

Replication files for Ignaszak and Sedláček “Customer Acquisition, Business Dynamism and Aggregate Growth”

2026-05-06

Table of contents

Data Availability Statement and Data Sources	1
Publicly available data	1
Proprietary data	2
Folder and File Structure	2
List of files	2
./bld/	2
./src/	3
Software Requirements/Settings	3
Replication Steps	4
Data Retrieval	4
Empirical Analysis	4
Model Analysis	4
Complete Set of Routines	4
Runtime	5

The code in this replication package reproduces the empirical and theoretical analysis in “Customer Acquisition, Business Dynamism and Aggregate Growth” by Marek Ignaszak and Petr Sedláček.

The empirical part utilizes publicly available data (BDS and St. Louis Fed’s FRED), replication packages for published papers (Field and Karpoff (2002); Loughran and Ritter (2004); De Loecker, Eeckhout, and Unger (2020)) as well as proprietary data collected and owned by Standard and Poor’s (Compustat and Capital IQ).

Data retrieval utilizes the Python WRDS interface for Compustat and a manual website interface for Capital IQ. The empirical analysis is conducted using the R programming language. The theoretical analysis is conducted using MATLAB. All results are based on data retrieved on May 6, 2026.

Data Availability Statement and Data Sources

The empirical analysis combines open-access aggregate data with proprietary firm-level data that require paid subscriptions to access.

Publicly available data

- Business Dynamics Statistics (BDS) 2023 vintage (although we only use data up to 2019); stored in the ./bld/data/ folder.
 - Firm-size statistics can be downloaded ([source](#))
 - Firm-age statistics ([source](#))
- Macroeconomic data for the US economy sourced from the [FRED database](#); stored in the ./bld/data/ folder.
 - GDP deflator, series id = “GDPDEF”
 - Nominal GDP, series id = “GDP”
 - GDP per capita, series id = “GDPC1”
 - Aggregate private sector employment, series id = “USPRIV”

- Field-Ritter IPO database; stored in `./src/original_data/` folder.
 - Available as part of the data appendix to Field and Karpoff (2002) and Loughran and Ritter (2004).
<https://site.warrington.ufl.edu/ritter/files/IPO-age.xlsx>
- De Loecker, Eeckhout, and Unger (2020) production function elasticities by sector; stored in `./src/original_data/DEU_markup` folder.
 - published as part of the replication files available here: <https://doi.org/10.7910/DVN/5GH8XO>

Proprietary data

- Compustat; stored in `./bld/data` once data retrieval routine is executed (see below)
 - accessed through the [Wharton Research Data Services \(WRDS\)](#) interface which requires a paid subscription to WRDS.
 - We use the North America Fundamentals Annual database.
 - See the source file `./src/data_analysis/get_compustat.py` for details.
- Capital IQ data and [Linking table](#) between Capital IQ and Compustat; stored in `./bld/data` once data retrieval routine is executed (see below)
 - Accessed through an interactive online interface which requires a paid subscription to Capital IQ.
 - See the source file `./src/data_analysis/get_compustat.py` for details.

Folder and File Structure

ROOT_PATH/

```
./bld/
  data/
  figures/
  tables/
./src/
  data_analysis/
  model/
  original_data/
install_packages.R
R_session_info.txt
requirements.txt
settings.json
Makefile
README.pdf
```

List of files

`./bld/`

The folder `./bld/` contains outputs generated by the source code in `./src/`, consisting of intermediate data files, output figures, and LaTeX tables.

`./bld/data/`

The subfolder `./bld/data/` contains datasets used in further analysis. These are:

- Four files `{agg_emp,gdp,GDPC1,gdp_def}.feather` contain US time series data on aggregate employment, nominal GDP, real GDP per capita, and the GDP deflator, respectively.
- Two files `{bds2023_fa,bds2023_fz}.csv` contain Business Dynamics Statistics data by firm age and firm size, respectively.

The remaining subfolders are empty and will be populated with figures and tables by the routines in the `./src/` folder.

`./src/`

The folder `./src/` contains routines used to generate empirical results, theoretical results, and “original data” that has to be downloaded manually.

`./src/original_data/`

This folder contains:

- `IPO-age.csv`: the comma-separated version of the original `.xlsx` (also included for reference) file of the Field-Ritter IPO dates dataset.
- `DEU_markup/data/PF/theta_W_s_window.dta`: De Loecker, Eeckhout, and Unger (2020) production function elasticities.
- `CapitalIQ/CapitalIQ-YearByYear-20251201.xlsx` (balance sheet data) and `CapitalIQ/CapIQ_year_established.xlsx` (firm establishment dates): both manually downloaded Capital IQ data.
 - the replication package includes mock versions of these files which show the required variables and structure but which are populated with random values.

`./src/data_analysis/`

- `prepare-compustat-data.R` cleans raw Compustat data and merges production function elasticities and firm founding dates datasets.
- `prepare-capitalIQ-data.R` cleans raw Capital IQ data.
- `merge-compustat-and-capitalIQ.R` merges Compustat and Capital IQ datasets.
- `calibration_moments_compustat_and_capitaliq.R` generates moments required to calibrate the theoretical model in Section 3, Table 1 and Table 2.
- `empirical_validation.R` generates results in Section 5 of the paper: Table 7 and Figure 4.
- `library.R`, `library_compustat_calibration.R`, and `library_validation.R` define auxiliary functions used in the empirical analysis.

`./src/model/`

- `A_mainscript.m` is the main file that executes all parts of the theoretical analysis. Key components are:
 - `B_calibration_and_fit.m` generates the results in Sections 3.2 and 3.3.
 - `C_counterfactual.m` generates the results in Sections 4.1 and 4.2.
 - `D_growth_policies_baseline.m` generates the results in Section 4.3.
 - `E_restricted.m` calibrates the restricted model in which firms do not have a customer base. Generates results in Section 4.3.
 - `F_restricted_policies.m` generates results in Section 4.3.
 - In addition, there is a list of auxiliary functions each defined in an individual file.

Software Requirements/Settings

We begin with a summary of required software and settings necessary for replication.

Summary:

1. Open `settings.json` and specify the path in which the replication package resides on your computer (`ROOT_PATH`) as well as your WRDS username, and optionally your FRED API key.
2. Install all R (version 4.5.3 or compatible) and packages specified in `install_packages.R`
 - necessary packages can be downloaded by running `Rscript install_packages.R` in the root folder
 - `R_session_info.txt` summarizes the working environment used to generate the empirical results
3. Install Python (version 3.11.6 or compatible) and all Python packages specified in `requirements.txt`
 - necessary packages can be downloaded by running `pip install -r requirements.txt` in the root folder
4. Install a MATLAB instance compatible with version 24.2 (2024b) and the following toolboxes:

Econometrics Toolbox	Version 24.2
Global Optimization Toolbox	Version 24.2
MATLAB Compiler	Version 24.2
MATLAB Compiler SDK	Version 24.2
Optimization Toolbox	Version 24.2
Parallel Computing Toolbox	Version 24.2
Statistics and Machine Learning Toolbox	Version 24.2

Replication Steps

Data Retrieval

Aggregate data

- use R to run `./src/data_analysis/library.R`
- this requires applying for an API key to the FRED database [here](#)
- for ease of replication, we include the downloaded files in this package

BDS data

- use R to run `./src/data_analysis/prepare-aggregate-data.R`
- for ease of replication, we include the downloaded files in this package

Compustat data

- use Python to run `./src/data_analysis/get_compustat.py`
- users will be prompted to enter their WRDS password. Note that users with active two-factor authentication may have to login to their WRDS account using an internet browser before executing the Python file. See [here](#) for more details.

Capital IQ data

- manually download and place the two required files in `./src/original_data/CapitalIQ/`
- follow the structure provided in the Capital IQ mock files in `./src/original_data/CapitalIQ/`

Empirical Analysis

- use R to run `./src/data_analysis/mainscript.R`
- this routine generates results for
 - Table 1, “Data” values in Table 2 and Table 3, Table 7
 - Figure 1 (“Data” values) and Figure 4
 - Appendix A (“Data” values in Table A1, Figure A1), Appendix D (Table A2, Figure A2-A4)

Model Analysis

- use MATLAB to run `./src/model/A_mainscript.m`
- this routine generates results for
 - “Model” values in Table 2 and Table 3, Table 4, Table 5, Table 6
 - Figure 1 (“Model” values), Figure 2-3
 - Appendix A (“Model” values in Table A1, Figure A1)

Complete Set of Routines

Alternatively, if you have GNU Make installed, open the terminal in the root path and execute `make`.

Runtime

The computational time required to replicate all the results in the paper is approximately 15 minutes on a standard laptop. Note that we use convenient initial values in the model analysis.

- De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. “The Rise of Market Power and the Macroeconomic Implications*.” *The Quarterly Journal of Economics* 135 (2): 561–644. <https://doi.org/10.1093/qje/qjz041>.
- Field, Laura Casares, and Jonathan M. Karpoff. 2002. “Takeover Defenses of IPO Firms.” *The Journal of Finance* 57 (5): 1857–89. <https://doi.org/https://doi.org/10.1111/0022-1082.00482>.
- Loughran, Tim, and Jay Ritter. 2004. “Why Has IPO Underpricing Changed over Time?” *Financial Management* 33 (3): 5–37. <http://www.jstor.org/stable/3666262>.