

User's guide to `ecol` for Stata

Frederik Siegumfeldt

December 31, 2004

Contents

1	Setting up <code>ecol</code>	1
1.1	Updating Stata	1
1.2	Installing the files using <code>net install</code>	2
1.3	Installing the files from <code>ecol.zip</code>	3
1.4	Checking the installation	3
2	Using <code>ecol</code>	4
2.1	The required data	4
2.2	Missing values	4
2.3	The <code>ecol</code> dialog	5
2.4	Interpreting the results	7
3	Syntax and settings	8
3.1	<code>Ecol</code> command syntax	8
3.2	<code>Ecol</code> settings	9

Introduction

Ecol estimates an $n \times m$ individual level table from a Stata file with aggregate data at the district level. The table shows the association between two individual level variables X and Y , where X has m categories and Y has n categories. The data needed is a Stata file with districts as units and the number of individuals in each category as variables.

For example, a table showing the flow of votes between the parties from one election to the next is estimated from election results aggregated by districts. Another example is the estimation of class voting in a table with classes as rows and parties as columns.

1 Setting up ecol

Ecol will only work under Stata 8.1 or newer, run on a fairly recent version of Windows (i.e. not Windows 98 or older). Installation is carried out in one of two ways: You can access the Stata-website of the Department of Political Science at Aarhus University using the net functions of Stata, or you can download the files in a zip-archive and put them in the proper locations yourself. Section 1.1 tells you how to update your version of Stata (if necessary), sections 1.2 and 1.3 describes both ways of installing **ecol**, and finally section 1.4 shows how to correct the few errors that may occur during installation.

1.1 Updating Stata

Ecol uses a recent feature of Stata called ‘plugins’, which was introduced in Stata 8.1, so before you install **ecol**, you will need to make sure that you are running Stata 8.1 or newer. Stata displays the version number on start-up (as shown in Fig. 1) or in response to the version command.¹ If you are already using version 8.1 or newer, no updates are necessary and you can proceed with the installation. Stata 8.0 can be updated to the latest version of Stata 8 by using the update all command. Updating the entire Stata package can be quite a lengthy process, so you may want to skip updating your command library by using update executable rather than update all. If you are using Stata 7 or older, you will not be able to use **ecol**.

¹Note that the output shown here may not resemble your Stata output exactly, since this guide uses 10pt Lucida Console with the ‘White background’ color scheme rather than the default ‘Stata 8×12’ on a black background.

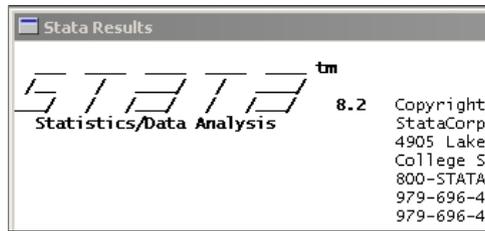


Figure 1: Stata version number

1.2 Installing the files using net install

The simplest way of installing **ecol** is to download it directly from the Stata-website of the Department of Political Science at the University of Aarhus. Run

```
net install http://www.ps.au.dk/stata/ecol_dir/ecol
```

in Stata to connect to the server and download the files. If you already have **ecol** installed, you can reinstall it by using

```
net install http://www.ps.au.dk/stata/ecol_dir/ecol, replace
```

If installation proceeds without Stata issuing any error messages, you can skip ahead to section 1.4.

This way of installing the files may, however, cause problems if you do not have permission to write to the hard drive of the machine you are working on. This may cause Stata (or Windows) to issue an error message along the lines of 'unable to write to the specified location' when you try to download the files. There are two ways to proceed if you lack the required permissions: You can consult your network administrator (or someone else with administrator privileges on your workstation), and they may be able to install the files to the appropriate location or grant you the required permissions to do so yourself.

Alternatively, you can install the files to some other location (e.g. a personal network drive) and make Stata look for them there. Say you have access to a network drive called N: and were trying to install the files with `net install` when things went wrong: The first thing to do is to create a directory on N: for Stata to install the files in, e.g. `N:\ado\plus`. Then run `net set ado N:\ado\plus` to make Stata download the files to the new directory. Finally, direct Stata's attention to the fact that add-ons are now present in this directory by running `sysdir set PLUS N:\ado\plus`. Note that this last command will have to be issued in *every* subsequent

Stata session where you want to use **ecol**, since the settings that determine where Stata looks for add-ons are reset between sessions.²

1.3 Installing the files from **ecol.zip**

Ecol can also be installed ‘manually’, by downloading the file **ecol.zip** from the website at www.ps.au.dk/stata/ecol.dir and putting the files contained in this archive in the appropriate locations. This zip-archive contains the core files of **ecol** (**ecol.ado**, **ecol.dlg**, **ecol.hlp**, and **ecol.s.plugin**) as well as the latest version of this manual and a sample data set. The core files should be placed in `C:\ado\personal\e`. If Windows does not object to this, skip ahead to section 1.4.

If Windows refuses to move the files to this location, you have two alternatives: One is to ask your network administrator or someone with administrator privileges on your workstation to install the files or grant you the required permissions to do so yourself. The other alternative is to install the files somewhere else (e.g. a personal network drive where you *do* have the required permissions) and then run `sysdir set PERSONAL [location]` in Stata, where *[location]* is the name of the directory where you moved the files. This command will have to be issued in every subsequent Stata session where you want to use **ecol**, since the settings that determine where Stata looks for add-ons are reset between sessions.³

1.4 Checking the installation

Having installed the files, you can run `ecol version` and `ecol` (with no parameters) in Stata as a simple way of confirming that all the components are present and working. Stata should respond to `ecol version` like this:

```
ecol: First run, menu item set up
      (ECOL for Stata v0.0)
```

The message in the first line will be displayed whenever you run **ecol** for the first time during a Stata session. It simply indicates that **ecol** has set up a menu item under `User > Statistics` and added a few environment variables that will preserve any changes made to **ecol**’s settings for the duration of the session (See section 3.2

²Stata may let you keep a `profile.do` on (the root of) a personal network drive and run the `sysdir` command automatically from there, depending on the value of the Windows `USERPROFILE` environment variable. See “Getting Started With Stata”, Appendix A.7 or www.stata.com/help.cgi?profilew for information on setting up a Stata profile.

³See footnote 2.

for more information on these settings). The second line of output is the version number of the **ecol** plugin. Running `ecol` with no parameters should bring up the **ecol** dialog box (click Cancel to close it again).

If Stata is able to do these things without any error messages being issued, you are ready to start doing ecological inference. However, if the core files of **ecol** are not available to Stata, an error will occur at some point: The error numbers⁴ associated with missing files are 199, 111, and 601, and they mean (respectively) that `ecol.ado`, `ecol.dlg`, and `ecol.s.plugin` could not be found. If any of these errors occur, you may wish to review the installation instructions above.

2 Using **ecol**

2.1 The required data

Ecol requires two sets of data describing the behavior of the same group of individuals at two different elections. (The tool can also be used for other purposes, such as estimating the degree of class voting, but for the present we shall restrict ourselves to voter mobility estimation, which is the most common application of ecological inference). The data should be in the form of absolute numbers, and may be thought of as the observed marginal counts of a number of cross tabulations – each line in the data set represents a district, so the marginal counts should sum to approximately the same total in each district. The crucial assumption is, that the data describes roughly the same individuals at two points in time. In the event that the data does not entirely sum to the same totals, **ecol** will fit the estimates to match the second set of observations specified.

The following example uses data describing voting behavior at two Danish general elections (those in 1998 and 2001), aggregated to 103 regions. The data file is called `ecol.example.dta`, and you can load it simply by running

```
use http://www.ps.au.dk/stata/ecol_dir/ecol_example.dta
```

or by downloading it from www.ps.au.dk/stata/ecol.dir.

2.2 Missing values

Ecol uses pairwise deletion of missing values, so the presence of several missing values in a variable does not necessarily affect the outcome of the analysis to any

⁴Stata error messages are followed by error numbers in parentheses

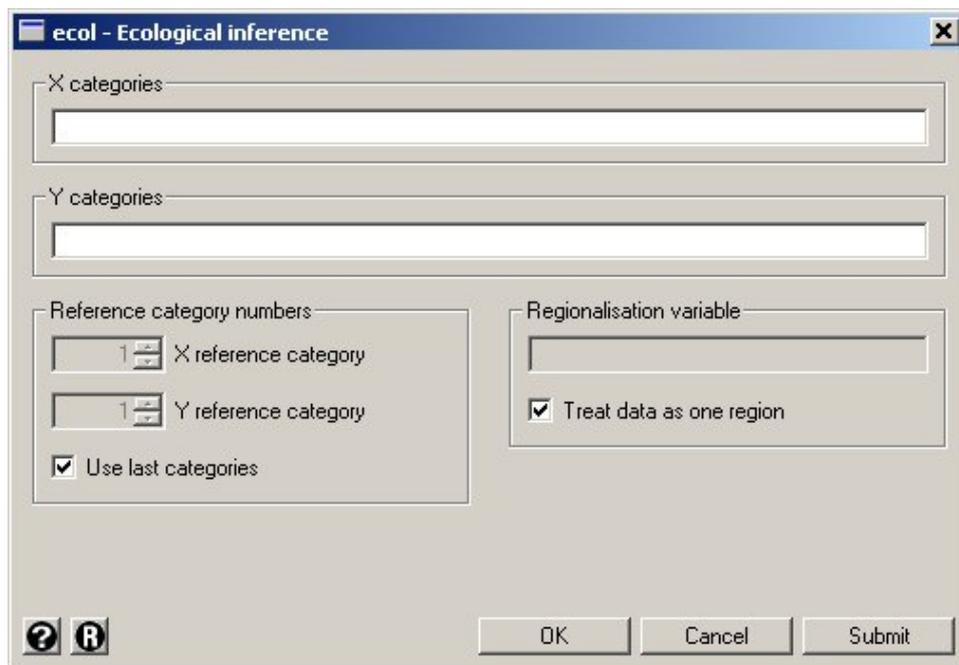


Figure 2: The **ecol** dialog

great extent. If a party does not run in every region, **ecol** will automatically exclude this party from the analysis in the regions where it does not run.

In the current dataset, the variables `x98` and `x01` contain the number of votes for independent candidates cast in each district, and these variables have missing values for entire regions. **ecol** is able to handle these variables, but as they represent a very small number of voters they will be excluded from the analysis.

2.3 The **ecol** dialog

Having installed **ecol** and prepared your data, you are now ready to run **ecol**. Like most other Stata commands, **ecol** comes equipped with a dialog box that allows the user to 'build' the command by clicking buttons and filling in text boxes rather than typing out the entire command on the Stata command line. Open the dialog box by selecting the menu item `User > Statistics > ecol` or by running `ecol` (with no parameters).

The dialog, which is shown in Figure 2, consists of four boxes: In the first two boxes, marked X and Y categories, you enter the names of the parties running at the two elections. The order in which the names are entered will determine the order in which **ecol** displays the results, so parties running at both elections should be

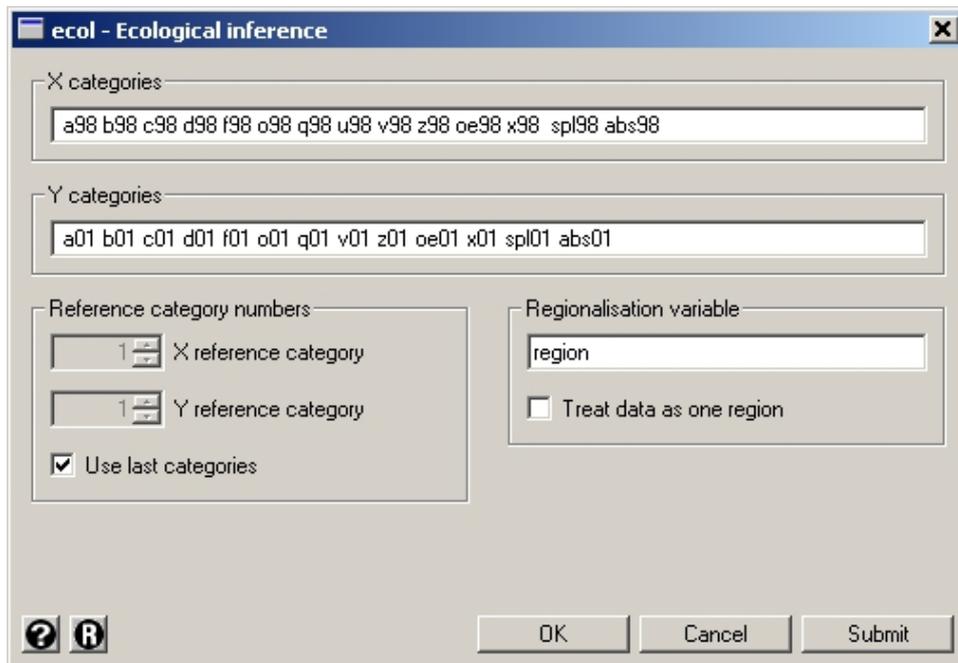


Figure 3: The **ecol** dialog, filled in

placed at identical positions (as this puts the proportions of stationary voters on the diagonal of the estimate matrix). Place the cursor in the X categories box and enter all the appropriate variables from the first election, remembering to exclude `val98` and `voters98`. This is most easily done by clicking the names in the Variables window. (If this window is not displayed, call it up by selecting Variables from the Window menu or by pressing [Ctrl+6]). When you're done, the list should read `a98 b98 c98 d98 f98 o98 q98 u98 v98 z98 oe98 x98 spl98 abs98`. Fill in the Y categories list in the same way.

The third box, marked Reference category numbers is where you tell **ecol** which party is to be used as reference category at each election. If you do not disable the Use last categories checkbox, **ecol** will assume that the reference parties are located at the end of each list. Checking the lists above, we find that the last categories are the number of abstainers `abs98` and `abs01`, and since this is usually rather a neutral category there is no reason to specify any other reference categories.

Finally, in the Regionalization variable box, you may specify the name of a variable that divides the observed districts into politically homogenous regions. Deactivate the Treat data as one region checkbox to activate the text box and enter `region`. The dialog box should now resemble Figure 3. Click OK to run the analysis.

Stata Results

```
. bysort region: ecol a98 b98 c98 d98 f98 o98 q98 u98 v98 z98 oe98 x98 spl98 abs98 / a01 b01 c01 d0
> 1 f01 o01 q01 v01 z01 oe01 x01 spl01 abs01
```

Regionalization by: 'region', 8 regions, 103 districts

Results for entire country (8 regions, 103 districts)

Total %	a01	b01	c01	d01	f01	o01	q01	v01	z01	oe01	x01	spl01	abs01	Total
a98	22.45	0.20	0.60	0.09	0.33	2.99	0.06	0.64	0.10	0.13	0.01	0.16	2.88	30.64
b98	0.05	2.12	0.24	0.29	0.04	0.04	0.01	0.35	0.00	0.02	0.00	0.03	0.09	3.29
c98	0.12	0.55	4.21	0.10	0.12	0.47	0.04	1.59	0.01	0.02	0.00	0.03	0.36	7.62
d98	0.05	0.46	0.81	0.57	0.07	0.13	0.10	1.28	0.00	0.01	0.00	0.01	0.18	3.68
f98	0.29	0.47	0.28	0.07	4.09	0.13	0.01	0.14	0.01	0.34	0.00	0.03	0.60	6.46
o98	0.56	0.06	0.22	0.03	0.04	4.12	0.05	0.82	0.04	0.00	0.00	0.10	0.28	6.32
q98	0.02	0.03	0.11	0.01	0.00	0.07	1.44	0.38	0.02	0.00	0.00	0.01	0.05	2.14
u98	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.10	0.01	0.01	0.00	0.01	0.07	0.27
v98	0.09	0.11	0.35	0.02	0.00	0.62	0.08	19.00	0.03	0.01	0.00	0.06	0.11	20.47
z98	0.32	0.02	0.34	0.05	0.02	0.23	0.02	0.43	0.07	0.04	0.00	0.10	0.43	2.06
oe98	0.05	0.12	0.03	0.04	0.46	0.01	0.00	0.08	0.00	1.34	0.00	0.00	0.17	2.31
x98	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	0.05
spl98	0.06	0.01	0.01	0.00	0.02	0.10	0.00	0.24	0.02	0.01	0.00	0.06	0.11	0.65
abs98	1.02	0.32	0.62	0.23	0.30	1.41	0.15	1.89	0.17	0.14	0.01	0.28	7.49	14.05
Total	25.09	4.48	7.82	1.53	5.50	10.35	1.97	26.95	0.48	2.07	0.03	0.88	12.85	100

Figure 4: Estimated voter transitions, Danish general elections of 1998 and 2001

2.4 Interpreting the results

The results of the analysis are shown in figure 4. This matrix describes the estimated mobility of voters between the parties. Aside from the high degree of loyalty indicated by the clustering of high values along the diagonal, they're not particularly informative. To see how the voters moved between the first and second elections, we need the estimates recalculated so that each row sums to 100%. To obtain the estimates in this form, we set the `pct` option to `rows` with the command `ecol pct rows` and run the analysis again. The previous `ecol` command may be recalled by opening the `ecol` dialog (run `ecol` or use the menu item).

This way of displaying the estimates (see Figure 5) reveals a number of interesting conclusions: The Liberals (party `v`) have been able to attract substantial proportions of voters from nearly all other parties, while at the same time keeping nearly all of those who voted for them in the previous election. The migration of 10% of Social Democratic voters (`a`) to the right-wing Danish People's Party (`o`) is also quite remarkable, particularly since the otherwise successful Liberal Party, located between these two parties on a traditional left-right scale, seems to have been almost completely unsuccessful in attracting any Social Democrats.

Row %	a01	b01	c01	d01	f01	o01	q01	v01	z01	oe01	x01	sp101	abs01	Total
a98	73.29	0.64	1.97	0.31	1.06	9.75	0.19	2.08	0.32	0.43	0.02	0.52	9.41	100
b98	1.51	64.55	7.18	8.78	1.32	1.31	0.38	10.72	0.11	0.52	0.01	0.82	2.79	100
c98	1.51	7.27	55.22	1.35	1.61	6.20	0.47	20.83	0.17	0.28	0.01	0.36	4.71	100
d98	1.41	12.46	22.05	15.62	1.86	3.63	2.61	34.72	0.13	0.31	0.01	0.36	4.83	100
f98	4.45	7.29	4.37	1.05	63.36	2.00	0.16	2.21	0.09	5.23	0.01	0.44	9.34	100
o98	8.82	0.99	3.44	0.45	0.64	65.17	0.76	13.02	0.59	0.06	0.02	1.58	4.44	100
q98	0.86	1.21	5.16	0.36	0.06	3.27	67.19	17.81	0.94	0.06	0.01	0.65	2.43	100
u98	6.90	5.81	2.81	3.87	2.98	5.32	1.08	37.49	2.18	2.18	0.08	2.29	27.00	100
v98	0.43	0.51	1.69	0.12	0.02	3.00	0.40	92.79	0.15	0.06	0.00	0.30	0.52	100
z98	15.51	0.77	16.69	2.32	0.90	11.36	1.05	20.70	3.32	1.82	0.02	4.71	20.83	100
oe98	2.21	5.04	1.21	1.81	19.73	0.56	0.17	3.66	0.07	57.83	0.01	0.18	7.51	100
x98	4.48	3.57	6.72	2.19	3.78	10.85	0.57	11.52	0.51	2.40	0.04	2.55	50.84	100
sp198	8.51	1.77	1.05	0.39	2.96	15.97	0.65	37.46	3.37	1.23	0.04	9.20	17.41	100
abs98	7.26	2.27	4.40	1.67	2.12	10.03	1.09	13.48	1.22	1.02	0.10	2.00	53.32	100
Total	25.09	4.48	7.82	1.53	5.50	10.35	1.97	26.95	0.48	2.07	0.03	0.88	12.85	100

Figure 5: Estimated voter transitions, Danish general elections of 1998 and 2001 (row percent)

3 Syntax and settings

3.1 Ecol command syntax

Unlike most Stata commands, **ecol** requires you to supply two sets of variables with at least two categories in each, the sets being separated by a forward slash (/). The minimal **ecol** syntax thus describes a double binary choice, e.g.

```
ecol right98 left98 / right01 left01
```

where right and left are sums of voters voting for right- and left-wing parties, respectively. As mentioned above, **ecol** will default to using the last categories listed in each set, but you may use any other category by adding the number of the category in parentheses at the end of each list, like this:

```
ecol abstain98 left98 right98 (1) / abstain01 left01 right01 (1)
```

This example uses abstainers as the reference 'party'. **Ecol** allows you to use abbreviated variable names, variable lists and wildcards, so the somewhat unwieldy **ecol** statement we constructed above might also have been written as

```
ecol a98-abs98 / a01-abs01
```

or even

```
ecol *98 / *01
```

Be careful when combining variable lists and wildcards with numbered reference parties: the position of any one party in the final, 'expanded' list may not be entirely obvious beforehand.

3.2 Ecol settings

Settings will be stored for the duration of a single Stata session, but will not be saved when you shut the program down. Common ways of dropping variables or estimates (i.e. `clear`) will not affect the settings, but running macro `drop .all` will reset them. You can also use `ecol reset` to restore default settings at any time.

The options available in **ecol** 0.0 are:

ecol version This option returns the version numbers for the **ecol** plugin.

ecol reset Resets all options to their default values.

ecol options Displays the current setting of all options.

ecol pct This option determines the way the estimates sum to 100%. By default, it is set to `total`, the alternatives are `rows` and `columns`.

ecol precision This option determines the precision with which the estimates are displayed. By default, it is set to `low` (displays two decimals), the alternative being `high` (which sets the number to seven). Note that this options does not affect the internal precision used by the plugin - changing it affects neither the numerical value of the estimates or the speed with which they're calculated.

ecol regions Setting this option to `display` will cause **ecol** to output the estimates for all the regions separately before displaying the total for all areas. The default setting is `hide`

ecol iterations # The estimates for each district are fitted to the marginals using an iterative process. The results produced by this process may fail to have converged to 100% within the first hundred iterations, and in this case **ecol** will normally end the fitting process and use the estimates as they are. If you want **ecol** to wait even longer for convergence to occur, you may set the `iterations` option to some number larger than 100.

Ecol allows you to abbreviate options to a minimum of three characters (i.e `ecol reg` `dis` is equivalent to `ecol regions display`).

If you will be using **ecol** regularly, it may be worth the effort to exchange the default setup with a customized one, run from your Stata profile (see “Getting Started With Stata”, Appendix A.7 or www.stata.com/help.cgi?profile for information on Stata profiles).

The options can *not* be set from your profile.do-file by running `ecol [option]`, but you can set them by using this a simple workaround: First, set all the options the way you want them. Next, run `macro list` and note the values of `ECOL_OPTIONS` and `ECOL_MAXITER`: these two global macros store your settings. Next, add these lines to profile.do:

```
global ECOL_OPTIONS = [value]
global ECOL_MAXITER = [value]
window menu append item "stUserStatistics" "ecol" "ecol"
window menu refresh
```

replacing the `[value]`'s with the values you recorded in the previous step (and remembering to leave a blank line at the bottom). These four lines will completely replace **ecol**'s default setup routine.