

## **Replication material for “The Size and Life-Cycle Growth of Plants: The Role of Productivity, Demand and Wedges” – Marcela Eslava, John Haltiwanger, Nicolás Urdaneta**

This folder contains all Stata codes and public domain datasets necessary to fully replicate the results. The empirical analysis in this paper uses proprietary establishment level data from a database constructed by the authors using information from the Colombian Annual Manufacturing Survey (AMS) and the Technological Development and Innovation Survey (EDIT) (DANE, 1982-2013 and 2003-2012 respectively).

### Data Availability and Provenance Statement

The AMS and EDIT are housed at the Colombian Departamento Administrativo Nacional de Estadística (DANE). For approved projects and for statistical purposes only, the plant-level and plant-product level data used for this project can be accessed at DANE. For access to the data, interested users must contact the Dirección de Metodología and Producción Estadística. Instructions are available [here](#). There is a publicly available version of the plant-level (but not the product-level) dataset from 1992 onwards that can be obtained from DANE’s website ([here](#) and [here](#)), with a few transformations relative to the original data, aimed at ensuring confidentiality.

### Software Requirements and Runtime

The programs require Stata. The code was run on Stata MP version 17.0. The codes should run in the computer server provided by DANE. The processing of all results, including processing the datasets, takes approximately one to two days.

### Description of Code and Data

This folder contains the codes to construct the dataset used in the project from the raw files as provided by DANE and generate the tables and figures in the paper. The folder “Public datasets” contains the external datasets that are required to construct the project’s dataset (PPI, CPI, crosswalks, etc.). For replication, follow these instructions:

1. Request access to the AMS (1982-2013) and EDIT (2003-2014) microdata from DANE: *Encuesta Anual Manufacturera* and *Encuesta de Desarrollo e Innovación en Tecnología* in Spanish. Make sure to request the EDIT for manufacturing only (there is also one for services, not used in this paper).
2. Unzip the package to the user assigned to you by DANE in the computer system in the folder that you desire, let’s call it “main folder”. This zip folder contains the subfolder structure necessary to replicate the results of the paper.
3. Setup the raw data files provided by DANE in folders (and subfolders) with the following names to create the required folder structure:
  - AMS files with the following folders within your “main folder”:

- “Primary Databases/Eam/Productos y materiales”: include each year’s Stata files in folders named with the form: “yAAAA” where AAAA corresponds to the year (e.g., “y2005”). The files should include the plant-level data and the product-level data for the years 1992-2013.
  - “Primary Databases/Eam/PLANOS EAM 1982\_1991”. This folder must include the 1982-1991 plant-level and product-level datasets provided by DANE as well as the product-level datasets (in .txt format).
  - “Primary Databases/Eam/PLANOS EAM 1982\_1991/id correlatives”. This folder must include the identifier crosswalks for the AMS provided by DANE.
  - Innovation survey in the following folder: “Primary Databases/EDIT INDUSTRIA/Planos”
4. Setup the publicly available datasets in the following folders:
- “ciiu”. This data file contains a crosswalk between 4-digit ISIC rev3 to ISIC rev2. Place the file in “Primary Databases/Other/correlativas”.
  - “CIIU3D”. This data file contains a crosswalk to switch from ISIC rev2 (3 digits) to ISIC rev3 (2 digits). It is needed to construct the dataset but not needed for the project. Place the file in “Primary Databases/Other/correlativas”.
  - “Correlativa\_CPC\_CIIU”. Crosswalk between CPC rev 1 and CIIU rev 2 product codes. Place the file in “Primary Databases/Other/correlativas”.
  - “deflactores\_ciiu\_3d\_r2”. Price deflators from PPI, CPI and Capital Formation PPI, only available at 3-digit sector from 1990 to 2014. Downloaded from *Banco de la República* website in 2015 (link [here](#)). Place the file in “Primary Databases/Other”.
  - “ipc\_1954\_2014”. CPI dataset by month. Downloaded from *Banco de la República* website in 2015 (link [here](#)). Place the file in “Primary Databases/Other”.
  - “ipp\_2d\_1970\_2014”. PPI dataset for 2digit sectors. Downloaded from *Banco de la República* website in 2015 (link [here](#)). Place the file in “Primary Databases/Other”.
  - “ipp\_uso\_destino\_econ\_1970\_2014”. PPI dataset by uses and destinations. Downloaded from *Banco de la República* website in 2015 (link [here](#)). Place the file in “Primary Databases/Other”.
5. Request DANE to install the following Stata packages: mmerge, estout, grc1leg, ivreg2, outreg2, ftools, parmes, quantiles, reghdfe, winsor2, scheme\_tufte, blindschemes, ranktest, unique.
6. Run the master file (or only file if the subfolder contains a single one) in each of the program subfolders in the order setup below. The first four folders prepare the EAM dataset (folders 1 and 2), the innovation survey (folder 3), and the plant level output

and input price indices (folder 4). The last folder sets up the dataset used specifically for the paper taking as inputs the datasets created in the previous folders and produces all tables and figures. In each master file, you will need to change the file directories to those which contain the raw data in your system username and where you will save processed data and results.

1. EAM 1982-2013
2. EAM processed dataset
3. EDIT
4. Plant level price indices
5. Size and Life Cycle Growth

Here is a description of the Stata globals and file directories that must be changed in each master or dofile:

- 1. EAM 1982-2013. This folder contains a master code that runs all the other necessary codes. You only need to adjust the file directories so that they correspond to the following:

<b>File directory</b>	<b>Target folder of file directory</b>
prog_dir	Programs/EAM 1982-2013/programas
save_db	Primary Databases/Eam/EAM_ME_82_13
logfile_dir	Primary Databases/Eam/EAM_ME_82_13/logfiles
graph_dir	Primary Databases/Eam/EAM_ME_82_13/graphs
eam	Primary Databases/Eam/
source_path	Primary Databases/Eam/PLANOS EAM 1982_1991/id_correlatives
cd1	Primary Databases/Other
correl_processing	Primary Databases/Eam/EAM_ME_82_13/Correl_processing
cd (line 29)	Primary Databases/Eam/Productos y materiales

Some important notes: i) Check that the files provided do not have the two last digits of the year in their names (i.e., “producto12.dta” for 2012 should only be “product.dta”). Edit the raw file names if needed. ii) Use the code “0.2 SAS to Stata.do” if needed to convert SAS files provided by DANE to Stata. iii) Run the code “0.3 Separate Products Materials 1982-1991” if in the files provided you have a file named “PROD” but not one named “MATE” (the file will contain both material input and product data that should be separated). For both programs you will need to adjust the file directories as well.

- 2. EAM processed dataset: This folder contains a single dofile. In this case, it is only necessary to change the following directories (with the line in the code in parenthesis):

<b>File directory</b>	<b>Target folder of file directory</b>
sale	Processed Datasets/
db	Primary Databases/Eam/EAM_ME_82_13

- 3. EDIT. This folder contains a single dofile. In this case, it is only necessary to change the following directories (with the line in the code in parenthesis):

<b>File directory</b>	<b>Target folder of file directory</b>
in	Primary Databases/EDIT INDUSTRIA/Planos
out	Primary Databases/EDIT INDUSTRIA/EDIT_I_VII
eam	Primary Databases/Eam/EAM_ME_82_13

- 4. Plant level price indices. This folder contains a master do file that runs all other programs in the folder. The programs in the “outputs” (“inputs”) subfolder calculate the price indices for outputs (inputs). You only need to adjust the file directories in the master so that they correspond to the following:

<b>File directory</b>	<b>Target folder of file directory</b>
do	Programs/Plant level price indices
cd	Processed Datasets/Plant level index price
db_route	Primary Databases/Eam/EAM_ME_82_13
other	Primary Databases/Other
initial_db	Processed Datasets
results	Processed Datasets/Growth decomposition/Results
db	Processed Datasets/Growth decomposition/Results/db
cd (line 31)	Primary Databases/Eam/Productos y materiales

- 5. Size and Life Cycle Growth. This folder contains a master dofile that runs all other programs in the folder. You only need to adjust the file directories in the master so that they correspond to the following:

<b>File directory</b>	<b>Target folder of file directory</b>
initial_db	Processed Datasets
db	Processed Datasets/Growth decomposition/Results/db
Growth_db	Processed Datasets/Growth decomposition/db
price_indices	Processed Datasets/Plant level index price
EAM_panel	Primary Databases/Eam/EAM_ME_82_13
programs	Programs/Size and Life Cycle Growth
outputs	Primary Databases/Eam/Productos y materiales
other	Primary Databases/Other/correlativas
temp	Processed Datasets/Growth decomposition/db/Temp
cd_Pind_out	Processed Datasets/Plant level index price/temp_database_outp
cd_Pind_inp	Processed Datasets/Plant level index price/temp_database
edit_db	Primary Databases/EDIT INDUSTRIA/EDIT_I_VII
cd (line 37)	Results

7. After running all files, the output produced for the paper includes the following:

- Results Restud.xlsx. This excel file contains all tables presented in the paper and additional numbers mentioned in the text.
  - Results Figures Restud.xlsx. This excel contains all numbers that plot the figures in the paper.
  - Results Restud Additional.xlsx. This excel contains additional results that are no longer present in the text or appendix.
  - Set of figures in gph and pdf formats (Fig1, Fig2, ..., FigI1)
8. List of tables, figures, and results mentioned in text by dofiles (in order of appearance). All outputs in the paper are produced in the codes found in the “Size and Life Cycle Growth” folder.

Table/Figure	Program
Table 1	1. Production Demand function estimation.do
Table 2	3. Correlations table.do
Figure 1	4. Contribution of fundamentals.do
Figure 2	4. Contribution of fundamentals.do
Table 3	5. Main decompositions.do
Table 4	6. Quality adjusted wages table.do
Table 5	7. Allocative efficiency.do
Table 6	6. Quality adjusted wages table.do
Figure 3	7. Allocative efficiency.do
Figure 4	8. Parameter distributions.do
Table 7	9. Alternative parameters decomposition.do + 10. Allocative efficiency alternative parameters.do
Table C1	1. Production Demand function estimation.do
Table C2	15. Additional statistics.do
Table C3	15. Additional statistics.do
Table C4	15. Additional statistics.do
Table C5	15. Additional statistics.do
Table E1	11. Correlates on observables.do
Table E2	12. Persistence and innovation cycle.do
Figure H1	13. Exiters and Continuers.do
Figure H2	13. Exiters and Continuers.do
Table H1	13. Exiters and Continuers.do
Table I1	14. Distortion adjusted user cost of capital.do
Figure I1	14. Distortion adjusted user cost of capital.do
Table L1	7. Allocative efficiency.do

There are a few additional results mentioned in the text of the paper that can be found in the following programs (the rest of numeric results mentioned can be found in their corresponding tables):

#### 1. Introduction.

- “Plants in the bottom(top) productivity quartile are 42% larger (24% smaller) than efficient”. These results are produced in *4. Contribution of fundamentals.do*. They are mentioned in section 5.2.
- “Efficient (i.e., in the absence of composite wedges) aggregate productivity is 158% larger than it would be in the absence of dispersion in quality-adjusted productivity.”. This estimate is produced in *7. Allocative efficiency.do*.

#### 3. Data.

- “Over 90% of plants in the AMS (i.e., over 90% manufacturing plants in Colombia with size over the inclusion threshold) belong to single-plant firms, so that the

distinction between plants and firms is not as crucial in our context as it is in others.”. This result can be found in *15. Additional statistics.do*.

- “There is exit in our sample, at a rate of approximately 7% per year.”. This result can be found in *15. Additional statistics.do*.

## 5.2. Plant attributes vs. size

- This section provides some main results behind Figures 1 and 2. All numbers are calculated as well in program *4. Contribution of fundamentals.do*.

## References

DANE. 1982-2013. “Encuesta Anual Manufacturera, EAM (AMS).” Departamento Administrativo Nacional de Estadística, DANE.

DANE. 2003-2013. “Encuesta de Desarrollo e Innovación Tecnológica, (EDIT).” Departamento Administrativo Nacional de Estadística, DANE.