

README: Replication Package for "Technology Sophistication Across Establishments"

Authors: Xavier Cirera (World Bank), Diego Comin (Dartmouth College), Marcio Cruz (IFC)

1. Overview

This replication package corresponds to the paper titled:

"Technology Sophistication Across Establishments"

by Xavier Cirera (World Bank), Diego Comin (Dartmouth College), Marcio Cruz (IFC), accepted for publication in the Quarterly Journal of Economics (QJE). The purpose of the package is to provide all materials required by the QJE data policy, enabling independent replication of the main empirical results.

Users of the database and codes, please cite its source as:

Cirera, X., Comin, D., & Cruz, M. (Forthcoming). Technology Sophistication Across Establishments. The Quarterly Journal of Economics.

2. Contents of Package

The folder structure of this package is as follows:

/1_data/

- a. [*FAT0_raw_data_qje.dta*] - This is the raw data from the survey used to construct all measures.
- b. [*FAT0_reshaped_firm_busfn_updated_new.dta*] - Data with technology measures at the firm-business function level. This dataset is derived from *FAT0_raw_data_qje.dta*.
- c. [*FAT0_firm_level.dta*] - Data with technology measures at the firm-level. This dataset is derived from *FAT0_raw_data_qje.dta*.
- d. [*relative_q_productivities.xlsx*] - Relative productivities of selected business functions.
- e. [*FAT0_reshaped_firm_busfn_consolidated.dta*] - Data with technology measures (both standard and relative technology measures) for selected business functions.
- f. [Figures_A12_A16 -- folder] – This folder provides a list of 27 .csv datasets to replicate figures A12-A16.
- g. [TAS_naukowiec.dta] – Data with financial and employment variables for Poland. This dataset is not available outside Statistics Poland.

/2_codes/

- a. [*0.master.do*] – Master do file.
- b. [*1.data_construction.do*] - Takes *FAT0_raw_data_qje.dta*, constructs all technology and financial measures, and generates the other datasets;
- c. [*2.main_tables*] - Generates Tables 2-10;
- d. [*3.appendix_tables*] - Generates Tables A.22, A.24-A.31, and C.1-C.9.
- e. [*4.main_figures*] - Generates Figures 3-11;
- f. [*5.appendix_figures*] - Generates Figure C.3 and C.4.
- g. [*0_data_merge_poland.do*] – This code is used to merge the financial and employment variables from Poland into the main database. This code is not needed when running the package without the restricted variables for Poland. These do-files are provided for documentation purposes only and are not expected to run with the public version of the package.

We recommend running the scripts in the order listed.

/4_documentation/

FAT_Survey.pdf. This file provides the original FAT questionnaire used for data collection, which includes the variable names for each question.

Implementation_Manual.pdf. The implementation manual includes all instructions for interviewers, training materials, and a full description of the technologies in the grid and variables used in the questionnaire.

Data_dictionary.xlsx. This file provides a description of the variables from the FAT survey used in the paper.

3. Software & Environment

- Primary software used: Stata [Version 18]
- Operating system: tested on [e.g., Windows 10 / macOS Ventura]
- Reproducibility note: We recommend running the scripts in the order specified in ## 2.

4. Data Availability and Access

The data used in this study come from the Firm-Level Adoption of Technology (FAT) survey, designed by Xavier Cirera (World Bank), Diego Comin (Dartmouth College), and Marcio Cruz (IFC), and implemented by the World Bank.

- Provider: The World Bank.
- Coverage: Multi-country firm-level survey on technology adoption. This data covers 15 countries.
- Access Restrictions: The database is proprietary to the World Bank.

The anonymized version of the dataset, including the variables used in this paper, and the documentation will be made available for replication purposes at <https://microdata.worldbank.org/index.php/home>

Data implementation was conducted in collaboration with National Statistical Offices (NSOs) and official agencies in various countries. In the case of Poland, variables related to productivity (e.g., revenue, cost) and workers (e.g., number of workers, education) are not made available outside the NSO.

Tables A26-A30 of the paper also rely on restricted-access information. Additionally, tables A31 and A32 of the appendix utilize a matched employer-employee database for Brazil, sourced from the *Relação Anual de Informações Sociais* (RAIS), which is not publicly available.

Therefore, a full replication of the results, including these variables, requires access to these datasets. The lists of “Tables (main text)” and “Tables (appendix)” indicate which tables contain restricted datasets. Except for tables A26-A32, for which restricted-access datasets are needed, the remaining tables with restricted datasets can be generated using data for the other 14 countries, excluding Poland.

Confidential — available only to researchers approved by the World Bank under strict data-use agreements.

The following table provides information on the source of each (raw) dataset that is used in the analysis.

Dataset	Filename	Description and citation	Access date	Data access
Firm-level Adoption of Technology (FAT) dataset	FAT0_raw_data_qje.dta	<p>This dataset generates the following intermediary data files: FAT0_reshaped_firm_busfn_consolidated.dta, FAT0_reshaped_firm_busfn_updated_new.dta, FAT0_firm_level.dta mgt_z.dta</p> <p>Citation: Cirera, Xavier; Comin, Diego, and Cruz, Marcio. “Technology Sophistication Across Establishments,” Quarterly Journal of Economics, Forthcoming.</p>	2026	<p>Forthcoming on: https://microdata.worldbank.org/index.php/home Available for download.</p>
Firm-level Adoption of Technology (FAT) – Poland dataset (restricted variables)	TAS_naukowiec.dta	<p>The FAT dataset for Poland is hosted in the NSO. Some variables (see data dictionary) are only accessible in this center.</p> <p>Citation: Cirera, Xavier; Comin, Diego, and Cruz, Marcio. “Technology Sophistication Across Establishments,” Quarterly Journal of Economics, Forthcoming.</p>	2026	<p>Confidential</p> <p>Access needs to be requested from Statistics Poland https://stat.gov.pl/en/</p>
Relative Productivities	relative_q_productivities.xlsx	<p>This dataset contains measures of the relative productivity of various technologies. The methodology used to compute these productivities is detailed in Appendix A3 of the accompanying paper.</p> <p>Citation: Cirera, Xavier; Comin, Diego, and Cruz, Marcio. “Technology Sophistication Across Establishments,” Quarterly Journal of Economics, Forthcoming.</p>	2026	<p>Forthcoming on: https://microdata.worldbank.org/index.php/home Available for download.</p>
World Development Indicators (WDI)	country_stats_2080.xlsx	<p>World Bank. (2024). Log of GDP per capita, PPP. <i>World Development Indicators</i> [Data set]. https://databank.worldbank.org/source/world-development-indicators</p>	2024	<p>Open data Forthcoming on: https://microdata.worldbank.org/index.php/home</p>

Brazil FAT dataset merged with RAIS	RAIS_firm_level_2018.dta RAIS_firm_level_2017.dta FAT0_brazil.dta FAT_treated.dta	Used for tables A31-A32. These datasets are confidential as they contain identifiable information for firms (ie, tax ID). The FAT0 Brazil and FAT_treated are derived from the Firm-level Adoption of Technology survey for Brazil, including the FIRM ID from RAIS.	2020	Confidential Access to microdata through the Ministério do Trabalho e Emprego, Brazil. https://www.gov.br/pt-br/servicos/solicitar-vinculos-empregaticios-da-rais
FAT dataset - interviewer data	replication_C28-C29.dta replication_tableC25-C27_brazil.dta replication_tableC25-C27_senegal.dta replication_tableC25-C27_vietnam.dta	These datasets include confidential information on interviewers from the Firm-level Adoption of Technology (FAT) dataset and information from the original sample frame provided by official agencies. Citation: Cirera, Xavier; Comin, Diego, and Cruz, Marcio. “Technology Sophistication Across Establishments,” Quarterly Journal of Economics, Forthcoming.	2026	Confidential These tables use information on enumerators and raw data from the sampling frame: https://www.gov.br/pt-br/servicos/solicitar-vinculos-empregaticios-da-rais https://www.nso.gov.vn/en/homepage/ https://www.ansd.sn/
ISIC classification	isic_data.dta	ISIC Rev. 4 sectoral classification based on the text description of the main product by establishment in the FAT dataset. See Appendix D of the paper. Citation: Cirera, Xavier; Comin, Diego, and Cruz, Marcio. “Technology Sophistication Across Establishments,” Quarterly Journal of Economics, Forthcoming.	2026	Forthcoming on: https://microdata.worldbank.org/index.php/home Available for download.
Firm-level Adoption of Technology (FAT) questionnaire (technology grid from experts) and ChatGPT	Figures_A12_A16.zip (see list of files included below)	Reflects the technology grid in the questionnaire used to collect the Firm-level Adoption of Technology (FAT) dataset, and the validation exercises with ChatGPT.	2026	Forthcoming on: https://microdata.worldbank.org/index.php/home Available for download.

		Citation: Cirera, Xavier; Comin, Diego, and Cruz, Marcio. "Technology Sophistication Across Establishments," Quarterly Journal of Economics, Forthcoming.		
Figures_A12_A16	Agriculture-Harvesting.csv Agriculture-Irrigation.csv Agriculture-Land preparation.csv Agriculture-Packing.csv Agriculture-Storage.csv Agriculture-Weeding.csv Apparel-Cutting.csv Apparel-Design.csv Apparel-Finishing.csv Apparel-Sewing.csv Business Administration.csv Food Processing- Packaging.csv Food Processing-Anti-bacter.csv Food Processing-Food storage.csv Food Processing-Input testin.csv Food Processing-Mixing.csv Marketing.csv Payment.csv Production or service operat.csv planning.csv Quality control.csv Retail-Advertisement.csv Retail-Customer service.csv Retail-Inventory.csv Retail-Merchandising.csv Retail-Pricing.csv Sales.csv Sourcing.csv	These .csv datasets provide information on the ranking of the technology sophistication grid by experts and on the validation exercises using ChatGPT. The methodology is provided in the appendix (sections A1 and A2.1). Citation: Cirera, Xavier; Comin, Diego, and Cruz, Marcio. "Technology Sophistication Across Establishments," Quarterly Journal of Economics, Forthcoming.	2026	Forthcoming on: https://microdata.worldbank.org/index.php/home Available for download.

List of Figures and Tables

Figures (Main Text)

File	Title	DO FILE
Figure_1	General Business Functions and Their Technologies	Text. No code.
Figure_2	Sector Specific Business Functions and Technologies in Agriculture	Text. No code.
Figure_3	MAX _{fj} in two establishments in retail services	5.main_figures.do
Figure_4	Distribution of technology sophistication in Food Processing (Fabrication)	5.main_figures.do
Figure_5	Distribution of MAX-MOST gap across establishments	5.main_figures.do
Figure_6	MAX-MOST GAP across countries	5.main_figures.do
Figure_7	Distribution of Years since Adoption of Top-tier Technologies Conditional on	5.main_figures.do
Figure_8	Distribution of MOST _{fj} , conditional on adopting Top-tier technology	5.main_figures.do
Figure_9	Distribution of MAX _j and MOST _j across establishments	5.main_figures.do
Figure_10	MAX and MOST across countries	5.main_figures.do
Figure_11	Within-country dispersion in technology sophistication and per capita income	5.main_figures.do

Tables (Main Text)

Number	Title	Responsible
Table 1	Comparison of Technology Categories: Business Administration	Text. No code.
Table 2*	Number of establishments in FAT by country, sector, and size	3.main_tables.do
Table 3*	Comparison of baseline and Q-measures of technology sophistication	3.main_tables.do
Table 4	Relationship between technology measures	3.main_tables.do
Table 5	Average GAP, MAX, and MOST, by age and country income groups	3.main_tables.do
Table 6*	Average MAX-MOST gap across establishment distribution across observable char-	3.main_tables.do
Table 7*	Technological sophistication and establishment characteristics	3.main_tables.do
Table 8	Cross-establishment distribution of MAX and MOST	3.main_tables.do
Table 9*	Percentage of cross-establishment variance accounted for by sector dummies	3.main_tables.do
Table 10*	Productivity and Sophistication	3.main_tables.do

Note: (*) These tables use indicators from Poland for which data is not included in this package. All results, excluding the restricted variables from Poland, can be replicated using the package. The sources of information for Table 1 are described in the main text.

Figures (Appendix)

Number	Title	Responsible
Figure A.1	Agriculture - Livestock: Business Functions and Technologies	Text. No code.
Figure A.2	Food Processing: Business Functions and Technologies	Text. No code.
Figure A.3	Wearing Apparel: Business Functions and Technologies	Text. No code.

Figure A.4	Leather and Footwear: Business Functions and Technologies	Text. No code.
Figure A.5	Automotive: Business Functions and Technologies	Text. No code.
Figure A.6	Pharmaceutical: Business Functions and Technologies	Text. No code.
Figure A.7	Wholesale and Retail: Business Functions and Technologies	Text. No code.
Figure A.8	Land Transportation: Business Functions and Technologies	Text. No code.
Figure A.9	Financial Services: Business Functions and Technologies	Text. No code.
Figure A.10	Health Services: Business Functions and Technologies	Text. No code.
Figure A.11	Other Manufacturing: Business Functions and Technologies	Text. No code.
Figure A.12	Comparison between expert's versus ChatGPT's in General Business Functions	6.appendix_figures.do
Figure A.13	Comparison between expert's versus ChatGPT's in Agriculture	6.appendix_figures.do
Figure A.14	Comparison between expert's versus ChatGPT's in Food Processing	6.appendix_figures.do
Figure A.15	Comparison between expert's versus ChatGPT's in Apparel	6.appendix_figures.do
Figure A.16	Comparison between expert's versus ChatGPT's in Retail	6.appendix_figures.do
Figure B.1	Example question for the presence of technologies and most-used technology	6.appendix_figures.do
Figure B.2	Question for Extensive and Intensive Margins of Tech. in Body Pressing	Text. No code.
Figure C.1	Average MAX and MOST GAP by Class of Business Function	6.appendix_figures.do
Figure C.2	Within-country dispersion in MAX-MOST Gaps	6.appendix_figures.do

Note: The source of information for Figures 1 and 2 is the FAT survey questionnaire.

Tables (Appendix)

Number	Title	Responsible
Table A.1	Coverage of Firm-Level Technology Surveys	Text. No code.
Table A.2	Technology Sophistication in Sourcing	Text. No code.
Table A.3	Technology Sophistication in Marketing	Text. No code.
Table A.4	Technology Sophistication in Sales	Text. No code.
Table A.5	Technology Sophistication in Payment	Text. No code.
Table A.6	Technology Sophistication in Quality Control	Text. No code.
Table A.7	Technology sophistication in Agriculture - Weeding	Text. No code.
Table A.8	Technology Sophistication in Agriculture - Harvesting	Text. No code.
Table A.9	Technology Sophistication in Apparel - Design	Text. No code.
Table A.10	Technology Sophistication in Apparel - Sewing	Text. No code.
Table A.11	Technology sophistication in Food Processing - Anti-Bacterial	Text. No code.
Table A.12	Technology Sophistication in Food Processing - Packaging	Text. No code.
Table A.13	Technology Sophistication in Retail - Pricing	Text. No code.
Table A.14	Technology Sophistication in Retail - Inventory	Text. No code.
Table A.15	Business Functions with Independent Measures of Relative Productivity	Text. No code.
Table A.16	Relative productivity in Business Administration - GBF	Text. No code.
Table A.17	Relative productivity in Agriculture	Text. No code.
Table A.18	Relative productivity in Retail	Text. No code.
Table A.19	Relative productivity in Banking	Text. No code.
Table A.20	Relative productivity in Apparel	Text. No code.
Table A.21	Relative productivity in Health	Text. No code.

Table A.22	Sampling frame by country	Text. No code.
Table A.23	Total number of firms in the universe covered by the survey	4.appendix_tables.do
Table A.24	Year and mode of data collection	Text. No code.
Table A.25	Response rates (by country)	Text. No code.
Table A.26	Comparison of establishment size between respondents vs non-respondents	4.appendix_tables.do
Table A.27	Comparison of technology sophistication between high and low numbers of attempts	4.appendix_tables.do
Table A.28	Comparison of technology sophistication between original and replacement samples	4.appendix_tables.do
Table A.29	Analysis of enumerator bias distribution	4.appendix_tables.do
Table A.30	Difference in technology sophistication in general business functions with and without outlying enumerators	4.appendix_tables.do
Table A.31	Relationship between FAT survey variables and the log of wages from administrative for Brazil	4.appendix_tables.do
Table A.32	Comparison between FAT sample and RAIS data (universe)	4.appendix_tables.do
Table B.2	Adjustment of Ranks in Automotive BFs	Text. No code.
Table C.1*	Summary Statistics	4.appendix_tables.do
Table C.2	Average level of technology measures	4.appendix_tables.do
Table C.3	Frequency of Sophistication Gaps by Business Function	4.appendix_tables.do
Table C.4*	Productivity and Technology Sophistication: Linearity	4.appendix_tables.do
Table C.5	Average GAP, MAX, and MOST along firm's age and size groups	4.appendix_tables.do
Table C.6*	Correlates of MAX-MOST Gaps across establishments by size and country income	4.appendix_tables.do
Table C.7*	Establishment Performance and Technology Sophistication	4.appendix_tables.do
Table C.8*	Establishment Performance and Technology Sophistication across Sectors	4.appendix_tables.do
Table C.9*	Establishment Performance and Technology Sophistication across Countries	4.appendix_tables.do

Note: (*) These tables use indicators from Poland for which data is not included in this package. All results, excluding the restricted variables from Poland, can be replicated using the package.

The source of information for Figures and Tables described as "Text. No code." is provided in the text. These exhibits do not use the FAT dataset.

5. Citation of Data & Code

- Users must cite the original paper when using or adapting this replication package.

If you use or base your work on this replication package, please cite it as:

Cirera, Xavier; Comin, Diego, and Cruz, Marcio. "Technology Sophistication Across Establishments," Quarterly Journal of Economics, Forthcoming, Replication package code and data, [DOI or persistent link to Dataverse].

Also cite any original data sources listed in Section 4 according to their recommended citation formats.

Thank you for your interest in replicating our work

Acknowledgements:

The preparation of the codes and replication package had significant contributions from Harneet Singh, Kyung Min Lee, Aman Mahajan, Charmaine Crisostomo, Yuheng Ding, and Magda Malec.

[End of README]