetical or prototypical cases. Stata 11 introduced new tools for making such calculations—factor variables and the `margins` command. These can do most of the things that were previously done by Stata's own `adjust` and `mfx` commands, and much more.

Unfortunately, the complexity of the `margins` syntax, the daunting 50-page reference manual entry that describes it, and a lack of understanding about what `margins` offers over older commands that have been widely used for years may have dissuaded some researchers from examining how the `margins` command could benefit them.

In this article, therefore, I explain what adjusted predictions and marginal effects are, and how they can contribute to the interpretation of results. I further explain why older commands, like `adjust` and `mfx`, can often produce incorrect results, and how factor variables and the `margins` command can avoid these errors. The relative merits of different methods for setting representative values for variables in the model (marginal effects at the means, average marginal effects, and marginal effects at representative values) are considered. I shows how the `marginsplot` command (introduced in Stata 12) provides a graphical and often much easier means for presenting and understanding the results from `margins`, and explain why `margins` does not present marginal effects for interaction terms.

Keywords: st0260, margins, marginsplot, adjusted predictions, marginal effects

1 Introduction

Many researchers and journals place a strong emphasis on the sign and statistical significance of effects—but often there is very little emphasis on the substantive and practical significance of the findings. As Long and Freese (2006) show, results can often be made more tangible by computing predicted or expected values for hypothetical or prototyp-

ical cases. For example, if we want to get a practical feel for the impact of gender in a logistic regression model, we might compare the predicted probabilities of success for a man and woman who both have low, average, or high values on the other variables in the model. Such predictions are sometimes referred to as margins, predictive margins, or (Stata's preferred terminology) adjusted predictions. Another useful aid to interpretation is marginal effects, which can succinctly show, for example, how the adjusted predictions for blacks differ from the adjusted predictions for whites.

Stata 11 introduced new tools for making such calculations—factor variables and the `margins` command. These can do most of the things that were previously done by Stata's own `adjust` and `mfx` commands, and much more. Unfortunately, the complexity of the `margins` syntax, the daunting 50-page reference manual entry that describes it, and a lack of understanding about what `margins` offers over older commands that have been widely used for years may have dissuaded some researchers from examining how the `margins` command could benefit them.

In this article, therefore, I illustrate and explain some of the most critical features and advantages of the `margins` command. I explain what adjusted predictions and marginal effects are, and how they can aid interpretation. I show how `margins` can replicate analyses done by older commands like `adjust` but can do so more easily. I demonstrate how, thanks to its support of factor variables that were introduced in Stata 11, `margins` can avoid mistakes made by earlier commands and provide a superior means for dealing with interdependent variables (for example, X and X^2; $X1$, $X2$, and $X1 \times X2$; and multiple dummies created from a single categorical variable). I illustrate the different strategies for defining "typical" cases and how `margins` can estimate them: marginal effects at the means (MEMs), average marginal effects (AMEs), and marginal effects at representative values (MERs); I also show some of the pros and cons of each approach. The output from `margins` can sometimes be overwhelming; I therefore show how the `marginsplot` command, introduced in Stata 12, provides an easy and convenient way of generating graphical results that can be much more understandable. Finally, I explain why, unlike older commands, `margins` does not report marginal effects for interaction terms and why it would be nonsensical to do so.

2 Data

We use `nhanes2f.dta`[1] (Second National Health and Nutrition Examination Survey), available from the StataCorp website. The examples examine how demographic variables are related to whether a person has diabetes.[2] We begin by retrieving the data, extracting the nonmissing cases we want, and then computing variables we will need later.

1. These data were collected in the 1980s. Rates of diabetes in the United States are much higher now.
2. To simplify the discussion and to facilitate our comparison of old and new commands, we do not use the sampling weights that come with the data. However, `margins` can handle those weights correctly.

```
. webuse nhanes2f, clear
. keep if !missing(diabetes, black, female, age, age2, agegrp)
(2 observations deleted)
. label variable age2 "age squared"
. describe diabetes black female age age2 agegrp
              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
diabetes        byte   %9.0g                  diabetes, 1=yes, 0=no
black           byte   %8.0g                  1 if race=black, 0 otherwise
female          byte   %8.0g                  1=female, 0=male
age             byte   %9.0g                  age in years
age2            float  %9.0g                  age squared
agegrp          byte   %8.0g       agegrp     Age groups 1-6
. * Compute the variables we will need
. tab1 agegrp, generate(agegrp)
-> tabulation of agegrp
 Age groups
       1-6 |      Freq.     Percent        Cum.
-----------+-----------------------------------
   age20-29 |      2,320       22.45       22.45
   age30-39 |      1,620       15.67       38.12
   age40-49 |      1,269       12.28       50.40
   age50-59 |      1,289       12.47       62.87
   age60-69 |      2,852       27.60       90.47
    age 70+ |        985        9.53      100.00
-----------+-----------------------------------
      Total |     10,335      100.00
. generate femage = female*age
. label variable femage "female * age interaction"
. summarize diabetes black female age age2 femage, separator(6)
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    diabetes |      10335    .0482825     .214373         0          1
       black |      10335    .1050798    .3066711         0          1
      female |      10335    .5250121    .4993982         0          1
         age |      10335    47.56584    17.21752        20         74
        age2 |      10335    2558.924    1616.804       400       5476
      femage |      10335    25.05031    26.91168         0         74
```

The observations in the sample range in age from 20 to 74, with an average age of 47.57. Slightly over half the sample (52.5%) is female and 10.5% is black.[3] Less than 5% of the respondents have diabetes, but as we will see, the likelihood of having diabetes differs by race, gender, and age. Note that the mean of `femage` (`female` × `age`) is about half the mean of `age`. This reflects the fact that men have a score of 0 on `femage` while for women, `femage` = `age`.

3. Less than two percent of the sample is coded `Other` on race, and their rates of diabetes are identical to whites. We therefore combine whites and others in the analysis.

3 Adjusted predictions for a basic model

We first fit a relatively uncomplicated model.

```
. * Basic model
. logit diabetes black female age, nolog
Logistic regression                                   Number of obs   =      10335
                                                      LR chi2(3)      =     374.17
                                                      Prob > chi2     =     0.0000
Log likelihood = -1811.9828                           Pseudo R2       =     0.0936

------------------------------------------------------------------------------
    diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |   .7179046   .1268061     5.66   0.000     .4693691      .96644
      female |   .1545569   .0942982     1.64   0.101    -.0302642    .3393779
         age |   .0594654   .0037333    15.93   0.000     .0521484    .0667825
       _cons |  -6.405437   .2372224   -27.00   0.000    -6.870384   -5.94049
------------------------------------------------------------------------------
```

According to the model, on an "all other things equal" basis, blacks are more likely to have diabetes than are whites, women are more likely to have diabetes than are men, and the probability of having diabetes increases with age. (The effect of being female is not significant in this model, but it will be significant in other models we test.) The coefficients tell us how the log odds of having diabetes are affected by each variable (for example, the log odds of a black having diabetes are 0.718 greater than the log odds for an otherwise-identical white). But because most people do not think in terms of log odds, many would find it more helpful if they could see how the probability of having diabetes was affected by each variable. For example, the positive and highly significant coefficient for age tells us that getting older is bad for one's health. This is hardly surprising, but just how bad is it? For most people, the coefficient for age of 0.059 has little intuitive or practical appeal.

Adjusted predictions can make these results more tangible. With adjusted predictions, you specify values for each of the independent variables in the model and then compute the probability of the event occurring for an individual who has those values. To illustrate, we will use the `adjust` command to compute the probability that an "average" 20-year-old will have diabetes and compare it to the probability that an "average" 70-year-old will.

```
. adjust age = 20 black female, pr
------------------------------------------------------------------------------
    Dependent variable: diabetes     Equation: diabetes     Command: logit
Covariates set to mean: black = .10507983, female = .52501209
Covariate set to value: age = 20
------------------------------------------------------------------------------

------------------
      All |     pr
----------+-------
          |  .006308
------------------

    Key:  pr  =  Probability
. adjust age = 70 black female, pr
------------------------------------------------------------------------------
    Dependent variable: diabetes     Equation: diabetes     Command: logit
Covariates set to mean: black = .10507983, female = .52501209
Covariate set to value: age = 70
------------------------------------------------------------------------------

------------------
      All |     pr
----------+-------
          |  .110438
------------------

    Key:  pr  =  Probability
```

The results show that an "average" 20-year-old has less than a 1% chance of having diabetes, while an otherwise-comparable 70-year-old has an 11% chance. Most people will find such results much more tangible and meaningful than the original coefficient for age. But what does "average" mean? In this case, we used the common, but not universal, practice of using the mean values for the other independent variables (`female`, `black`) that are in the model; for example, the value of `female` is set to 0.525, while the value for `black` is fixed at 0.105. Later, when discussing marginal effects, I show other options for defining "average".

With `margins`, it is even easier to get these results, and more. We use the `at()` option to fix a variable at a specific value or set of values. The `atmeans` option tells `margins` to fix all other variables at their means. (Unlike `adjust`, this is not the default for `margins`.) If we wanted to see how the probability of having diabetes for average individuals differs across age groups, we could do something like this:[4]

4. The `vsquish` option suppresses blank lines between terms. An even more compact display can be obtained by using the `noatlegend` option, which suppresses the display of the values that variables were fixed at. However, be careful using `noatlegend` because not having that information may make output harder to interpret.

```
. margins, at(age=(20 30 40 50 60 70)) atmeans vsquish
Adjusted predictions                              Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
1._at        : black           =    .1050798 (mean)
               female          =    .5250121 (mean)
               age             =          20
2._at        : black           =    .1050798 (mean)
               female          =    .5250121 (mean)
               age             =          30
3._at        : black           =    .1050798 (mean)
               female          =    .5250121 (mean)
               age             =          40
4._at        : black           =    .1050798 (mean)
               female          =    .5250121 (mean)
               age             =          50
5._at        : black           =    .1050798 (mean)
               female          =    .5250121 (mean)
               age             =          60
6._at        : black           =    .1050798 (mean)
               female          =    .5250121 (mean)
               age             =          70
```

	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
_at						
1	.0063084	.0009888	6.38	0.000	.0043703	.0082465
2	.0113751	.0013794	8.25	0.000	.0086715	.0140786
3	.0204274	.0017892	11.42	0.000	.0169206	.0239342
4	.0364184	.0021437	16.99	0.000	.0322167	.04062
5	.0641081	.0028498	22.50	0.000	.0585226	.0696935
6	.1104379	.005868	18.82	0.000	.0989369	.121939

According to these results, an average 70-year-old (who is again 0.105 `black` and 0.525 `female`) is almost 18 times as likely to have diabetes as an average 20-year-old (11.04% compared with 0.63%). Further, we see that there is a large increase in the predicted probability of diabetes between ages 50 and 60 and an even bigger jump between 60 and 70.

4 Factor variables

Suppose, instead, we wanted to compare the average female with the average male, and the average black with the average nonblack. We could give the commands

```
. margins, at(black = (0 1)) atmeans
. margins, at(female = (0 1)) atmeans
```

Using factor variables, introduced in Stata 11, can make things easier. We need to rerun the `logit` command first.

```
. logit diabetes i.black i.female age, nolog
Logistic regression                             Number of obs   =      10335
                                                LR chi2(3)      =     374.17
                                                Prob > chi2     =     0.0000
Log likelihood = -1811.9828                     Pseudo R2       =     0.0936

------------------------------------------------------------------------------
    diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.black |   .7179046   .1268061     5.66   0.000     .4693691      .96644
    1.female |   .1545569   .0942982     1.64   0.101    -.0302642    .3393779
         age |   .0594654   .0037333    15.93   0.000     .0521484    .0667825
       _cons |  -6.405437   .2372224   -27.00   0.000    -6.870384    -5.94049
------------------------------------------------------------------------------

. margins black female, atmeans
Adjusted predictions                            Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
at           : 0.black         =    .8949202 (mean)
               1.black         =    .1050798 (mean)
               0.female        =    .4749879 (mean)
               1.female        =    .5250121 (mean)
               age             =    47.56584 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |
           0 |   .0294328   .0020089    14.65   0.000     .0254955    .0333702
           1 |   .0585321   .0067984     8.61   0.000     .0452076    .0718566
             |
      female |
           0 |   .0292703   .0024257    12.07   0.000      .024516    .0340245
           1 |   .0339962   .0025912    13.12   0.000     .0289175    .0390748
------------------------------------------------------------------------------
```

The `i.` notation tells Stata that `black` and `female` are categorical variables rather than continuous. As the Stata 12 *User's Guide* (StataCorp 2011) explains in section 11.4.3.1, "`i.group` is called a factor variable, although more correctly, we should say that `group` is a categorical variable to which factor-variable operators have been applied .... When you type `i.group`, it forms the indicators for the unique values of `group`."

In other words, Stata, in effect, creates dummy variables coded 0 or 1 from the categorical variable. In this case, of course, `black` and `female` are already coded 0 or 1—but `margins` and other postestimation commands still like you to use the `i.` notation so they know the variable is categorical (rather than, say, being a continuous variable that just happens to only have the values of 0 or 1 in this sample). But if, say, we had the variable `race` coded 1 = `white` and 2 = `black`, then the new variable would be coded 0 = `white` and 1 = `black`.

Or if the variable `religion` was coded 1 = `Catholic`, 2 = `Protestant`, 3 = `Jewish`, and 4 = `Other`, then saying `i.religion` would cause Stata to create three 0 or 1 dummies. By default, the first category (in this case, `Catholic`) is the reference category, but we can easily change that; for example, `ib2.religion` would make `Protestant` the reference category, or `ib(last).religion` would make the last category, `Other`, the reference.

Factor variables can also be used to include squared terms and interaction terms in models. For example,

```
. logit diabetes i.black i.female age c.age#c.age, nolog
. logit diabetes i.black i.female age i.female#c.age, nolog
```

The `#` (pronounced cross) operator is used for interactions and product terms. The use of `#` implies the `i.` prefix; that is, unless you indicate otherwise, Stata will assume that the variables on both sides of the `#` operator are categorical and will compute interaction terms accordingly. Hence, we use the `c.` notation to override the default and tell Stata that `age` is a continuous variable. So, `c.age#c.age` tells Stata to include $\texttt{age}^2$ in the model; we do not want or need to compute the variable separately. Similarly, `i.female#c.age` produces the `female` $\times$ `age` interaction term. Stata also offers a `##` notation, called factorial cross. It can save some typing and provide an alternative parameterization of the results.

At first glance, the use of factor variables might seem like a minor convenience at best: they save you the trouble of computing dummy variables and interaction terms beforehand. However, the advantages of factor variables become much more apparent when used in conjunction with the `margins` command.

5 Adjusted predictions when there are interdependencies among variables

Sometimes the value of one variable or variables perfectly determines the value of another. For example, if a model includes both X and X^2, then if $X = 10$, X^2 must equal 100. Or if $X1 = 0$, then the interaction $X1 \times X2$ must also equal 0; or if $X1 = 1$, then the interaction $X1 \times X2$ must equal $X2$. If multiple dummies have been created from the same categorical variable (for example, black, white, and other have been created from the variable `race`), then if `black` = 1, the other race dummies must equal 0.

Older Stata commands generally do not recognize such interdependencies between variables. This can lead to incorrect results when computing adjusted predictions. Factor variables and the `margins` command can avoid these errors. Following are some examples.

5.1 Squared terms

Suppose, for example, that our model includes an age^2 term (that is, the `age2` variable), and we want to see what the predicted value is for a 70-year-old who has average (mean) values on the other variables in the model. We can do the following:

```
. * Squared term in model
. logit diabetes black female age age2, nolog
Logistic regression                                   Number of obs   =      10335
                                                      LR chi2(4)      =     381.03
                                                      Prob > chi2     =     0.0000
Log likelihood = -1808.5522                           Pseudo R2       =     0.0953

------------------------------------------------------------------------------
    diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |   .7207406   .1266509     5.69   0.000     .4725093    .9689718
      female |   .1566863   .0942032     1.66   0.096    -.0279486    .3413212
         age |   .1324622   .0291223     4.55   0.000     .0753836    .1895408
        age2 |  -.0007031   .0002753    -2.55   0.011    -.0012428   -.0001635
       _cons |  -8.14958    .7455986   -10.93   0.000    -9.610926   -6.688233
------------------------------------------------------------------------------

. adjust age = 70 black female age2, pr
------------------------------------------------------------------------------
     Dependent variable: diabetes     Equation: diabetes     Command: logit
 Covariates set to mean: black = .10507983, female = .52501209,
                         age2 = 2558.9238
 Covariate set to value: age = 70
------------------------------------------------------------------------------

------------------------------
         All |           pr
-------------+----------------
             |      .373211
------------------------------

     Key:  pr  =  Probability
```

`adjust` yields a predicted probability of 37.3%. That is pretty grim compared with our earlier estimate of 11%! The problem is that `age2` (which has a negative effect) is not being handled correctly. Because the `adjust` command does not know that `age2` is a function of age, it simply uses the mean of `age2`, which is 2558.92. But for a 70-year-old, `age2` $= 4900$.

If we instead use factor variables and the `margins` command, the correct results are easily obtained.

```
. logit diabetes i.black i.female age c.age#c.age, nolog
Logistic regression                             Number of obs   =      10335
                                                LR chi2(4)      =     381.03
                                                Prob > chi2     =     0.0000
Log likelihood = -1808.5522                     Pseudo R2       =     0.0953

------------------------------------------------------------------------------
    diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.black |   .7207406   .1266509     5.69   0.000     .4725093    .9689718
    1.female |   .1566863   .0942032     1.66   0.096    -.0279486    .3413212
         age |   .1324622   .0291223     4.55   0.000     .0753836    .1895408
             |
 c.age#c.age |  -.0007031   .0002753    -2.55   0.011    -.0012428   -.0001635
             |
       _cons |   -8.14958   .7455986   -10.93   0.000    -9.610926   -6.688233
------------------------------------------------------------------------------

. margins, at(age = 70) atmeans
Adjusted predictions                            Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
at           : 0.black         =    .8949202 (mean)
               1.black         =    .1050798 (mean)
               0.female        =    .4749879 (mean)
               1.female        =    .5250121 (mean)
               age             =          70

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .1029814   .0063178    16.30   0.000     .0905988     .115364
------------------------------------------------------------------------------
```

By using factor-variable notation, we let the `margins` command know that if age=70, then $\text{age}^2 = 4900$, and it hence computes the predicted values correctly.

5.2 Interaction terms

Now suppose we have an interaction term in our model, for example, `female` × `age` (`femage`). We want to compute the predicted probability of diabetes for a male who has average values on the other variables. We might do something like this:

```
. * Interaction term
. logit diabetes black female age femage, nolog
Logistic regression                                 Number of obs   =      10335
                                                    LR chi2(4)      =     380.85
                                                    Prob > chi2     =     0.0000
Log likelihood = -1808.6405                         Pseudo R2       =     0.0953

------------------------------------------------------------------------------
    diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |   .7112782   .1268575     5.61   0.000     .4626421    .9599144
      female |   1.358331   .4851999     2.80   0.005     .4073562    2.309305
         age |   .0715351   .0063037    11.35   0.000       .05918    .0838902
      femage |  -.0199143   .0078292    -2.54   0.011    -.0352593   -.0045693
       _cons |  -7.140004   .3961599   -18.02   0.000    -7.916463   -6.363545
------------------------------------------------------------------------------

. adjust female = 0 black age femage, pr
------------------------------------------------------------------------------
    Dependent variable: diabetes     Equation: diabetes     Command: logit
Covariates set to mean: black = .10507983, age = 47.565844, femage = 25.050314
Covariate set to value: female = 0
------------------------------------------------------------------------------

------------------------
         All |         pr
-------------+----------
             |    .015345
------------------------

    Key:  pr  =  Probability
```

Note that the `adjust` command is using `femage` = 25.05, which, as we saw earlier, is the mean value of `femage` in the sample. But that is obviously wrong: if `female` = 0, then `femage` also = 0.

Now let's use factor variables and the `margins` command:

```
. logit diabetes i.black i.female age i.female#c.age, nolog
Logistic regression                             Number of obs   =      10335
                                                LR chi2(4)      =     380.85
                                                Prob > chi2     =     0.0000
Log likelihood = -1808.6405                     Pseudo R2       =     0.0953

------------------------------------------------------------------------------
    diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.black |   .7112782   .1268575     5.61   0.000     .4626421    .9599144
    1.female |   1.358331   .4851999     2.80   0.005     .4073562    2.309305
         age |   .0715351   .0063037    11.35   0.000       .05918    .0838902
             |
female#c.age |
          1  |  -.0199143   .0078292    -2.54   0.011    -.0352593   -.0045693
             |
       _cons |  -7.140004   .3961599   -18.02   0.000    -7.916463   -6.363545
------------------------------------------------------------------------------

. margins female black female#black, atmeans
Adjusted predictions                            Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
at           : 0.black         =    .8949202 (mean)
               1.black         =    .1050798 (mean)
               0.female        =    .4749879 (mean)
               1.female        =    .5250121 (mean)
               age             =    47.56584 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |
          0  |   .0250225   .0027872     8.98   0.000     .0195597    .0304854
          1  |   .0372713   .0029632    12.58   0.000     .0314635    .0430791
             |
       black |
          0  |   .0287052   .0020278    14.16   0.000     .0247307    .0326797
          1  |   .0567715   .0067009     8.47   0.000     .0436379    .0699051
             |
female#black |
        0 0  |   .0232624   .0026348     8.83   0.000     .0180983    .0284265
        0 1  |   .0462606   .0068486     6.75   0.000     .0328376    .0596835
        1 0  |   .0346803   .0028544    12.15   0.000     .0290857    .0402748
        1 1  |   .0681786   .0083774     8.14   0.000     .0517592     .084598
------------------------------------------------------------------------------
```

This tells us that the average male (who is 0.105 `black` and 47.57 years old) has a predicted 2.5% chance of having diabetes. The average female (also 0.105 `black` and 47.57 years old) has a 3.7% chance. If we fail to take into account the fact that `femage` is a function of age, we underestimate the likelihood that men will have diabetes; that is, if we do it wrong, we estimate that the average male has a 1.5% probability of having diabetes when the correct estimate is 2.5%.

We also asked for information pertaining to race. This shows that the average white (who is 0.525 `female` and 47.57 years old) has a 2.9% chance of having diabetes, while for the average black the figure is nearly twice as high at 5.7%. The `female#black` notation on the `margins` command does not mean that an interaction term for race and gender has been added to the model. Rather, it simply causes the adjusted predictions for each combination of race and gender (based on the model that was fit) to be included in the output.

5.3 Multiple dummies

One other sort of interdependency not handled well by older commands is when multiple dummies are computed from a single categorical variable. For example, suppose we do not have the continuous variable `age` and instead have to use the categorical `agegrp` variables. We want to estimate the probability that the average person aged 70 or above has diabetes:

```
. * Multiple dummies
. logit diabetes black female agegrp2 agegrp3 agegrp4 agegrp5 agegrp6, nolog
Logistic regression                             Number of obs   =      10335
                                                LR chi2(7)      =     368.98
                                                Prob > chi2     =     0.0000
Log likelihood =  -1814.575                     Pseudo R2       =     0.0923

------------------------------------------------------------------------------
    diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |   .7250941   .1265946     5.73   0.000     .4769733    .9732148
      female |   .1578264   .0941559     1.68   0.094    -.0267158    .3423686
     agegrp2 |   .7139572   .3397881     2.10   0.036     .0479847     1.37993
     agegrp3 |   1.685402   .3031107     5.56   0.000     1.091316    2.279488
     agegrp4 |   2.223236   .2862673     7.77   0.000     1.662162    2.784309
     agegrp5 |   2.674737   .2680303     9.98   0.000     2.149407    3.200066
     agegrp6 |   2.999892   .2783041    10.78   0.000     2.454426    3.545358
       _cons |  -5.242579   .2658865   -19.72   0.000    -5.763707   -4.721451
------------------------------------------------------------------------------

. adjust agegrp6 = 1 black female agegrp2 agegrp3 agegrp4 agegrp5, pr
------------------------------------------------------------------------------
    Dependent variable: diabetes     Equation: diabetes     Command: logit
 Covariates set to mean: black = .10507983, female = .52501209,
                         agegrp2 = .15674891, agegrp3 = .12278665,
                         agegrp4 = .12472182, agegrp5 = .27595549
 Covariate set to value: agegrp6 = 1
------------------------------------------------------------------------------

------------------------
         All |         pr
-------------+----------
             |   .320956
------------------------

    Key:  pr  =  Probability
```

According to the `adjust` command, the average person (average meaning 0.105 `black` and 0.525 `female`) who is age 70 or above has a 32.1% chance of having diabetes—far higher than our earlier estimates of around 10 or 11%. Being old may be bad for your health, but it is not that bad! As the `logit` results show, each older age group is more likely to have diabetes than the youngest age group. But each person only belongs to one age group; that is, if you have a score of 1 on `agegrp6`, you have to have a score of 0 on all the other age-group variables. `adjust`, on the other hand, is using the mean values for all the other age dummies (rather than 0), which causes the probability of having diabetes for somebody aged 70 or above to be greatly overestimated.

Factor variables and `margins` again provide an easy means of doing things correctly.

```
. logit diabetes i.black i.female i.agegrp, nolog
Logistic regression                             Number of obs   =      10335
                                                LR chi2(7)      =     368.98
                                                Prob > chi2     =     0.0000
Log likelihood =  -1814.575                     Pseudo R2       =     0.0923

------------------------------------------------------------------------------
    diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.black |   .7250941   .1265946     5.73   0.000     .4769733    .9732148
    1.female |   .1578264   .0941559     1.68   0.094    -.0267158    .3423686
             |
      agegrp |
          2  |   .7139572   .3397881     2.10   0.036     .0479847     1.37993
          3  |   1.685402   .3031107     5.56   0.000     1.091316    2.279488
          4  |   2.223236   .2862673     7.77   0.000     1.662162    2.784309
          5  |   2.674737   .2680303     9.98   0.000     2.149407    3.200066
          6  |   2.999892   .2783041    10.78   0.000     2.454426    3.545358
             |
       _cons |  -5.242579   .2658865   -19.72   0.000    -5.763707   -4.721451
------------------------------------------------------------------------------
```

```
. margins female black agegrp, atmeans
Adjusted predictions                              Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
at           : 0.black         =    .8949202 (mean)
               1.black         =    .1050798 (mean)
               0.female        =    .4749879 (mean)
               1.female        =    .5250121 (mean)
               1.agegrp        =    .2244799 (mean)
               2.agegrp        =    .1567489 (mean)
               3.agegrp        =    .1227866 (mean)
               4.agegrp        =    .1247218 (mean)
               5.agegrp        =    .2759555 (mean)
               6.agegrp        =    .0953072 (mean)
```

	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
female						
0	.0280253	.0025121	11.16	0.000	.0231016	.0329489
1	.03266	.0027212	12.00	0.000	.0273266	.0379935
black						
0	.0282075	.0021515	13.11	0.000	.0239906	.0324244
1	.0565477	.006821	8.29	0.000	.0431787	.0699166
agegrp						
1	.0061598	.0015891	3.88	0.000	.0030453	.0092744
2	.0124985	.002717	4.60	0.000	.0071733	.0178238
3	.0323541	.0049292	6.56	0.000	.0226932	.0420151
4	.0541518	.0062521	8.66	0.000	.041898	.0664056
5	.082505	.0051629	15.98	0.000	.0723859	.092624
6	.1106978	.009985	11.09	0.000	.0911276	.130268

Similarly to our earlier results, the probability of having diabetes is much greater for an otherwise-average person aged 70 or above than it is for a similar person in his or her 20s.

We also got the predicted values for average females, males, blacks, and whites. While these numbers are similar to before, the average person is no longer 47.57 years old. Rather, the average person now has a score of 0.224 on `agegrp1`, 0.157 on `agegrp2`, and so on.

To sum up, for many purposes both older and newer, Stata commands like `adjust` and `margins` will work well, but `margins` is usually easier to use and more flexible. When variables are interdependent, for example, when the value of one or more variables completely determines the value of another, the `margins` command is clearly superior. You can try to include options with older commands to take into account the interdependencies, but it is generally easier (and probably less error-prone) if you use the new `margins` command instead.

6 Marginal effects

Marginal effects are another popular means by which the effects of variables in nonlinear models can be made more intuitively meaningful. As Cameron and Trivedi (2010, 343) note, "A marginal effect (ME), or partial effect, most often measures the effect on the conditional mean of y of a change in one of the regressors, say, x_j. In the linear regression model, the ME equals the relevant slope coefficient, greatly simplifying analysis. For nonlinear models, this is no longer the case, leading to remarkably many different methods for calculating MEs."

Marginal effects for categorical independent variables are especially easy to understand.[5] The ME for categorical variables shows how $\mathrm{P}(Y = 1)$ changes as the categorical variable changes from 0 to 1, after controlling in some way for the other variables in the model.With a dichotomous independent variable, the ME is the difference in the adjusted predictions for the two groups, for example, for blacks and whites.

There are different ways of controlling for the other variables in the model. Older Stata commands (for example, `adjust` and `mfx`) generally default to using the means for variables whose values have not been otherwise specified, that is, they estimate marginal effects at the means (MEMs). Presumably, the mean reflects the "average" or "typical" person on the variable. However, at least two other approaches are also possible with the `margins` command: average marginal effects (AMEs) and marginal effects at representative values (MERs). We now illustrate each of these approaches, with each building off of the following basic model.

```
. * Back to basic model
. logit diabetes i.black i.female age, nolog
Logistic regression                             Number of obs   =      10335
                                                LR chi2(3)      =     374.17
                                                Prob > chi2     =     0.0000
Log likelihood = -1811.9828                     Pseudo R2       =     0.0936

------------------------------------------------------------------------------
    diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.black |   .7179046   .1268061     5.66   0.000     .4693691      .96644
   1.female |   .1545569   .0942982     1.64   0.101    -.0302642     .3393779
         age |   .0594654   .0037333    15.93   0.000     .0521484     .0667825
       _cons |  -6.405437   .2372224   -27.00   0.000    -6.870384    -5.94049
------------------------------------------------------------------------------
```

6.1 MEMs

MEMs are easily estimated with the `margins` command. The `dydx()` option tells `margins` which variables to compute MEs for. The `atmeans` option tells `margins` to use the mean values for other variables when computing the ME for a variable. For the same reasons as given before, it is important to use factor-variable notation so that Stata recognizes any interdependencies between variables. It is also important because MEs are computed differently for discrete and continuous independent variables.

5. See Cameron and Trivedi (2010) for a discussion of marginal effects for continuous variables.

```
. * MEMs - Marginal Effects at the Means
. margins, dydx(black female) atmeans
Conditional marginal effects                      Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
dy/dx w.r.t. : 1.black 1.female
at           : 0.black         =    .8949202 (mean)
               1.black         =    .1050798 (mean)
               0.female        =    .4749879 (mean)
               1.female        =    .5250121 (mean)
               age             =    47.56584 (mean)
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.black	.0290993	.0066198	4.40	0.000	.0161246	.0420739
1.female	.0047259	.0028785	1.64	0.101	-.0009158	.0103677

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

The results tell us that if you had two otherwise-average individuals, one white, one black, the black's probability of having diabetes would be 2.9 percentage points higher. And what do we mean by "average"? With MEMs, average is defined as having the mean value for the other independent variables in the model, that is, 47.57 years old, 10.5% black, and 52.5% female.

MEMs are easy to explain. With the `atmeans` option, we fix some variable values (for example, `black` = 1), compute the mean values for the other variables, and then use the fixed and mean values to compute predicted probabilities. The predicted values show us how the average female compares with the average male, where average is defined as having mean values on the other variables in the model.

MEMs have been widely used. Indeed, for a long time, MEMs were the only option with Stata, because that is all the old `mfx` command supported. But many do not like MEMs. While there are people who are 47.57 years old, there is nobody who is 10.5% black or 52.5% female. Further, the means are only one of many possible sets of values that could be used—and a set of values that no real person could actually have seems troublesome. For these and other reasons, many researchers prefer AMEs, which I describe next.

6.2 AMEs

Rather than use the means when computing predicted values, some argue it is best to use the actual observed values for the variables whose values are not otherwise fixed (which is the default `asobserved` option for the `margins` command). With `atmeans`, we fix the values of some variables (for example, `black` = 1) and then use the means for the other variables to compute predicted probabilities. With `asobserved`, we again fix the values for some variables, but for the other variables we use the observed values for each case. We then compute a predicted probability for each case with the fixed and observed values of variables, and then we average the predicted values.

```
. * AMEs - Average Marginal Effects
. margins, dydx(black female)
Average marginal effects                        Number of obs   =      10335
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
dy/dx w.r.t. : 1.black 1.female
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.black	.0400922	.0087055	4.61	0.000	.0230297	.0571547
1.female	.0067987	.0041282	1.65	0.100	-.0012924	.0148898

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

Intuitively, the AME for being black is computed as follows:

- Go to the first case. Treat that person as though he or she were white, regardless of what the person's race actually is. Leave all other independent variable values as is. Compute the probability that this person (if he or she were white) would have diabetes.
- Now do the same thing but this time treating the person as though he or she were black.
- The difference in the two probabilities just computed is the ME for that case.
- Repeat the process for every case in the sample.
- Compute the average of all the MEs you have computed. This gives you the AME for being black.

If the `margins` command did not exist, it would be fairly straightforward to do the same computations using other Stata commands. Indeed, doing so can yield additional insights of interest.

```
. * Replicate AME for black without using margins
. clonevar xblack = black
. quietly logit diabetes i.xblack i.female age, nolog
. replace xblack = 0
(1086 real changes made)
. predict adjpredwhite
(option pr assumed; Pr(diabetes))
. replace xblack = 1
(10335 real changes made)
. predict adjpredblack
(option pr assumed; Pr(diabetes))
. generate meblack = adjpredblack - adjpredwhite
```

```
. summarize adjpredwhite adjpredblack meblack
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
adjpredwhite |      10335    .0443248    .0362422    .005399   .1358214
adjpredblack |      10335     .084417    .0663927   .0110063   .2436938
     meblack |      10335    .0400922    .0301892   .0056073   .1078724
```

With AMEs, you are in effect comparing two hypothetical populations—one all white, one all black—that have the exact same values on the other independent variables in the model. The logic is similar to that of a matching study, where subjects have identical values on every independent variable except one. Because the only difference between these two populations is their races, race must be the cause of the difference in their probabilities of having diabetes.

Many people like the fact that all the data are being used, not just the means, and feel that this leads to superior estimates. Many, perhaps most, authors seem to prefer AMEs over MEMs (for example, Bartus [2005] and Cameron and Trivedi [2010]). Others, however, are not convinced that treating men as though they are women and women as though they are men really is a better way of computing MEs.

The biggest problem with both of the last two approaches, however, may be that they only produce a single estimate of the ME. No matter how "average" is defined, averages can obscure differences in effects across cases. In reality, the effect that variables like race have on the probability of success varies with the characteristics of the person; for example, racial differences could be much greater for older people than for younger. Indeed, in the example above, the summary statistics showed that while the AME for being black was 0.04, the ME for individual cases ranged between 0.006 and 0.108; that is, at the individual level, the largest ME for being black was almost 20 times as large as the smallest.

For these and other reasons, MERs will often be preferable to either of the alternatives already discussed.

6.3 MERs

With MERs, you choose ranges of values for one or more independent variables and then see how the MEs differ across that range. MERs can be intuitively meaningful, while showing how the effects of variables vary by other characteristics of the individual. The use of the `at()` option makes this possible.

```
. * Section 6.3: MERs - Marginal Effects at Representative Values
. quietly logit diabetes i.black i.female age, nolog
. margins, dydx(black female) at(age=(20 30 40 50 60 70)) vsquish
Average marginal effects                        Number of obs   =      10335
Model VCE    : OIM
Expression   : Pr(diabetes), predict()
dy/dx w.r.t. : 1.black 1.female
1._at        : age             =          20
2._at        : age             =          30
3._at        : age             =          40
4._at        : age             =          50
5._at        : age             =          60
6._at        : age             =          70
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.black						
_at						
1	.0060899	.0016303	3.74	0.000	.0028946	.0092852
2	.0108784	.0027129	4.01	0.000	.0055612	.0161956
3	.0192101	.0045185	4.25	0.000	.0103541	.0280662
4	.0332459	.0074944	4.44	0.000	.018557	.0479347
5	.0555816	.0121843	4.56	0.000	.0317008	.0794625
6	.0877803	.0187859	4.67	0.000	.0509606	.1245999
1.female						
_at						
1	.0009933	.0006215	1.60	0.110	-.0002248	.0022114
2	.00178	.0010993	1.62	0.105	-.0003746	.0039345
3	.003161	.0019339	1.63	0.102	-.0006294	.0069514
4	.0055253	.0033615	1.64	0.100	-.001063	.0121137
5	.0093981	.0057063	1.65	0.100	-.001786	.0205821
6	.0152754	.0092827	1.65	0.100	-.0029184	.0334692

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

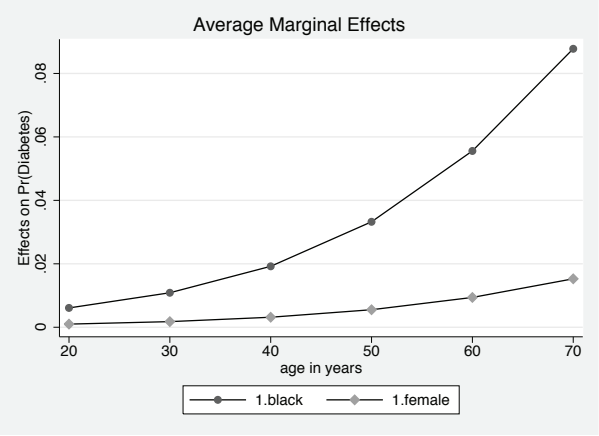
Earlier, the AME for being black was 4%; that is, on average blacks' probability of having diabetes is four percentage points higher than it is for whites. But when we estimate MEs for different ages, we see that the effect of being black differs greatly by age. It is less than one percentage point for 20-year-olds and almost nine percentage points for those aged 70. This makes sense, because the probability of diabetes differs greatly by age; it would be unreasonable to expect every white to be four percentage points less likely to get diabetes than every black regardless of age. Similarly, while the AME for gender was only 0.6%, at different ages the effect is much smaller or much higher than that.

In a large model, it may be cumbersome to specify representative values for every variable, but you can do so for those of greatest interest. The `atmeans` or `asobserved` options can then be used to set the values of the other variables in the model.

7 Graphic displays of margins results: The marginsplot command

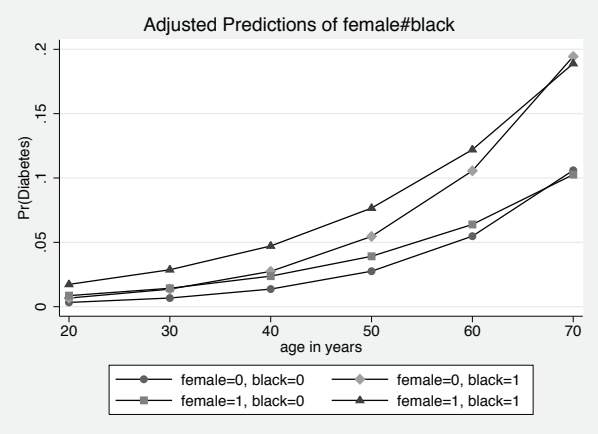
The output from the `margins` command can be very difficult to read. Because of space constraints, numbers are used to label categories rather than value labels. The `marginsplot` command introduced in Stata 12 makes it easy to create a visual display of results. Here are two simple examples:

```
. quietly logit diabetes i.black i.female age, nolog
. quietly margins, dydx(black female) at(age=(20 30 40 50 60 70)) vsquish
. marginsplot, noci
```



The graph makes it clear how the differences between blacks and whites, and between men and women, increase with age. Here is a slightly more complicated example that illustrates how `marginsplot` can also be used with adjusted predictions.

```
. * Plot of adjusted predictions
. quietly logit diabetes i.black i.female age i.female#c.age, nolog
. quietly margins female#black, at(age=(20 30 40 50 60 70))
. marginsplot, noci
```



The differences between blacks and whites and men and women are again clear. This model also included an interaction term for gender and age. The graphic shows that up until about age 70, women are more likely to get diabetes than their same-race male counterparts, but after that men are slightly more likely.

8 Marginal effects for interaction terms

People often ask what the ME of an interaction term is. Stata's `margins` command replies: there is not one. You just have the MEs of the component terms. The value of the interaction term cannot change independently of the values of the component terms, so you cannot estimate a separate effect for the interaction. The older `mfx` command will report MEs for interaction terms, but the numbers it gives are wrong because `mfx` is not aware of the interdependencies between the interaction term itself and the variables used to compute the interaction term.

```
. quietly logit diabetes i.black i.female age i.female#c.age, nolog
. margins, dydx(*)
Average marginal effects                          Number of obs   =      10335
Model VCE    : OIM
Expression   : Pr(diabetes), predict()
dy/dx w.r.t. : 1.black 1.female age
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.black	.0396176	.0086693	4.57	0.000	.022626	.0566092
1.female	.0067791	.0041302	1.64	0.101	-.001316	.0148743
age	.0026632	.0001904	13.99	0.000	.0022901	.0030364

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

9 Other points

`margins` would also give the wrong answers if you did not use factor variables. You should use `margins` because older commands, like `adjust` and `mfx`, do not support the use of factor variables. `margins` supports the use of the `svy:` prefix with `svyset` data. Some older commands do not. `margins` is, unfortunately, more difficult to use with multiple-outcome commands like `ologit` or `mlogit`. You have to specify a different `margins` command for each possible outcome of the dependent variable. But this is also true of many older commands. It is my hope that future versions of `margins` will overcome this limitation. The ability to compute adjusted predictions and MEs for individual cases would also be a welcome addition to `margins`. Finally, both `margins` and `marginsplot` include numerous other options that can be used to further refine the analysis and the presentation of results.

10 Conclusion

Adjusted predictions and marginal effects can make the results from many analyses much more intuitive and easier to interpret. The `margins` command offers a generally superior alternative to the `adjust` and `mfx` commands that preceded it. It can estimate the same models and can generally do so more easily. Interdependencies between variables are easily handled, and the user has a choice between the `atmeans` and `asobserved` options.

The relative merits of `atmeans` versus `asobserved` continue to be debated. Clearly, many prefer the `asobserved` approach. They would rather compare hypothetical populations that have values that real people actually do have than compare hypothetical persons with mean values on variables that no real person could ever have. But however "typical" or "average" is defined, any approach that only looks at "typical" values is going to miss variability in effects across cases. Presenting MERs can make results easier to interpret and provide a better feel for how the effects of variables differ across cases.

11 Acknowledgment

Sarah Mustillo offered numerous helpful comments on earlier versions of this article.

12 References

Bartus, T. 2005. Estimation of marginal effects using margeff. *Stata Journal* 5: 309–329.

Cameron, A. C., and P. K. Trivedi. 2010. *Microeconometrics Using Stata*. Rev. ed. College Station, TX: Stata Press.

Long, J. S., and J. Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd ed. College Station, TX: Stata Press.

StataCorp. 2011. *Stata 12 User's Guide*. College Station, TX: Stata Press.

About the author

Richard Williams is an associate professor and a former chairman of the Department of Sociology at the University of Notre Dame. His teaching and research interests include Methods and Statistics, Demography, and Urban Sociology. His work has appeared in the *American Sociological Review*, *Social Forces*, *Stata Journal*, *Social Problems*, *Demography*, *Sociology of Education*, *Journal of Urban Affairs*, *Cityscape*, *Journal of Marriage and the Family*, and *Sociological Methods and Research*. His recent research, which has been funded by grants from the Department of Housing and Urban Development and the National Science Foundation, focuses on the causes and consequences of inequality in American home ownership. He is a frequent contributor to Statalist.