REGULARISATION IN STATA

Regularisation methods, like the Least Absolute Shrinkage Operator (LASSO), have become increasingly popular in applied econometric research. Stata Lasso (written by Achim Ahrens, Christian Hansen and Mark Schaffer) is a recent package that allows researchers to estimate this type of routine in Stata. The econometrics behind the package goes beyond the scope of this post, which just provides a "mostly harmless" teaser into the clear documentation provided by the authors of the package and related papers.

Simply put, regularisation methods are useful for dealing with situations when the number of observations in our dataset might be relatively small compared to the number of candidate variables we wish to use in our statistical model. A succinct introduction to regularisation (or, penalised least-squares regression) methods is summarised in a working paper by Ahrens, Hansen and Schaffer, the authors of the package. The package's website contains many more resources, commands, and points to a more detailed set of references behind these methods.

These methods can help solving problems of:

- prediction (e.g. what is the household's predicted consumption level given their assets);
- model-selection (e.g. what genotypes are most relevant for a particular trait), and even;
- estimation of the effect of a particular variable on an outcome, with many potential controls (&/or instruments) available.

A neat command in the package, **pdslasso** can solve the last type of problems, especially in quasi-experimental research settings.

pdslasso estimates the parameters (and the appropriate standard errors) for one or few regressors of particular interest to us, in settings where we need to pick control variables (and/or instruments) from a potentially large number of candidates, maybe even larger than the number of observations in our dataset. Belloni, Chernozhukov and Hansen (2014) present the estimators behind the command and settings that are most appropriate for their use.

Roughly speaking, the routine implemented by **pdlasso** entails two components: a first component that selects a small number of controls from the large set of potential candidate variables (using LASSO-type models), and a second component that estimates the effects of interest of a particular treatment variable, having controlled for the selected variables from the first component.

The syntax in Stata:

Install **pdslasso** and **lassopack** using **ssc** install, since **pdlasso** invokes a command that is part of **lassopack** as part of its routine). The syntax for **pdslasso** is:

```
pdslasso [outcome var] [treatment var(s)] ( [(large) list of controls] ) , [options]
```

The usual cluster and robust options are available to take into account heteroscedasticity in our data, though be aware that using these options may also influence the value of the final coefficients of interest, because they affect the selection component of the routine.

Some of the command options allow users to add their own judgement on the relevance of certain controls from the potentially large set of controls available, rather than leaving all the work of picking them to the routine.

For example:

> **`aset()`** allows you to choose a subset of potential controls that should be included in the model estimating the effect(s) of interest, regardless of whether this subset of potential controls were selected by the routine and in addition to any of those controls that were selected by the routine.
>
> **`partial()`** allows you to specify a subset of controls that should be partialled out (in the Frisch-Waugh-Lowell sense) before starting the routine.
>
> **`pnotpen()`** allows you to specify a subset of controls that should also be included in the models estimated in every component of the routine.

The documentation of this command (**`help pdslasso`**) provides a clear step-by-step demonstration of its most important options applied to the data from Acemoglu, Johnson and Robinson (2001)'s highly-cited paper examining the effect of institutions on economic performance. Have fun learning!

Giulio Schinaia, Research Assistant in Development Economics, Blavatnik School of Government, Oxford
10 June 2019