

Guide: Estimating International Trade Margin Shares by Mode of Transport for the GTAP Data Base

December 2018

In this guide we briefly present a description of the data used during our estimations, the code needed to obtain our estimations and figures, and the folders where they are contained.

1. Folders

Folder “codeForEstimation” contains the R code we use to estimate the modal shares and generate our document’s figures.

Folder “Figures and tables” contains the figures produced by the code in folder “codeForEstimation” and reported in the paper.

Folder “DATA” contains the source data, data generated by our R code files, and the final database of trade margin modal shares. This folder contains within itself the following folders:

- Folder: RawData – This folder contains the source data. The description of each file can be found in the R file that uses it, most of them are utilized by the R file load.r.
- Folder: my_data – This folder is used by the R file codes to store any “intermediate” data generated as a byproduct of our methodology.
- Folder: output – In this folder, the R file do.r stores the final trade margins by mode as generated by our methodology (OURSHARES.csv).

2. Code

Below, we briefly explain the tasks that each R program file performs. For more details, please see the program files themselves. To successfully execute the R code in these files, the following R packages are necessary: countrycode, foreign, reshape, dplyr, stringr, plyr, gdata, reshape2, tidyr, xtable, ggplot2, gridExtra, and brindrcpp. The R code files themselves are located in a folder named “codeForEstimation”. We describe the files in the other in which they ought to be executed.

- File: load.r – This file takes the ALADI, Eurostats, Census, and NATFD databases and prepares them to be worked with. Among other tasks, this file removes unnecessary variables; matches commodity-specific codes with their corresponding GTAP sectors; executes currency conversions, and ensures that all origin and destination countries are referred to with their ISO code. For the American data, specifically, in this file we obtain weight shipped by rail and road (see appendix A in the document). The output is one file for each: American, Latin American, and European data.
- File: clean.r – This file takes the American, Latin American, and European Data to create a single database where all the relevant computations will be executed with another code file. In clean.r, we create useful identifiers within the database (such as origin-destination, and origin-destination-sector identifying columns); incorporate bilateral distances, and merge all databases into a single base.

- File do.r – This file applies equations 4 to 7 and estimates the modal shares proper. This file generates a csv file named OURSHARES.csv (which will be placed in folder DATA/output). This is our final estimations. The variables in this file are:
 - odgm: Unique id. Origin-destination-sector-mode combination
 - o: Country of origin (exporter) ISO 3c code.
 - d: Country of destination (importer). ISO 3c code.
 - value_usd: Value in USD
 - weight_kg: Weight in kilograms
 - dist: Distance in thousands of kilometers.
 - g: GTAP sector code.
 - share: Our estimated modal share for the given Origin-destination-sector-mode combination.
 - wv_ratio: Weight value ratio as defined in the paper.
 - m: Mode of transportation (water, air, or other).
- File figures.r – This file generates the table and figures reported in the paper.