The Great Separation – Replication Package

The *Great Separation* is based on fine-grained administrative data that cannot be published for obvious privacy reasons. Therefore, you will not find the original raw data in this replication package. However, you will find the generic scripts to produce estimates in a given country (provided you get access to the data from the relevant statistical authority), the aggregated country estimates, and the estimation script based on the aggregated country data.

Data folder

The data folder contains the aggregated country estimates in two formats: csv and rds (for the R software). The *Variables_Categories_list.xls* file provides a short description of the data, the variables and the categories.

Scripts folder

The *Estimates_On_Aggregated_Data.R* R script generates the article's figures and tables based on the aggregated country datasets in the Figures and Tables folders. All you have to do is install the necessary packages and run the script with *R* and *R Studio*.

The scripts in the *Scripts for LEED data* folder come in two sets. The *COIN Shared programs - All countries* folder contains the generic scripts in SAS (France, Canada) or Stata language. Note that these scripts require some adaptation in each country. The *France - Specific scripts* folder contains the scripts for France used in particular for Figure 5 and various robustness checks in the supplementary files.

Figures and Tables folders

These folders remain empty until you run *Estimates_On_Aggregated_Data.R*.