Replication for "Management and Misallocation in Mexico" by Nicholas Bloom, Leonardo Iacovone, Mariana Pereira-Lopez, and John Van Reenen

Overview

This replication package constructs the analysis file using data from the 2015 National Survey on Productivity and Competitiveness of Micro, Small, and Medium-sized Enterprises (ENAPROCE), Mexico's first large-scale representative management survey. A second wave, conducted in 2018, targeted the same sample. We also incorporate data from the 2014 Economic Census to test the robustness of our estimates and calculate export shares at the 6-digit NAICS level.

For the analysis of frictions, we use additional external data, including population estimates from the 2010 Population and Housing Census, combined with household income and expenditure data from the National Survey of Household Income and Expenses (ENIGH), using Small Area Estimates. We also use estimated drive times to the U.S. border, derived from OpenStreetMap data.

Graphs comparing Mexico with U.S. data from the Management and Organizational Practices Survey (MOPS) were created using Plot Digitizer to extract information from graphs in Bloom et al. (2019), as we did not have direct access to the original U.S. MOPS data.

The code is written in Stata. A master script (Master.do) runs the code to generate 12 main figures and 4 tables in the paper, along with 14 tables and 11 figures in the appendix.[1] The replication process takes approximately 90 minutes using the original data and 5 minutes using the mock data provided for testing. All mock data files have the prefix ejem_ in their names.

Data Availability and Provenance Statements

Data Accessibility Statement

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.

Data sources

This paper relies on two main data sources that are **confidential under Articles 37 and 38** of the **Mexican Law on the National System of Statistical and Geographical Information**. According to this law, these datasets can only be accessed within the **Microdata Lab** at the **National Statistical Office**. Results—including descriptive statistics, graphs, and estimations—can only be

[1] The Appendix Includes 13 Figures but the first two (A1 and A2) are generated manually.

released if the statistical authorities overseeing each project determine that they comply with the confidentiality principles established in the law. Therefore, the microdata used in this study **cannot be made publicly available**.[2]

Additionally, as our team contributed to the **survey design**, we had access to **additional variables** related to **data collection monitoring** and **explicitly requested identifiers** for matching with other statistical projects.

1. ***ENAPROCE data:*** Information about the two waves of this survey, including some aggregate descriptive statistics, can be found at https://www.inegi.org.mx/programas/enaproce/2015/ and https://www.inegi.org.mx/programas/enaproce/2018/ , respectively.
2. ***Economic Census:*** Information about the Economic Census for 2014, which was used given that it was the sample frame for the survey, along with the most recent Economic Censuses can be found at https://www.inegi.org.mx/programas/ce/2014/

Summary of Availability

■ **No data can be made** publicly available.

The microdata for this project is used in compliance with the **Mexican Law on the National System of Statistical and Geographical Information**, strictly under **authorization** and within the premises of the **National Statistical Office's Microdata Lab**.

Details on each Data Source

Table 1. Data sources

Data.Name	Data.Files	Citation
"ENAPROCE"	***SMEs and large firms*** ejem_base_pyme_ciega_2015.dta ejem_base_pyme_ciega_2018.dta ***Microenterprises*** ejem_base_micro_ciega2015.dta ejem_base_micro_ciega2018.dta ***Survey operations info*** ejem_operation_info2015	**ENAPROCE (2015) and ENAPROCE (2018)**
"Economic Census"	ejem_CE_2014.dta ejem_CE_2019.dta	**Economic Census (2014)**

[2] Log files are not among the types of files that can be submitted for clearance.

Data.Name	Data.Files	Citation
		Economic Census (2019)
"Municipality data"	Municipality_data_all.dta	2010 National Population and Household Census, INEGI. Municipality level data based on various sources documented in Appendix B. Additional datasets
"Tradability and KIBS" (NAICS 6 digits dataset)	Tradability and KIBS.dta	KIBS defined according to **Eurostat** and Tradability according to **Nayyar et al. (2021)**
"Share exports clase" (NAICS 6 digits dataset)	Share exports clase.dta	Share of exports at the NAICS 6 digits level using data from the **Economic Census (2014)**
"Data extracted from graphs in Bloom et al. (2019)	Age_bins_US.dta Data_US_bins.dta	Data on age bins and management score bins, estimated using the graphs provided in **Bloom et al. (2019).** No access to the underlying data.
Maps data state level	SHAPEFILES Entidades_2010_5A.dbf Entidades_2010_5A.prj Entidades_2010_5A.sbn Entidades_2010_5A.sbx Entidades_2010_5A.shp Entidades_2010_5A.shx	https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825292805
"Performance measures at the decile level (*for Figure 1*)"	Management and performance.dta	Not applicable
"Management score densities Mexico and U.S. (*for Figure 2*)"	Densities MX vs US.dta	Not applicable

Data.Name	Data.Files	Citation
"Data binscatter size management U.S. and Mexico (*for Figure 3*)"	Data Binscatter US.dta Binscatters size management MX vs US.dta	Not applicable
"Age bins and mean and sd of management MX (*For Figure 4 and 5*)"	Age bin mean sd manufacturing.dta	Not applicable
"Data management practices by state (*for Figure 6*)"	Map management state.dta	Not applicable
"Data vingtiles of management practices ENAPROCE 2015 vs. ENAPROCE 2018 (*for Appendix Figure A3*)"	Vingtiles management 2014 2017.dta	Not applicable
"Data management distributions by sectors comparing 2014 and 2017 *(for Appendix Figures A4 and A5*)"	Datos histograma updated 2019.dta	Not applicable
"Data comparing employment and firm distribution of the Economic Census and ENAPROCE 2015(*For Appendix Figure A6*)"	Distributions overlapping sample.dta	Not applicable
"Data age bins and mean and sd of TFP and Value added (*for Appendix Figures A7-A10*)'"	Data age TFP Value added.dta	Not applicable

Description of folder contents

This folder structure of this package is as follows:

- **Dofiles:** Holds the Stata code to produce all the different stages of the analysis.
- **data:** Holds example datasets, mostly mock data with expanded observations based on examples provided for coding by the Statistical Office (INEGI).
- **working:** Stores the datasets produced for estimations and graphs. It currently includes the intermediate data cleared to produce graphs, but the folder working/Temp includes the datasets assembled for estimations and graphs.
- **Outputs: Stores all the results from the analysis, including tables and figures.**

- **Outputs/figures:** Includes real figures produced from intermediate data cleared by INEGI that are included in the paper.
- **Outputs/figures/mock:** Includes the figures produced based on the mock data provided just for coding purposes.
- **Outputs/tables:** Includes all the tables produced in the analysis based on the mock data provided.
- **Outputs/appendix:** Includes all the tables and figures in the Online Appendix
    - **Outputs/appendix/figures:** Includes real figures produced from intermediate data cleared by INEGI that are included in the paper.
    - **Outputs/appendix/figures/mock:** Includes the figures produced based on the mock data provided just for coding purposes.
    - **Outputs/appendix/tables/mock:** Includes all the tables produced for the Online Appendix based on the mock data provided.

Computational requirements

Software requirements

Stata is used in the analysis of this project. This project was written using Stata 18 and has not been tested on older versions.

Memory and Runtime Requirements

Summary

Approximate time required to reproduce the analyses on a standard (CURRENT YEAR) desktop machine: 90 minutes if starting from the raw data (not provided for confidentiality). Approximately 5 minutes using the mock data provided to test the replication package.

Details

The code was last run on a laptop with an **11th Gen Intel(R) Core(TM) i5-1145G7 @ 2.60GHz (1.50 GHz), 16GB RAM,** and **Windows 10 Enterprise.** Computation using the mock data took **10 minutes.**

The original code was executed at **INEGI's Microdata Lab** on a **virtual desktop** with an additional **32GB of allocated memory**, taking **90 minutes** to run the full code.

Description of programs/code

-Program **Data prepare 2015 mock data.do and Data prepare 2018 mock data.do** organizes and clean the confidential data from ENAPROCE 2015 and 2018.

-Program **Data assembly and merge.do**, merges the data from ENAPROCE 2015 and ENAPROCE 2018, as well as additional data at the municipality and sectoral level and produces the long and wide versions of the dataset that is used for estimations.

-Program **Regressions.do** estimates all the regressions included in the main paper based on the long dataset including clean data from ENAPROCE 2015 and ENAPROCE 2018. The

current version relies on mock data provided for coding purposes and the results, of course, differ from those based on the original datasets.

-Program **Figures.do** generates all the figures included in the paper.

-Program **Municipality variables.do** generates municipality level variables to be merged with the firm-level datasets.

-Program **Microenterprisesdata.do** builds the dataset for microenterprises (not included in the main analysis), to test the robustness of the results for firms of this size.

-Program **Appendix Figures.do** generates all the figures included in the Online Appendix of the Paper.

-Program **Appendix Tables.do** generates all the tables included in the Online Appendix of the paper.

-Program **Appendix Table A4.do** generates Appendix Table A4 and is called by the Appendix Tables program.

Instructions to Replicators

1. Open Stata
2. Open the Master do file.
3. Set the working directory (line 33 of the code).
4. Run the Master do file in Stata (to create all the tables and figures). This is the only do-file you should need to run.

All required commands are installed before running the rest of the Master file. It checks each command and installs it if missing. This prevents errors and ensures replicability across different machines. If something pops up that you need to download, please download it.

List of tables and programs

The provided code reproduces:

- ■ All numbers provided in text in the paper
- ■ All tables and figures in the paper
- ☐ Selected tables and figures in the paper, as explained and justified below.

Table 2. Index of code to reproduce tables and figures (main paper)

Figure/Table # Object	Program	Line number	Output file	Notes
Table 1	Regressions.do	24	Table1.tex Table1.csv	MOCK Table
Table 2	Regressions.do	116	Table2.tex Table2.csv	MOCK Table
Table 3	Regressions.do	203	Table3.tex Table3.csv	MOCK Table
Table 4	Regressions.do	285	Table4.tex Table4.csv	MOCK Table
Figure 1	Figures.do	11	Figure_1.gph Figure_1.jpg	Real Figure based on intermediate data
Figure 2	Figures.do	110	Figure_2.gph Figure_2.jpg Figure_2.png	Real Figure based on intermediate data
Figure 3	Figures.do	140	Figure_3.gph Figure_3.jpg Figure_3.png	Real Figure based on intermediate data
Figure 4	Figures.do	169	Figure_4a.gph Figure_4a.jpg Figure_4a.png Figure_4b.gph Figure_4b.jpg Figure_4b.png Figure_4c.gph Figure_4c.jpg Figure_4c.png	Real Figure based on intermediate data
Figure 5	Figures.do	218	Figure_5a.gph Figure_5a.jpg Figure_5a.png Figure_5b.gph Figure_5b.jpg Figure_5b.png Figure_5c.gph Figure_5c.jpg Figure_5c.png	Real Figure based on intermediate data
Figure 6	Figures.do	268	Figure_6a.gph Figure_6a.jpg Figure_6a.png Figure_6b.gph Figure_6b.jpg Figure_6b.png	Real Figure based on intermediate data

Figure/Table # Object	Program	Line number	Output file	Notes
Figure 7	Figures.do	354	Figure_7a.gph Figure_7a.jpg Figure_7a.png Figure_7b.gph Figure_7b.jpg Figure_7b.png	MOCK Figure
Figure 8	Figures.do	400	Figure_8a.gph Figure_8a.jpg Figure_8a.png Figure_8b.gph Figure_8b.jpg Figure_8b.png	MOCK Figure
Figure 9	Figures.do	438	Figure_9a.gph Figure_9a.jpg Figure_9a.png Figure_9b.gph Figure_9b.jpg Figure_9b.png	MOCK Figure
Figure 10	Figures.do	484	Figure_10a.gph Figure_10a.jpg Figure_10a.png Figure_10b.gph Figure_10b.jpg Figure_10b.png	MOCK Figure
Figure 11	Figures.do	521	Figure_11a.gph Figure_11a.jpg Figure_11a.png Figure_11b.gph Figure_11b.jpg Figure_11b.png Figure_11c.gph Figure_11c.jpg Figure_11c.png Figure_11d.gph Figure_11d.jpg Figure_11d.png	MOCK Figure
Figure 12	Figures.do	605	Figure_12a.gph Figure_12a.jpg Figure_12a.png Figure_12b.gph Figure_12b.jpg Figure_12b.png Figure_12c.gph Figure_12c.jpg Figure_12c.png Figure_12d.gph Figure_12d.jpg Figure_12d.png	MOCK Figure

Table 3. Index of code to reproduce Online Appendix tables and figures

Figure/Table # Object	Program	Line number	Output file	Notes
Table A1	Appendix Tables.do	9	TableA1.xlsx TableA1p1.tex TableA1p2.tex TableA1p3.tex TableA1p4.tex	MOCK Table
Table A2	Appendix Tables.do	268	TableA2.tex TableA2.csv	MOCK Table
Table A3	Appendix Tables.do	339	TableA3.tex TableA3.csv	MOCK Table
Table A4	Appendix Tables.do Appendix tables A4	398	TableA4PanelA1Man.xls TableA4PanelA1Man.tex TableA4PanelA1Ser.xls TableA4PanelA1Ser.tex TableA4PanelA2Man.xls TableA4PanelA2Man.tex TableA4PanelA2Ser.xls TableA4PanelA2Ser.tex TableA4PanelB1Man.xls TableA4PanelB1Man.tex TableA4PanelB1Ser.xls TableA4PanelB1Ser.tex TableA4PanelB2Man.xls TableA4PanelB2Man.tex TableA4PanelB2Ser.xls TableA4PanelB2Ser.tex combinedPanelA.xls combinedPanelB.xls **combinedTablefull.xls**	MOCK Table Combines all panels into 1 file
Table A5	Appendix Tables.do	404	TableA5.tex TableA5.xls	MOCK Table
Table A6	Appendix Tables.do	527	TableA6.tex TableA6.csv	MOCK Table
Table A7	Appendix Tables.do	626	TableA7.tex TableA7.csv	MOCK Table
Table A8	Appendix Tables.do	709	TableA8.tex TableA8.csv	MOCK Table
Table A9	Appendix Tables.do	815	TableA9.tex TableA9.csv	MOCK Table
Table A10	Appendix Tables.do	919	TableA10.tex TableA10.csv	MOCK Table
Table A11	Appendix Tables.do	965	TableA11.tex TableA11.csv	MOCK Table
Table A12	Appendix Tables.do	1075	TableA12.tex TableA12.csv	MOCK Table

Figure/Table # Object	Program	Line number	Output file	Notes
Table A13	Appendix Tables.do	1680	TableA13p1.tex TableA13p1.csv TableA13p2.tex TableA13p2.csv	MOCK Table
Table A14	Appendix Tables.do	1840	TableA14.tex TableA14.csv	MOCK Table
Figure A1	Not applicable (manual-screenshot)			
Figure A2	Not applicable (manual)			
Figure A3	Appendix Figures.do	15	Figure_A3.gph Figure_A3.jpg	Real Figure based on intermediate data
Figure A4	Appendix Figures.do	38	Figure_A4.gph Figure_A4.jpg	Real Figure based on intermediate data
Figure A5	Appendix Figures.do	38	Figure_A5.gph Figure_A5.jpg	Real Figure based on intermediate data
Figure A6	Appendix Figures.do	72	Figure_A6_a.gph Figure_A6_a.jpg Figure_A6_b.gph Figure_A6_b.jpg	Real Figure based on intermediate data
Figure A7	Appendix Figures.do	101	Figure_A7_a.gph Figure_A7_a.jpg Figure_A7_b.gph Figure_A7_b.jpg	Real Figure based on intermediate data
Figure A8	Appendix Figures.do	118	Figure_A8_a.gph Figure_A8_a.jpg Figure_A8_b.gph Figure_A8_b.jpg	Real Figure based on intermediate data
Figure A9	Appendix Figures.do	131	Figure_A9_a.gph Figure_A9_a.jpg Figure_A9_b.gph Figure_A9_b.jpg	Real Figure based on intermediate data
Figure A10	Appendix Figures.do	145	Figure_A10_a.gph Figure_A10_a.jpg Figure_A10_b.gph Figure_A10_b.jpg	Real Figure based on intermediate data
Figure A11	Appendix Figures.do	160	Figure_A11_a.gph Figure_A11_a.jpg	MOCK Figure

Figure/Table # Object	Program	Line number	Output file	Notes
			Figure_A11_b.gph Figure_A11_b.jpg	
Figure A12	Appendix Figures.do	207	Figure_A12_a.gph Figure_A12_a.jpg Figure_A12_b.gph Figure_A12_b.jpg	MOCK Figure
Figure A13	Appendix Figures.do	251	Figure_A13_a.gph Figure_A13_a.jpg Figure_A13_b.gph Figure_A13_b.jpg Figure_A13_c.gph Figure_A13_c.jpg Figure_A13_d.gph Figure_A13_d.jpg	MOCK Figure

An auxiliary TEX file was created in Overleaf to visualize the tables. However, it is important to note that not all tables are produced correctly due to the limitations of the mock dataset, which includes many repeated observations. As a result, some regressions omit variables that are included in the real estimations.

The above-mentioned mock tables and figures were generated using mock data due to data restrictions. The original outputs for these exhibits are provided in the "Original results statistical office" folder.

References

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I., and Van Reenen, J. (2019). What drives differences in management practices? *American Economic Review*, *109*(5), 1648-83.

Eurostat. (2019). *High-tech employment: Annex 8* (HTEC_ESMS). Eurostat. https://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an_8.pdf

Instituto Nacional de Estadística y Geografía (INEGI). (2015). Encuesta Nacional sobre Productividad y Competitividad de las Micro, Pequeñas y Medianas Empresas 2015 (ENAPROCE). INEGI. https://www.inegi.org.mx/programas/enaproce/2015/

Instituto Nacional de Estadística y Geografía (INEGI). (2018). Encuesta Nacional sobre Productividad y Competitividad de las Micro, Pequeñas y Medianas Empresas 2018 (ENAPROCE). INEGI. https://www.inegi.org.mx/programas/enaproce/2018/

Instituto Nacional de Estadística y Geografía (INEGI). (2014). Censo Económico 2014. INEGI. https://www.inegi.org.mx/programas/ce/2014/

Instituto Nacional de Estadística y Geografía (INEGI). (2019). Censo Económico 2019. INEGI. https://www.inegi.org.mx/programas/ce/2019/

Nayyar, G., Hallward-Driemeier, M., & Davies, E. (2021). *At Your Service?: The Promise of Services-Led Development.* Washington, DC: World Bank. http://hdl.handle.net/10986/35599 License: CC BY 3.0 IGO.