

Data and Code for Income Differences and Prices of Tradables: Insights from an Online Retailer

Ina Simonovska*

University of California, Davis and NBER

This Version: March 13, 2015

Abstract

This document describes the construction of the dataset and provides instructions to implement the code used to generate all empirical results in “Income Differences and Prices of Tradables: Insights from an Online Retailer.”

1 Introduction

In this document, I describe the construction of all variables used in the empirical analysis. Additionally, I outline the necessary steps to replicate all empirical results. The document has two parts. In the first part, I focus on the empirical analysis that uses data from an online retailer—Mango, while in the second, I focus on the results that involve data from the International Comparison Program—ICP.

2 Mango Data and Code

The code that reproduces the empirical results that pertain to Mango is contained in file `mango_code.do`. All variables necessary to run the code are contained in files `mango_data.dta`, `mango_sales_data.dta`, `mango_market_share.dta`. To see the description of a variable, consult the “label” column in the data file. The data source of each variable is described in the main body of the paper.

There are two types of variables: original and constructed. Original variables appear as reported in source data. Constructed variables are a combination of original variables. All constructed variables, with the exception of the ones described in the next subsection, are obtained by performing

*Contact: Department of Economics, University of California - Davis, One Shields Avenue, Davis, California 95616. Email: inasimonovska@ucdavis.edu.

basic algebraic manipulations, explained in the “label” column of the constructed variable, to the original variables. For example: consider original variable “gni_atl_2009” with label “GNI per capita, Atlas method (current US\$) 2009” in the data file “mango_data.dta”, which constitutes the fourth column of the same data file. This variable is identical to the corresponding variable in the World Development Indicators (WDI) database. Each row of this variable corresponds to the value in year 2009 for the country listed in the third column of the data file under variable “country_name”. Now consider the constructed variables “sp_gni_atl_2009” with label “GNI per capita, Atlas method (current US\$) 2009, Spain” and “log_gni_atl_2009” with label “GNI per capita, Atlas method (current US\$) 2009, relative to Spain, in logs”. The first is simply the original variable corresponding to the country Spain. This is a constructed variable because the observation for Spain appears in every row. The second variable is computed as follows: $\log_gni_atl_2009 = \ln(gni_atl_2009/sp_gni_atl_2009)$, corresponding to the description in the label column, which reads “GNI per capita, Atlas method (current US\$) 2009, relative to Spain, in logs”.

2.1 Constructed variables: log_icp and log_go

Two constructed variables in the data file “mango_data.dta” involve steps in addition to the above described rule. These variables are log_icp and log_go.

log_icp represents the log of the ratio of import share of country j from Spain to Spain’s domestic expenditure share, both in apparel and footwear industries only, in year 2010. To construct the import and expenditure shares in apparel and footwear, two pieces of data are needed. The first is trade flows in these goods and the second is expenditures on these goods. The data source for trade flows is described in the main text of the paper. File titled flows2010.txt contains these observations, which are publicly available. Expenditures in year 2010 are constructed using two variables. The first is the share of a country’s expenditures on apparel and footwear in year 2005 obtained from the ICP database. The second is the GDP in current USD in year 2010 obtained from the WDI database.

The ICP data are available to all researchers upon request from the World Bank. After obtaining the data, to compute the expenditure share of a country in year 2005, compute the sum of expenditures in year 2005 on categories 1103111 Clothing materials and accessories, 1103121 Garments, and 1103211 Footwear and divide by Total GDP. The GDP data for year 2010 is available from WDI. After obtaining the variable, multiply GDP by the expenditure share computed in the previous step to obtain the nominal value of expenditure on apparel and footwear in year 2010. Finally, to compute log_icp run Matlab file computing_log_go_log_icp.m, which relies on input file flows2010.txt, Matlab file construct_tradeshare_go.m and the input file containing ICP expenditures that are to be computed as described above.

log_go represents the log of the ratio of import share of country j from Spain to Spain’s domestic expenditure share, both in apparel and footwear industries only, in year 2008. To construct the

import and expenditure shares in apparel and footwear, two pieces of data are needed. The first is trade flows in these goods and the second is expenditures on these goods. The data source for trade flows is described in the main text of the paper. File titled `flows2008.txt` contains these observations, which are publicly available. The accompanying file `trade_flows_2008.xls` computes the trade flow observations from HS 2 digit level data.

Expenditures in year 2008 are proxied by gross output in apparel and footwear in 2008. File `gross_out_data_with_source.xls` contains the observation for each country and the source of the statistic. The accompanying file `go_29.txt` contains the same observations in a format that can be imported into Matlab.

To compute `log_go` run Matlab file `computing_log_go_log_icp.m`, which relies on input file `flows2008.txt`, `go_29.txt`, and Matlab file `construct_tradeshare_go.m`.

2.2 Variance Decomposition

The code to conduct the variance decomposition exercise is contained in file `var_dec_code.m`, which uses as inputs the data files `data_decomp_may_2014.txt`, `betas_bench_decomp_may_2014.txt`, `betas_dist_decomp_may_2014.txt`.

3 ICP Data and Code

The code that reproduces the empirical results that pertain to ICP (Section 5 of manuscript) is contained in file `icp_code.do`. All variables (with the exception of ICP prices) necessary to run the code are contained in files `wdi_2004.dta`, `geo_2004.dta`, `gini_cia.dta`, `mean_tariff.dta`, and `sales_tax.dta`. To see the description of a variable, consult the “label” column in the data file. The data source of each variable is described in the main body of the paper.

There are two types of variables: original and constructed. Original variables appear as reported in source data. Constructed variables are a combination of original variables. The code for all (except for one) constructed variables is contained in file `icp_code.do`. The exception is the variable titled `weighted_sitc4_tariff`. To construct it, follow these steps: (i) obtain bilateral tariff data at SITC 4-digit level for year 2004 from [Feenstra and Romalis \(2014\)](#) and bilateral trade flow data at SITC 4-digit level for year 2004 from [Feenstra et al. \(2005\)](#); (ii) relabel country codes using concordance file `country_codes.txt`; (iii) run `tariff_code.do`. The outcome is file `mean_tariff.dta`.

References

- FEENSTRA, R. C., R. E. LIPSEY, H. DENG, A. C. MA, AND H. MO (2005): “World trade flows: 1962-2000,” Tech. rep., National Bureau of Economic Research.
- FEENSTRA, R. C. AND J. ROMALIS (2014): “International Prices and Endogenous Quality,” *The Quarterly Journal of Economics*, 129, 477–527.