

## AGRODEP Stata Training

April 2013

### Module 4

## Bivariate Regressions

Manuel Barron<sup>1</sup> and Pia Basurto<sup>2</sup>

<sup>1</sup> University of California, Berkeley, Department of Agricultural and Resource Economics

<sup>2</sup> University of California, Santa Cruz, Department of Economics

AGRODEP Stata Training documents are designed to give AGRODEP members a brief overview of basic Stata commands needed in AGRODEP training courses. These documents have been reviewed but have not been subject to a formal external peer review via IFPRI's Publications Review Committee; any opinions expressed are those of the author(s) and do not necessarily reflect the opinions of AGRODEP or of IFPRI.

## Module 4 – Bivariate Regressions

This module will introduce the commands required to run bivariate regressions, with particular emphasis on *probit* and *logit*. Since these are non-linear models, it is important to calculate the marginal effects adequately, which we will do through the *mfxf* command. We will end the module with an illustration of how to export the results with *outreg*.

For this module we will use [hhmembers\\_2.dta, available in the AGRODEP website](#).

### 1. probit

The *probit* command will run a probit regression. The syntax is similar to *regress*. First you type the command name, then the left-hand-side variable followed by the right-hand-side variables. You may use *if*, *in* to constrain the estimation to a subset of the sample, as well as *weights* and other advanced options that will not be covered here.

#### \* Do-file or Command Window

```
help probit
```

#### \*Help File

```
probit depvar [indepvars] [if] [in] [weight] [, options]
```

#### \*Do-file or command window

```
probit family_work sex age
```

#### \*Stata output

```
Iteration 0:  log likelihood = -11473.134
Iteration 1:  log likelihood = -10810.857
Iteration 2:  log likelihood = -10805.545
Iteration 3:  log likelihood = -10805.544
Iteration 4:  log likelihood = -10805.544
```

```
Probit regression                Number of obs   =       23127
                                LR chi2(2)         =       1335.18
                                Prob > chi2          =         0.0000
Log likelihood = -10805.544      Pseudo R2       =         0.0582
```

```
-----+-----
family_work |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      sex |   .3913636   .0196327    19.93  0.000   .3528842   .4298431
      age |   .104986   .0035399    29.66  0.000   .0980479   .1119241
      _cons | -2.078091   .0378471   -54.91  0.000  -2.15227  -2.003913
-----+-----
```

To calculate the marginal effects from your probit regression, type *mfx* immediately after you ran the probit regression. The *mfx* command uses the stored output that Stata saves in its temporary memory (for more information on how Stata saves the results in memory and how to access them, type “help return”). If you are familiar with probit regressions you will know that the marginal effects are not constant. Stata calculates the marginal effects at the average values of the explanatory variables. You may change this with the *at()* option. This is an advanced feature (see *help mfx* for details, especially the *at(atlist)* section).

**\*Do-file or command window**

```
mfx
```

**\*Stata Output**

```
Marginal effects after probit
      y = Pr(family_work) (predict)
      = .17270865
-----
variable |      dy/dx   Std. Err.    z    P>|z|    [   95% C.I.   ]      X
-----+-----
      sex*|   .1046784   .00502   20.84   0.000   .094835   .114522   .511213
      age |   .0294173   .00091   32.29   0.000   .027632   .031203   9.27055
-----
(*) dy/dx is for discrete change of dummy variable from 0 to 1
.
```

**2. Logit**

To run a logit regression, use the *logit* command. The syntax is similar to that of *regress* and *probit*. First you type the command name, then the left-hand-side variable followed by the right-hand-side variables. Again, you may use *if*, *in*, and *weights*, and some advanced options that will not be covered in these notes.

**\* Do-file or Command Window**

```
help logit
```

**\*Help File**

```
logit depvar [indepvars] [if] [in] [weight] [, options]
```

**\*Do-file or command window**

```
logit family_work sex age
```

### \*Stata output

```
Iteration 0: log likelihood = -11132.912
Iteration 1: log likelihood = -10420.177
Iteration 2: log likelihood = -10392.673
Iteration 3: log likelihood = -10392.608
Iteration 4: log likelihood = -10392.608
```

```
Logistic regression                               Number of obs =      22920
                                                  LR chi2(2)      =      1480.61
                                                  Prob > chi2     =          0.0000
Log likelihood = -10392.608                    Pseudo R2      =          0.0665
```

family_work	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
sex	.7369376	.0359242	20.51	0.000	.6665274	.8073478
age	.2004067	.0064456	31.09	0.000	.1877736	.2130399
_cons	-3.823348	.0721334	-53.00	0.000	-3.964727	-3.681969

```
.
end of do-file
```

As in the case of probit, you may use the *mfx* to obtain the marginal effects.

### \*Do-file or command window

```
mfx
```

### \*Stata output

```
Marginal effects after logit
y = Pr(family_work) (predict)
= .1695619
```

variable	dy/dx	Std. Err.	z	P> z	[ 95% C.I. ]		X
sex*	.1035623	.00494	20.97	0.000	.093883	.113242	.511213
age	.0282194	.00086	32.72	0.000	.026529	.02991	9.27055

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

To check the accuracy in the predictive power of your model, type: *estat classification*

**\*Do-file or command window**

```
estat classification
```

**\*Stata output**

Logistic model for family\_work

Classified	True		Total
	D	~D	
+	0	0	0
-	4347	18573	22920
Total	4347	18573	22920

Classified + if predicted Pr(D) >= .5  
True D defined as family\_work != 0

Sensitivity	Pr( +  D)	0.00%
Specificity	Pr( - ~D)	100.00%
Positive predictive value	Pr( D  +)	.%
Negative predictive value	Pr(~D  -)	81.03%
False + rate for true ~D	Pr( + ~D)	0.00%
False - rate for true D	Pr( -  D)	100.00%
False + rate for classified +	Pr(~D  +)	.%
False - rate for classified -	Pr( D  -)	18.97%
Correctly classified		81.03%

.

**3. outreg**

To store your results in a Word file use outreg as in the previous module.

**\*Do-file or command window**

```
probit family_work sex age
margeff,replace
outreg using reg_module4,replace se ctitle("Probit") title("Family work")

logit family_work sex age
margeff,replace
outreg using reg_module4,append se ctitle("Logit")
```

Your Word file will look like this:

Bivariate Regressions		
	(1)	(2)
	Probit	Logit
Sex	0.076 (0.004)**	0.077 (0.004)**
Age	0.021 (0.000)**	0.021 (0.000)**
Observations	22920	22920

Standard errors in parentheses  
\* significant at 5%; \*\* significant at 1%

#### 4. Wrapping Up

This module presented *probit* and *logit*, the two most commonly used commands for bivariate regressions. We introduced the *mf* command to calculate the marginal effects, and we finished the module showing how to export the estimation results with *outreg*.