README and Guidance

Overview

This replication package contains a description of the data sources used in the paper *"The Causal Effects of R&D Grants: Evidence from a Regression Discontinuity"*, as well as the Stata code used to generate the tables and figures of the paper and online appendix.

Data Availability and Provenance Statements

The two main sources of data used in the paper are the EASME's database on the SME Instrument and Bureau Van Dijk's (BvD) ORBIS database.

1. Data on SME Instrument competitions are confidential and proprietary to the Executive Agency for Small and Medium Enterprises (EASME), which is now called European Innovation Council and SMEs Executive Agency (EISMEA). Information on grant-winning firms is public (https://sme.easme-web.eu) whereas information on the rest of applicant firms is not.

EASME granted the authors access to the data for the competitions held during 2014-2017 (EASME, 2018) under the agreement that they could not be made publicly available. Researchers interested in accessing the data should contact the agency at https://eismea.ec.europa.eu/contact-form_en. As the data source is not widely used and the process is not standardized, it may take several months to sign data use agreements and gain access to the data.

Datafile: `SME_Instrument.dta` (not provided)

2. Firm patent, balance-sheet, private equity and exit data used in the paper are drawn from ORBIS provided by Bureau Van Dijk (BvD) (Bureau Van Dijk, 2021). ORBIS contains information on around 300 million companies across the globe. Given the proprietary nature of the data, we restrain from making them available. Researchers who possess (themselves or through the institution they belong to) a valid BvD-ORBIS license can have access to the data. The authors accessed the data through the licence owned by Scuola Superiore Sant'Anna in Pisa (Italy). Conditional on subscription, a researcher can access the BvD-ORBIS database in three ways:

- BvD proprietary browser online (orbis.bvdinfo.com);
- BvD historic (CD/DVD-ROM, Blu-Ray) disks;
- Through the Wharton Research Data Services (WRDS) from the Wharton School at the University of Pennsylvania.

Data from ORBIS were last extracted in March 2020. The analysis is based on data for all SME Instrument competitions during the period 2014-2017. We use information on company names and country of origin from `SME_Instrument.dta` to match participants firms with data available from ORBIS. Given the absence of VAT numbers for participant firms, to link SME Instrument applicants with ORBIS records we employed the BvD's Batch Search string functionality. In more detail, the search was performed using applicants' name and country. To adopt a conservative approach and avoid false positives, we only retained those matches featuring the highest quality

possible, that is, those with "excellent" quality according to ORBIS (>=95% correspondence). Every step was vetted with hand-matching to ensure accuracy of the procedure.

Datafile: `ORBIS.dta` (not provided)

In Table 7 and in the Online Appendix of the paper we make use of additional (public-use) data which come from the following sources:

3. GDP per capita and the ratio of domestic credit to the private sector to GDP are retrievable, respectively, from the World Bank Development Indicators dataset and the Global Financial Development Database (v. October 2019). Data refer to the year 2013. GDP per capita data can be downloaded at https://data.worldbank.org/indicator/NY.GDP.PCAP.KD, whereas the Global Financial Development Database is retrievable from https://www.worldbank.org/en/publication/gfdr/data/global-financial-development-database.

Datafile: `October2019globalfinancialdevelopmentdatabase.xlsx` (can be provided)

Datafile: `World_Bank_GDP_1313.xlsx` (can be provided)

4. Data on GDP per capita at the regional level (NUTS2) are from Eurostat and refer to 2013. They are downloadable at https://ec.europa.eu/eurostat/web/regions/data/database.

Datafile: `NUTS_GDP.xlsx` (can be provided)

5. Data on entry rates at the regional level (NUTS3) and sectoral level (NACE main section) are from Eurostat. They are downloadable at https://ec.europa.eu/eurostat/web/structural-business-statistics/business-demography.

Datafile: `bd_hgnace2_r3_NACE.xlsx` (can be provided)

Statement about rights: I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Summary of availability: Data mentioned in point 1 and 2 cannot be made publicly available. Data mentioned in point 3 and 4 can be made publicly available.

Dataset list

Because of the confidentiality issues discussed in the DAS, this replication package does not contain the data. However, this is a summary of the data structure needed to run the replication code we provide, conditional on obtaining access to the SME Instrument and ORBIS data as discussed above:

Data file	Source	Notes	Provided
`SME_Instrument.dta`	EASME	Confidential	No
`ORBIS.dta`	BvD	Proprietary	No
`NUTS_GDP.xlsx`	Eurostat	Public-use	Can be provided
`World_Bank_GDP_1313...`	World Bank	Public-use	Can be provided

Data file	Source	Notes	Provided
`October2019globalf... DATA_ALL_FINAL.dta`	World Bank	Public-use Combines the data sources, serves as input for the analysis.	Can be provided No

Computational requirements

Software Requirements

- Stata (code was last run with Stata MP version 16.1)
- List of user-written packages:
  - `blindschemes` (version 1.3.0, as of 2021-06-21)
  - `coefplot` (version 1.8.4, 17-12-2020)
  - `erepost` (version 1.0.2, 15-06-2015)
  - `estout` (version 3.24 30-04-2021)
  - `grstyle` (version 1.1.1, 15-09-2020)
  - `ppmlhdfe` (version 2.2.0, 02-08-2019)
  - `reghdfe` (version 5.7.3, 13-11-2019)
  - `rdrobust` (version 8.0.2, 04-03-2020)
  - `rdlocrand` (version 0.9, 18-05-2021)
  - `stddiff` (version 2.1, 8-03-2021)
  - `tuples` (version 4.0.3, 16-05-2021)
  - `winsor2` (version 1.1, 16-12-2014)

Memory and Runtime Requirements

The code was last run on a **4-core Intel-based laptop with MacOS version 11.4**. Computation takes roughly 3 minutes to run for the main analysis. Computation takes 2 hours and 10 minutes to run for both the main analysis and the appendix.

Description of programs

The programs contained are the following:

- `scripts/0.run_file.do`: sets up the environment and runs all do files in appropriate order to generate output. There is no need to change the paths to run all codes.
- All .do files contained in the folder `scripts` are named after specific table(s) and figure(s) in the manuscript and replicate the results contained in such table(s) and figure(s).
- The .do file `scripts/appendix.do` will generate all tables and figures in the same order they appear in the online appendix.

References

Bureau Van Dijk (2021) Company Account Statistics (ORBIS) [dataset] Retrieved from https://www.bvdinfo.com/en-gb/our-products/data/international/orbis

EASME (2018) SME Instrument competition data 2014-2017 [dataset]. Executive Agency for Small and Medium Enterprises (EASME).