Stata How-to:
Export regression results using `esttab`

Patrick Blanchenay

Contents

Often, we would like to report the results of regressions in an article or a report. Instead of creating tables by hand, Stata can automatically generate Microsoft Word documents with the table already formatted. This is done using the **estout** package, which provides a command **esttab** for exporting results to Word. It allows to create a table reporting results of one or several regressions.[1]

1. Installation (do only once)

If this is the first time you use the package **estout**, you first need to install it. You need to be connected to the Internet. To install, type:

```
ssc install estout, replace
```

and let Stata install the package. Once installed, we are able to use the **esttab** command.

---

[1] The **esttab** command can also create tables to be embedded directly in LaTeX, a typesetting syste more commonly used in economics than MS Word, by using the **tex** option instead **rtf** option in what follows.

2. Use

2.1. Specify the working directory

Like any command that creates new files, **esttab** will create a Word document in the working directory. It is therefore important to make sure we specify the working directory for Stata (otherwise your file will get lost):

```
cd "C:/Users/Patrick/Desktop"
```

2.2. Loading data

For this, we are going to use a fictitious dataset provided by Princeton University. Make sure you are connected to the Internet and run:

```
use "https://dss.princeton.edu/training/Panel101.dta", clear
```

This will automatically download the dataset and load it into Stata memory.

2.3. Recommended Workflow

Run the regressions that we are interested in, storing after each regression the estimates (which contains the coefficients, standard errors, R-squared, etc.), under a suitable name. Once we have run all the regressions we need, we export the results of all these estimates in one pass to a single Word document.

3. Storing regression estimates for later use

Regression results are stored using **eststo** followed by a name that will identify the regression. Each regression result can then be recalled using that name.

For instance, imagine we estimate

$$y_i = \alpha + \beta x_{1i} + u_i \qquad \text{(Restricted OLS)}$$

and store the estimate under the name "OLSRestricted". We then include an additional regressor x2, estimating

$$y_i = \alpha + \beta x_{1i} + \gamma x_{2i} + e_i \qquad \text{(Full OLS)}$$

and store the estimate under the name "OLSFull".

```
reg y x1, robust
eststo OLSRestricted

reg y x1 x2, robust
eststo OLSFull
```

4. Exporting estimates using `esttab`

Once this is done, to export our OLS estimates "OLSRestricted" and "OLSFull" to a Word document called **myOLSresults.rtf**, we use the command:

```
esttab OLSRestricted OLSFull using "myOLSresults", rtf replace se label
```

Note the following options:

- **rtf** tells Stata to create a Word document (with extension .rtf).
- **replace** tells Stata that if the file **myOLSresults.rtf** already exists, it should erase it first and replace it with this one.
- **se** tells Stata to display standard errors below the coefficients.
- **label** tells Stata to use variable descriptions instead of variable names; this is usually easier to understand for readers who are not familiar with the dataset, but can be omitted if variables have very long descriptions.

We obtain a Word document with the following table:

	(1) Outcome Y	(2) Outcome Y
Predictor x1	494988913.9 (690156733.6)	551342424.2 (686895122.1)
Predictor x2		38082005.6 (247779427.1)
Constant	1.52432e+09* (663633136.7)	1.48270e+09* (659519033.0)
Observations	70	70

Standard errors in parentheses
$^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Each column represents one of the regression we ran and exported. The title of the column, by default, is a number. Below is the dependent variable, here y in both regressions.

4.1. Mixing OLS and IV (2SLS)

Imagine that we then want to rerun the last model but instrumenting variable x1 with instrument x3, and store this as "IVFull":

```
ivregress 2sls y (x1 = x3) x2, robust
eststo IVFull
```

Nothing prevents us from exporting together OLS and 2SLS results, perhaps in a new file:

```
esttab OLSRestricted OLSFull IVFull using "allresults", rtf replace se label // we add estimates IVFull
```

	(1) Outcome Y	(2) Outcome Y	(3) Outcome Y
Predictor x1	494988913.9	551342424.2	-1.02418e+10
	(690156733.6)	(686895122.1)	(4.92967e+10)
Predictor x2		38082005.6	-1.82329e+09
		(247779427.1)	(8.48021e+09)
Constant	1.52432e+09*	1.48270e+09*	8.72586e+09
	(663633136.7)	(659519033.0)	(3.30563e+10)
Observations	70	70	70

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: It is important to remember that for the IV model in column (3), the coefficient reported for Predictor x1 is not the OLS coefficient, but the two-stage least square coefficient (which if the 3 requirements for IV are satisfied, is the LATE). It is not obvious from the default way **esttab** labels columns which ones are OLS, which ones are 2SLS.

In general we would like to indicate to the reader which columns are OLS models, and which are 2SLS (see below).

5. Important additional options

5.1. VERY IMPORTANT: Specifying model titles

As you can see from the table above, nothing alerts the reader that column (1) and (2) were computed using OLS, while column (3) reports the results of the IV estimation (two-stage least squares).

To add specific titles to each model, we can use the option **mtitles()**, where each model name should be enclosed in quotes " " and separated by space.

```
// Note: we use "///" to make commands carry on several lines
esttab OLSRestricted OLSFull IVFull using "allresults", ///
        rtf replace se label ///
        mtitles("Restricted OLS" "Full OLS" "IV")
```

produces a Word document **allresults.rtf** containing the following table, which contains informative titles about the different models:

	(1) Restricted OLS	(2) Full OLS	(3) IV
Predictor x1	494988913.9 (690156733.6)	551342424.2 (686895122.1)	-1.02418e+10 (4.92967e+10)
Predictor x2		38082005.6 (247779427.1)	-1.82329e+09 (8.48021e+09)
Constant	1.52432e+09* (663633136.7)	1.48270e+09* (659519033.0)	8.72586e+09 (3.30563e+10)
Observations	70	70	70

Standard errors in parentheses
$^*\,p < 0.05$, $^{**}\,p < 0.01$, $^{***}\,p < 0.001$

5.2. Adding notes

Since our table does not feature the dependent variable anymore, it might be useful to specifiy as a note. We will also remind the reader that we are using robust standard errors (since we are). We can accomplish this with the option **addnotes()**, where each additional note should be enclosed in quotes " " and separated by space.

```
// Adding notes
esttab OLSRestricted OLSFull IVFull using "allresults", ///
    rtf replace se label /// our usual options
    mtitles("Restricted OLS" "Full OLS" "IV") /// adding model titles
    addnotes("In all regressions, the dependent variable is y (Outcome Y)." /// adding notes
     "Robust standard errors.")
```

This produces the following table:

	(1) Restricted OLS	(2) Full OLS	(3) IV
Predictor x1	494988913.9 (690156733.6)	551342424.2 (686895122.1)	-1.02418e+10 (4.92967e+10)
Predictor x2		38082005.6 (247779427.1)	-1.82329e+09 (8.48021e+09)
Constant	1.52432e+09* (663633136.7)	1.48270e+09* (659519033.0)	8.72586e+09 (3.30563e+10)
Observations	70	70	70

Standard errors in parentheses
In all regressions, the dependent variable is y (Outcome Y).
Robust standard errors.
$^*\,p < 0.05$, $^{**}\,p < 0.01$, $^{***}\,p < 0.001$

5.3. Adding R^2 and F-statistic

It is possible to add summary statistics to estimates of linear regressions, using the **stats()** options, with **r2** for R^2, and **F** for the F-statistic testing the null hypothesis that coefficients are *all* zero. We keep including **N** as it is

always important to report the number of observations. Note the use of **labels()** inside **stats()**, to give these stats readable labels.

```
// Adding R2 and F-statistic
// Note the labels for those
esttab OLSRestricted OLSFull IVFull using "allresults", ///
    rtf replace se label /// our usual options
    mtitles("Restricted OLS" "Full OLS" "IV") /// adding model titles
    addnotes("In all regressions, the dependent variable is y (Outcome Y)." /// and notes
    "Robust standard errors.") ///
    stats(N r2 F, labels("Observations" "R-squared" "F-statistic")) // adding stats, with nice
        readable labels
```

This produces the following table:

	(1) Restricted OLS	(2) Full OLS	(3) IV
Predictor x1	494988913.9	551342424.2	-1.02418e+10
	(690156733.6)	(686895122.1)	(4.92967e+10)
Predictor x2		38082005.6	-1.82329e+09
		(247779427.1)	(8.48021e+09)
Constant	1.52432e+09*	1.48270e+09*	8.72586e+09
	(663633136.7)	(659519033.0)	(3.30563e+10)
Observations	70	70	70
R-squared	0.00590	0.00613	.
F-statistic	0.514	0.323	

Standard errors in parentheses
In all regressions, the dependent variable is y (Outcome Y).
Robust standard errors.
$^* p < 0.05, ^{**} p < 0.01, ^{***} p < 0.001$

Since R^2, and F-statistic are not defined for IV estimations, they remain unreported for column (3). They can be nonetheless important where reporting separately the first-stage of an IV estimation, which is a linear regression.

5.4. Different tables for different models

In each Word document, there is a single table. If we would like to produce one table that reports "OLSRestricted" and "OLSFull" and one table that reports "OLSFull" and "IVFull", we would export them in different files. This is easy to do by adjusting the list of estimates that we export, just after the **esttab** command, and using different filenames:

```
// Exporting different models to different files
esttab OLSRestricted OLSFull using "justOLS", /// the file will be justOLS.rtf
    rtf replace se label ///
    mtitles("Restricted OLS" "Full OLS") ///
    addnotes("In all regressions, the dependent variable is y (Outcome Y)." ///
    "Robust standard errors.") ///
    stats(N r2 F, labels("Observations" "R-squared" "F-statistic"))
esttab OLSFull IVFull using "fullRegressions_only", /// the file will be fullRegressions_only.rtf
    rtf replace se label ///
    mtitles("Full OLS" "2SLS") ///
```

```
        addnotes("In all regressions, the dependent variable is y (Outcome Y)." ///
        "Robust standard errors.") ///
        stats(N r2 F, labels("Observations" "R-squared" "F-statistic"))
```

6. An example with IV

To illustrate the flexibility of the process, we perform below an example which reports in the same table the first stage of an IV and the 2SLS estimates, where the x1 is the treatment variable, which we instrument by x2, controlling for x3.

```
// IV first stage
// regressing x1 on its instrument x2, controlling for x2
reg x1 x2 x3, robust
eststo IVfirststage

// IV approach using two-stage least squares
ivregress 2sls y (x1 = x2) x3, robust
eststo IVFull

esttab IVfirststage IVFull using "IV_combined", ///
        rtf replace se label ///
        mtitles("IV-First Stage" "2SLS") ///
        addnotes("Robust standard errors." "Dependent variable in (1) is x1 (Predictor x1). Dependent
            variable in (2) is y (Outcome 1).") ///
        stats(N r2 F, labels("Observations" "R-squared" "F-statistic"))
```

which produces:

	(1) IV-First Stage	(2) 2SLS
Predictor x2	-0.177***	
	(0.0420)	
Predictor x3	-0.00858	88383446.0
	(0.0332)	(197609858.3)
Predictor x1		64756677.9
		(1.61364e+09)
Constant	0.678***	1.73577e+09
	(0.0575)	(1.07220e+09)
Observations	70	70
R-squared	0.256	0.00324
F-statistic	11.43	

Robust standard errors in parentheses.
Dependent variable in (1) is x1 (Predictor x1). Dependent variable in (2) is y (Outcome 1).

(Note that here, we are just doing this for the sake of the example. Any econometric approach, in particular IV, must be carefully argumented.)