Advanced Applied Econometrics – selected OLS topics

Stata exercise on Regression Anatomy Theorem (Frisch-Waugh)

By Felix Weinhardt

In this exercise we will apply the Frish-Waugh theorem to gain a better understanding of the variation that is used in OLS regressions for the coefficient estimation, in the case of multiple regressions. Try to be tidy and write everything into a do-file.

For this, open up the automobile sample dataset that comes with Stata by typing the following command:

```
sysuse auto
```

This is one of the many in-build datasets, which is often used for simple examples or illustrations of Stata commands.

We are interested in the effect of vehicle length on the price.

1. Run a simple bivariate regression and comment on the findings
2. Add as control variables weight headroom and mpg. Comment on the findings
3. According to the Frish-Waugh theorem we can also obtain the OLS estimate from the multiple regression by running a bivariate regression of the price on a residualised version of the length. Can you show this?
4. The package "reganat" offers a nice way of graphically showing partial relations, with and without co-variate controls. First, you will likely need to install this package to your version of Stata (note, this only needs to be done once). Many additional packages (often in .ado-files) can be loaded into Stata. Over time, with updates, important and successful packages may be included with the main version so that user-installation is not necessary.
    a. Install this by package by typing "`ssc install reganat`". You might also just type "`help reganat`" and click the install button there. If the package is installed already, you will get to see the help file directly.
    b. Now generate the plots (bivariate and multivariate) for the relations between vehicle length and price.

Which variables should be included? Ideally, theory drives these choices. Often, data availability is key. There are also ML tools that choose control variables for you, often based on the idea of extracting "orthogonal" variation in the variable of interest. None of these methods solves issues with respect to unobservable selection.