Replication Files Guide for "Testing the Production Approach to Markup Estimation"

Devesh Raval

Federal Trade Commission

devesh.raval@gmail.com

October 16, 2022

I divide the replication files into three components. In Section 1, labeled "Data Cleaning Files", I include programs that construct datasets for the main analysis. In Section 2, labeled "Estimation Files", I include all programs that use these datasets to estimate markups. In Section 3, labeled "Analysis Files", I incljude all programs that evaluate the properties of these markup estimates. Finally, in Section 4, labeled "Supplemental Files", I include the results of supplemental analysis such as Monte Carlo simulations.

1 Data Cleaning Files

The paper uses 7 datasets. For each dataset, I detail how to obtain the dataset as well as programs to clean the raw data. Please see Appendix A, and Appendix G in the Web Appendix, for data notes for these datasets. All of these files are in the `Data Cleaning`

subfolder.

1.1 Chile

All of the files to construct the Chile dataset are in the `Chile` subfolder. For Chile, I rely on an original dataset constructed by David Greenstreet, which is included in the subfolder. The `documentation` subfolder contains Greenstreet's data notes. In the `programs` subfolder, the program file `create_chile_data.do` cleans and creates variables before creating the Chile dataset.

1.2 Colombia

All of the files to construct the Colombia dataset are in the `Colombia` subfolder. In the `programs` subfolder, the program file `make_colombia_dataset_new.do` cleans and creates variables using the original data files in the `data` subfolder before creating the Colombia dataset.

1.3 India

All of the files to construct the India dataset are in the `India` subfolder. For India, I do not include the raw data; however, that data is available for purchase from the Government of India.

The raw files I obtained came in text format and would be located in the `India\01. Data\ASI\raw data\` folder by year; the file `file list.csv` in that folder contains the file list of all raw data files. I build on a replication package by Alcott, Collard-Wexler,

and O'Connell (2016), available at https://www.aeaweb.org/articles?id=10.1257/aer.20140389. In the 02. Programs subfolder, the programs dev_1e_Load and Merge Blocks of 1998-2014.do, dev_1f_Select fields and append 1998-2014.do, and dev_1g_Clean stacked dataset 1998-2014.do have to be run in succession to construct the dataset.

1.4 Indonesia

All of the files to construct the Indonesia dataset are in the Indonesia subfolder. For Indonesia, I do not include the raw data; however, that data is available for purchase from Statistik Indonesia.

The raw files I obtained came in DBF format, which would be located in the Raw Data\DBF folder. The file DBF file list.csv in the Raw Data folder contains the file list of all of the raw files. I first convert to Stata format using the R file inputRawDataIndonesia.R located in the Raw Data folder.

I then use the data_construction_new.do program, in the programs folder, to construct the dataset. This file calls several subprograms located in the same folder.

1.5 US

All of the files to construct the US dataset are in the US subfolder. For US, I do not include the raw data; however, that data is from Compustat and available from Wharton Research Data Services (WRDS).

I then use the load_data_us.R program, in the programs folder, to construct the dataset.

1.6 Southern Europe

All of the files to construct the Southern Europe dataset are in the ORBIS subfolder. For Southern Europe, I do not include the raw data; however, that data is from ORBIS and available from Bureau van Dijk. Using the ORBIS web interface, I created data files by country and 2 digit industry NACE code (so a sample file would be data_es_11.xlsx for Spain, NACE industry 11). Due to download limits from the ORBIS web interface, I sometimes have to chunk these raw files into multiple subfiles.

I then use the process_raw_data.R program and then create_deflators.R, in the programs folder, to construct the dataset by first processing the raw data and then merging in deflators.

1.7 Retailer

In order to preserve the confidentiality of the retailer, I do not provide either the raw data or data cleaning files for the retailer dataset. However, this data is available from the Federal Trade Commission for anyone who can arrange with the FTC to access the data.

2 Estimation Files

All of these files are in the Estimation Files subfolder.

For each dataset, I run a program that first creates a final set of variables before production function estimation, saves the resulting dataset, and then runs several programs to estimate markups. These programs are located in the programs subfolder; the programs

are estimateMarkupChile.do, estimateMarkupColombia.do, estimateMarkupIndia.do, estimateMarkupIndonesia.do, estimateMarkupUS.do, and estimateMarkupSoEurope.do. (An equivalent file for the Retailer is not included.)

All programs call estimateMarkupAll.do. This program trims the data by industry and then calls programs estimate_markup_acf_wt.do, estimate_markup_acf_agg_wt.do, and estimateCostShares.do to estimate production functions and markups for each estimate of output elasticities using the ACF control function estimator for a production function with capital, labor and materials; the ACF control function estimator for a production function estimator with capital and a composite variable input; and cost share based estimators.

For Chile, Colombia, India, and Indonesia, the program estimateMarkupAllEnergy.do is called next. It trims the data by industry and then calls the program estimate_markup_acf_en_wt.do to estimate production functions and markups for each estimate of output elasticities using the ACF control function estimator for a production function with capital, labor, raw materials, and energy.

For all datasets, estimateMarkupDP.do is called next. It trims the data by industry and then calls programs estimate_markup_dp_wt.do and estimate_markup_dp_agg_wt.do to estimate production functions and markups for each estimate of output elasticities using the dynamic panel estimator for a production function with capital, labor and materials; and the dynamic panel estimator for a production function estimator with capital and a composite variable input.

For all datasets, estimateMarkupFGT.do is called last. It trims the data by industry and then calls programs estimate_markup_fgt_wt.do and estimate_markup_fgt_agg_wt.do to estimate production functions and markups for each estimate of output elasticities using

the FGT control function estimator for a production function with capital, labor and materials; and the FGT control function estimator for a production function estimator with capital and a composite variable input.

In the data subfolder, I include the data files produced by these programs for the Chile and Colombia datasets.

3 Analysis Files

All of these files are in the Analysis Files subfolder.

The file doMarkupStats.do calls createStatisticsMarkup.do to estimate several different statistics (such as time series changes, dispersion, cross-sectional correlations, stylized facts) for each dataset for several differnt specifications. These programs are located in the programs subfolder, and create tex and csv files located in the results subfolder, with separate folders for each dataset. The equivalent file for the retailer for createStatisticsMarkup.do is createStatisticsMarkupRetailer.do.

The file create_tex_tables.R develops tex tables for the paper that combine results from the different datasets, which are then saved in a combined subfolder in the results subfolder.

The file make_graphs_new.R creates a datafile (all_time.csv, in the results\combined\ subfolder) combining the time series statistics for all of the datasets. The file make_graphs.do then creates time series plots with this file, as seen in the paper in Figures 2, 3, and 5-15.

The file chile_scatter.do creates plots using the Chilean markups as seen in Figure 1 and Figure 4 of the paper. All graphs are located in the results\graphs subfolder.

Several files implement the product-level estimation described in Appendices C.2 and C.3. First, `create_soleproduct_india.do` examines only plants that list a single product produced, and estimates markups at the product level for sole-product firms using cost shares. Second, `create_product_codes_india.do` creates a set of homogenous products (described in Appendix G.7) using Indian product data. `estimateMarkupProduct.do` then estimates markups for these homogenous products using ACF Cobb-Douglas and Translog estimators, calling `estimate_markup_prod_wt.do` and `estimate_markup_prod_agg_wt.do`.

3.1 Supplemental Files

All of these files are in the `Supplemental Files` subfolder.

The file `monte_carlo_panel.do` runs the Monte Carlo simulation described in Appendix F. `estimateMarkupMonteCarloPanel.do` estimates ACF control function methods to estimate markups on the Monte Carlo datasets, and `monte_carlo_results_acf.R` compiles markup statistics for these Monte Carlo simulations.