Replication Package for the Data Sections of "Bottom-Up Markup Fluctuations"

By A. Burstein, V.M. Carvalho and B. Grassi

Version: June 4, 2025

System Requirements for Replication: Windows 2012 R2, Stata/MP17, Python: WPy64-3740

Preamble: Data Access

The firm-level data can be accessed through the *Centre d'accès sécurisé aux données* (CASD), a public French organization that manages access to confidential datasets. To access the datasets used in this project, researchers must obtain authorization from the Statistical Secrecy Committee (www.comite-du-secret.fr/). The procedure for gaining access is outlined at the following link: www.casd.eu/gerer-son-projet/procedures-dhabilitation/.

The required datasets are:

- **FICUS**: *Statistique structurelle annuelle d'entreprises* from the SUSE system
- **FARE**: *Statistique structurelle annuelle d'entreprises* from the ESANE system
- **EAP**: *Enquête Annuelle de Production*
- **EACEI:** *Enquête sur les consommations d'énergie dans l'industrie*

As described in the Data section of the main paper and in Section 1.1. below, to construct the main panel of firms from 1994 to 2019, interested researchers will need to (i) merge the FICUS (until 2007) and FARE (from 2008 onwards) datasets, and (ii) generate a consistent sectoral classification.

The codes used to construct this panel from raw data contain unique firm-level identifiers and cannot be shared in this replication package outside the secure environment of the CASD. However, once access to the data is granted by the CASD, it will also provide the codes that cannot be publicly distributed. The relevant codes to construct the panel from the raw data are:

- Master_100222.do
- Import_BalanceSheet_100222.do
- Clean_BalanceSheet_OriginalProfilage_Check040222.do

Given access to the raw datasets above and these three codes via CASD, interested researchers can then use all other codes and auxiliary datasets - described below, from Section 2 and onwards - made available with this package to replicate all empirical results in the paper.

Note: Further information about this project at CASD can be found at: www.casd.eu/en/project/microeconomic-determining-factors-affecting-growth-volatility-and-sectorial-margin-rates/

1) Data Construction

In this section, we describe how the datasets used for estimating markups and conducting our various empirical exercises are constructed.

1.1) FICUS-FARE Firm-Level Panel

As described above, once access is granted, the raw datasets referenced below can be generated with the following do-files, which we can transfer from our CASD account:

- Master_230421.do,
- Import_BalanceSheet_230421.do
- Clean_BalanceSheet_OriginalProfilage_CheckApril21.do

  To obtain these do-files, please contact us with the name of your CASD project. The CASD team will then add the files to your secure environment.

Input: raw data from FICUS and FARE

Output: data_deep/ficusfare_profil_wof_19942019_v0222

1.2) EAP Dataset

The raw EAP data are also available through CASD. The following two do-files generate the dataset referenced below:

- sasimport_april21: a SAS routine that converts the EAP data into CSV format
- loadEAP.do: a Stata script that computes product-level and firm-level prices

Input: raw data from EAP

Output: data_deep\EAP\EAP_firmlevel.dta

2) Data Preparation

Run prepare_datasets.do

Input:

- data_deep/ficusfare_profil_wof_19942019_v0222
- data_deep/gdp_deflator_euklems_nov21 data with deflator from EUKLEMS (generate_deflator_april21.do)
- data_deep\EAP\EAP_firmlevel

Output: This file produces the following sub-datasets:

- data\ficusfare_reduced_9419_v0222

- data\ficusfare_reduced_9419_sec2_v0222
- data\ficusfare_reduced_9419_sec2_year2_v0222
- data\ficusfare_reduced_9419_sec2_year2_firm_va_sevyears_v0222
- data\ficusfare_reduced_9419_sec2_year2_firm_va_sevyears_uniproduct_v0222

3) Markup Estimation

Run markup_estimation_XX.do (set XX=baseline or uniprod)

This script formats and saves the data used by the Python routine, and runs the Python code. The resulting files correspond to different markup specifications.

Note that if the do-file is unable to run the Python routine, you can run the Python script manually after executing the do-file.

The following Python scripts estimates markups for each XX specification (where XX=baseline or uniprod): run_for_stata_FSQ_XX.py and run_for_stata_BFSR_baseline.py. The latter run the estimation on quantity data while the former on revenue data.

Input:

- data/ficusfare_reduced_9419_sec2_year2_firm_va_sevyears_v022 (for baseline)
- data/ficusfare_reduced_9419_sec2_year2_firm_va_sevyears_uniproduct_v0222 (for uniprod)

Output:

- data\data_smallest_for_python_XX.dta
- coefficients_byind_Python_FSQ_XX.dta
- coefficients_byind_Python_BFSR_baseline.dta

4) Baseline Specification Analysis

BCG\BCG_tables_baseline.do

This file produces many of the baseline results reported in the main text (Tables 3, 5, 6, 7, 8, 9, 10, 11, A6, A7, A9, A11, A12, A13, A14, A15, A16).

Note that this file creates intermediate .dta files, which are constructed and used by this script.

Input:

- data\ficusfare_reduced_9419_sec2_v0222.dta
- data\coefficients_byind_Python_FSQ_baseline.dta

Output:

- latex tables in the folder BCG\results_sl: MarkupMShare_FD_03_baseline.tex, MarkupHhi_FD_03_baseline.tex, MarkupSectorOutput_03_baseline.tex, marketShare_SectorOutput_03_baseline.tex, HHI_SectorOutput_03_baseline.tex, SectorMarkup_SectorOutput_03_baseline.tex, SectorMarkup_AggregateOutput_03_baseline.tex MarkupMShare_level_03_baseline.tex, MarkupHhi_level_03_baseline.tex

5) Robustness Analysis

Each file below produces the results for a specific robustness exercise (Table A3). The structure is identical to that of the baseline file described above.

5.1) Outlier treatment

(BCG/BCG_tables_outliers.do)

Input: same as baseline

Output: same folder as baseline BCG\results_sl : file finishing with "01_outliers" instead of "03_baseline".

5.2) Lerner Markup

(BCG/BCG_tables_lerner.do)

Input: only data\ficusfare_reduced_9419_sec2_v0222.dta
Output: same folder as baseline BCG\results_sl: file finishing with "_lerner" instead of "_baseline".

5.3) Period 2009-2019

(BCG/BCG_tables_20092019.do)

Input: same as baseline

Output: same folder as baseline BCG\results_sl: file finishing with "_20092019" instead of "_baseline".

5.4) Revenue-Based Markup

(BCG/BCG_tables_revenue.do)

Input: same as baseline and data\coefficients_byind_Python_BFSR_p30.dta

Output: same folder as baseline BCG\results_sl: file finishing with "_revenue" instead of "_baseline".

5.5) Single Product Firms

(BCG/BCG_tables_singleproduct.do)

Input: data\ficusfare_reduced_9419_sec2_v0222.dta and data\
coefficients_byind_Python_FSQ_uniprod.dta

Output: same folder as baseline BCG\results_sl: file finishing with "_uniprod" instead of "_baseline".

5.6) Estimation Sample

(BCG/BCG_tables_estimsample.do)

Input: same as baseline

Output: same folder as baseline BCG\results_sl: file finishing with "_estimsample" instead of "_baseline".

5.7) Manufacturing firms

(BCG/BCG_tables_manuf.do)

Input: same as baseline

Output: same folder as baseline BCG\results_sl: file finishing with "_manuf" instead of "_baseline".

6) Generate results on the role of entry and exit.

To produce the Table A4 run (BCG/BCG_tables_EntryExit.do)

Input: same as baseline

Output: BCG\results_sl: file finishing with "_EntryExit" instead of "_baseline".

7) Descriptive Statistics and Size–Volatility Relationship

BCG\descriptive_stats_size_vol.do
This script generates the descriptive statistics table (Table A1), the size–volatility relationship used in the calibration (Table A5), and the results reported in Table A2.

Input:

- data\ficusfare_reduced_9419_sec2_v0222.dta
- data\coefficients_byind_Python_FSQ_p30.dta
- data\coefficients_byind_Python_BFSR_p30.dta

Output:

- Produces a stata log file: BCG/logs/BCG_statdes_sizevol_03_baseline.log, replace

8) Energy Data Construction

BCG\construct_eacei.do
This script constructs the sector-level energy dataset and generates intermediate .dta files used in subsequent analyses.

Input:

- data_deep/eacei/eacei`year' with year from 2008 to 2020
- data/ficusfare_reduced_9419_sec2_v0222.dta

Output:

- data/eacei_sector2dlevel.dta

9) Analysis of Marginal Costs, Market Shares, and Markups

BCG\price_markup_marginalcost_energy.do
This script runs the regressions on the sample with price and quantity data, analyzing the relationship between marginal cost, market share, and markups. The results are reported in Table 4, Table A7, Table A8, and Table A10.

Input:

- data\ficusfare_reduced_9419_sec2_v0222.dta
- data\coefficients_byind_Python_FSQ_p30.dta
- data\eacei_sector2dlevel.dta

Output:

- multiple latex tables in the folder BCG\results: FDmarkup_FDmc_03_baseline.tex, FDmarketshare_FDmc_03_baseline.tex, markup_mc_03_baseline.tex, marketshare_mc_03_baseline.tex, mc_IV_mc_ene_gene2_hat_lag_w_03_baseline.tex

10) Produce Data used for Calibration

BCG\BCG_export_for_calibration.do
This script produce the excel file used in the calibration of the model. Note that this file need to be run after running the baseline specification analysis (BCG\BCG_tables_baseline.do).

Input :

- BCG/data/full_firm_baseline_03.dta (produced by BCG\BCG_tables_baseline.do)

Output:

- BCG\output\exportdata_naf_rev2022.xls (NB: This file need to be converted to xlsx format to be used by matlab)