Documentation of data and the programs to reproduce the results of

Assortative Matching of Exporters and Importers

Review of Economics and Statistics

Yoichi Sugita, Kensuke Teshima, and Enrique Seira

Overview

The code in this replication package constructs the analysis file using Stata. The master do file `"master.do"` executes the do-files that generate the 3 figures and 4 tables in the main text of the paper as well as 3 figures and 24 tables in the online appendix.

This README file has four sections. Section 1 "Dataset List" explains the raw data and how each dataset can be obtained. Section 2 "Computational Requirements" explains the environment under which the codes were run. Section 3 "Description of Programs" gives an overview of the program files. Section 4 "List of Tables and Programs" provides the correspondence table of which results in the paper are produced by which program.

1. Dataset list

1.1 Mexican Customs Transaction Record

The main data used for this study is the custom transaction data of Mexico, which have been provided by the Mexican Ministry of Economy (Secretaria de Economia) to Instituto Tecnológico Autónomo de México (ITAM) as a result of an agreement between ITAM and the Ministry of Economy of Mexico. The data have been provided under the condition that the data are hosted securely at ITAM and that the researchers would not disclose any information of individual firms. Therefore, the raw data cannot be uploaded. Researchers who are granted ITAM affiliate status can have access the micro data at ITAM. Granting affiliate status is at the discretion of ITAM. There are several researchers who have used data hosted at ITAM under a similar affiliate arrangement and interested researchers should contact ITAM or/and the authors. In addition, we understand that the customs transaction data are also hosted at the tax authority (Servicio de Administración Tributaria), the Central Bank of Mexico and the statistical agency (INEGI), and we are aware of researchers who have access to the Mexican customs transaction data through these channels.

The Mexican customs transaction record in our study period had three formats depending on the period, (i) 2004-2011, (ii) 2000-2004 and (iii) 1998-1999. The customs transaction

record of format (i) has information on the transaction partners, while others do not. (i) is used for all the types of analysis, while (ii) and (iii) are used only for the analysis of capability cutoff changes.

1.1.1 Mexican Customs Transaction Record : 2004-2011

The data are organized by the type of information and month. YYYY and MM below denote year and month, respectively.

501-MMYYYY.txt
The files of this format have basic information on the Mexican exporter and transactions.

502-MMYYYY.txt
The files of this format have information on the transportation companies.

505-MMYYYY.txt
The files of this format have information on the transaction partners.

551-MMYYYY.txt
The files of this format have information on the goods traded such as their product code, value and quantity.

These files can be linked by transaction ID. We aggregate 501/505/551 data so that we can have trade volumes and average unit prices at the exporter-importer-product-year level. We use the 502 data to exclude the exporters that are transportation companies.

1.1.2 Mexican Customs Transaction Record : 2000-2004

All the variables are in stored in the files of format "XFEPYYMM.DTA", where YY is the last two digit of year, and MM is month.

1.1.3 Mexican Customs Transaction Record : 1999-2000

All the variables for all the periods are stored in one file: export1993_1999.dta.

1.2 US importer ID and characteristics

We constructed importer ID for the Mexican customs record for 2004-2011, using Orbis by Bureau van Dijik and using the procedure explained in Appendix B2. We are not allowed to post the Orbis data. The accessed to the Orbis data can be purchased from Bureau van Dijik. The resulting dataset is "matched_hsSS_common.dta" where SS is from 50 to 63, which are the HS chapters for the textile and apparel products. In this process, we also use the concordance file between zipcode and city/country to clean the address of the US importers, extracted from ZP4, an address standardization software we used to standardize the address. ZP4 was available until 2017, and was replaced by MAF, the Master Address File, which offers now free downloading of the software.
https://sites.google.com/site/masteraddressfile/maf/maf?authuser=0

1.3 MFA data

We downloaded the MFA made available at Peter Schott's website.
https://faculty.som.yale.edu/peterschott/international-trade-data/
and go to the data constructed and used by Brambilla, Khandelwal and Schott (2010).

We downloaded HS-level import data in 2004 also available at Peter Schott's website, constructed and used by Schott (2008).
https://spinup-000d1a-wp-offload-media.s3.amazonaws.com/faculty/wp-content/uploads/sites/47/2019/10/imp_detl_yearly_104n.zip

1.4 Product-level Trade data

We downloaded the product-level trade from Comtrade. In addition, we manually created the product characteristics data for the apparel products judging by names.

2. Computational requirements

Software Requirements
-Stata (code was last run with version 16)

The following commands should be installed.

```
-egenmore
-regsave
-estout
-tabform
-grc1leg2
```

The code was run on a 8-core Intel-based Lenovo server operating with Windows and with 128 GB of RAM. Computation took approximately 6 hours. Step 1 of the process explained the data takes most of the time. The rest will run within 5 minutes.

3. Description of programs

Perliminaries

`-master.do:` This do-file executes all the other codes run all steps in sequence. This do file lists the input files and output files of all the other do files.

`-set_paths.do`: This do-file set the path. adjust the paths. This has to be run before moving to other codes.

Replicators would need to create the subfolders in described in this do file to store intermediate and final outputs.

Step 1: From "Raw Data" to "Main Data"

1) `construct_binding_measures.do:` This do-file constructs the main independent variable, the Binding, which equals one if the Chinese exports of a HS6 product faced a binding quota.
2) `prepare_match_data.do`: This do-file constructs exporter-importer-product-year-level data from the customs transaction record data raw data from 2004 to 2011.
3) `prepare_data_for_exit.do:` This do-file constructs exporter-product-year-level data from the three types of the customs transaction record raw data from 2004 to 2011, from 2000 to 2004 and from 1998 to 1999.

The list of the dataset created, and the other data files together constitute the "Main Data", with the respective lists of variables, is attached at the end of this section.

4) `check_sample_selection.do:` This do-file performs the sample selection analysis conducted in Appendix B.3 and prepares the data for Appendix B.4.

Step 2: From "Main Data" to "Analysis Data"

1) `prepare_summary_statistics.do`: This do-file prepares the data for analyzing summary statistics. Do-file contains steps for:
    a. Make and clean data for summary statistics (Line20-249)
    b. Make and clean data for summary statistics restricting to main regression sample (Line252-477)
    c. Make and clean data for checking characteristics differences (Line481-669)

2) `create_ex_importer_rank.do`: This do-file contains two procedures of making own rank from "Main data."

    a. Make data on importer and exporter own rank (Line24-68)
    b. Make data on importer and exporter own rank based on five, ten, and twenty bins (72-133)

3) `create_rankings_main.do`: This do-file creates firms' main partner's ranking. The data made in this code are used in most of the analysis of the partner change.

4) `create_rankings_total.do`: This do-file creates the data on firms' ranking based on total trade. The data made in this code are used in robustness checks shown in Table A24.

5) `create_rankings_price.do:` This do-file creates the data on firms' ranking based on price measure. The data made in this code are used in robustness checks shown in Table A24.

6) `create_rankings_quality.do`: This do-file creates the data on firms' ranking based on the quality measure. The data made in this code are used in robustness checks shown in Table A24.

7) `create_rankings_bin.do`: This do-file creates the data on firms' ranking based on five, ten, and twenty bins. The data made in this code are used in robustness checks shown in Table A18.

8) `prepare_partner_change.do`: This do-file prepares further data for analyzing partner changes. This do-file contains steps for:

a. Creating 6-digit level ranking data, 2007-2011 2008-2011 2009-2011 for placebo checks shown in Table 3 Panel B (Line34-278)
b. Preparing control variables for regressions with controls shown in Table A12 and A13 (Line282-396)
c. Creating data for analyzing partner changes within and across initial partners shown in Table A15 (Line400-559)

Step 3: Producing the results of the paper.

1) `analyze_mfa_effects_aggregate.do`: This do-file generates the graphs related to the MFA's end effects on Mexican and Chinese exports to US (Figure 1).

2) `analyze_summary_statistics.do`: This do-file generates the results related to summary statistics (Table 1, Table A4-A6 and A9-A11). The do-file contains steps for:

    a. Making Table 1 and Table A4 from data generated in prepare_summary_statistics.do (Lines22-51)
    b. Making Table A5 and A6 from data generated in prepare_summary_statistics.do (Line54-63)
    c. Running regressions for checking differences of product level characteristics between treatments (Binding) and controls, i.e., Table A9, A10, and A11 (Lines66-183)

3) `analyze_export_change.do`: This do-file generates the results related to changes in Mexican exports to the United States from 2004 to 2007 (Table 2, 3, and A7). The do-file contains steps for:

    a. Decomposing the export changes into intensive and extensive margins, Table 2 (Line11-101)
    b. Conducting the same exercise as step a for the exports with main partner, Table 3 (Line105-197)
    c. Decomposing extensive margin into leaving and product dropping, Table A7 (Line200-236)

4) `analyze_partner_change.do`: This do-file generates the results related to partner changes (Table 3 Panel A, Table A1, A8, A12-A19, A21, A23, A24 Panel A, and Figures 3

and A4). The do-file contains steps for:

a. Running the benchmark regressions, Table 3 Panel A, from data generated in prepare_partner_change.do (Line39-79)
b. Conducting placebo checks using data on different periods, Table 3 Panel B (Line83-101)
c. Conducting weighted regressions as a robustness check, Table A1 (Line105-156)
d. Running probit instead of linear probability model as a robustness check, Table A8 (Line160-205)
e. Running regressions with more control variables, Table A12 and A13 (Line209-334)
f. Running regressions controlling for multi-product firms, Table A14 (Line338-383)
g. Running regressions that distinguish main partner switching beyond and across initial partners, Table A15 (Line386-401)
h. Running regressions that distinguish the firms with single partner from those with multiple partners, Table A16 and A17 (Line404-461)
i. Run regressions using rankings based on five, ten and twenty bins, Table A18 (Line465-507)
j. Running regressions using a continuous binding measure, Table A19 (Line511-549)
k. Running regressions using an alternative binding measure, Table A21 (Line553-591)
l. Checking gradual partner changes using data on 2004-06, 2004-07, and 2004-08, Table A23 (Line595-638)
m. Running regressions using alternative capability measures: total trade, price, and quality, Table A24 Panel A. (Line642-726)
   These measures are made in create_rankings_total.do, create_rankings_price.do, and create_rankings_quality.do
n. Analyzing the relation of partner change and initial ranks for Apparel and non-Apparel, Figure 3 and A4 (Line731-911)

5) `analyze_new_old_rank_correlation.do`: This do-file generates the results related to new and old partner rank correlations (Table 4, A5, and A24 Panel B). The do-file contains steps for:

a. Calculating the number of importers (exporters) by hs6 in 2004 (Line11-40)
b. Plotting the correlations of new and old ranks, Figure 4 and A5, without and with restricting sample to Binding=1, respectively. (Line43-131)
c. Analyzing the correlations using alternative capability measures: total trade, price and quality, Table A24 Panel B (Line135-201)

6) `analyze_firm_exit.do`: This do-file generates the results related to Mexican exporters' capability cutoff changes and exit from the market (Table 4, A20 and A22). The do-file contains steps for:

    a. Deriving the export cutoff and exit status, then analyze the effects of MFA's end using DiD strategy, Table 4 Panel A (Line11-420)
    b. Conducting the same exercise as step a using a continuous binding measure, Table A20 (Line424-625)
    c. Conducting the same exercise as step a using an alternative binding measure, Table A22 (Line630-829)

7) `analyze_number_of_partners.do`: This do-file generates the results related to (changes in) the number of partners (Table 4 Panel B and Table A3). The do-file contains steps for:

    a. Calculating the number of partners and regress on Binding dummy, Table 4 Panel B (Line9-89)
    b. Checking the robustness under different match definitions, Table A3 (Line94-138)

8) `analyze_import_growth.do`: This do-file generates the results for import growth of US firms during 2004-07 (Table A2).

The variable lists for the set of the Main data

matched_exp_imp_hs6_junedecember_maquila_restrict.dta	
Variable Name in Stata	Explanation/note
rfc	Exporter ID

newclid2	Importer ID
hs6	
year	
value_exporter_importer	Value of trade at the exporter-importer-hs6-year level.
maquilavalue	Volume of Maquiladora/IMMEX trade

matched_exp_imp_hs6_junedecember_maquila_restrict.dta	
Variable Name in Stata	Explanation/note
rfc	Exporter ID
newclid2	Importer ID
hs6	
year	
value_exporter_importer	Value of trade at the exporter-importer-hs6-year level.
cveunidad	ID for the unit of the quantity
Canto	Quantity

exp_rfc_prod_textile_us20042011.dta	
Variable Name in Stata	Explanation/note
year	
hs6	
rfc	Exporter ID
maquila	An indicator of a duty-free processing reexport program (Maquiladora/IMMEX)
valorendolares	Value of exports

exp_rfc_prod_textile20002004.dta

Variable Name in Stata	Explanation/note
year	
hs6	
rfc	Exporter ID
maquila	An indicator of a duty-free processing reexport program (Maquiladora/IMMEX)
FEPA_VAL	Trade Value

export1993_1999.dta	
Variable Name in Stata	Explanation/note
v2	Month/Year
v3	Hs6
v4	Exporter ID
v6	Importing country
v7	Trade value

description_hs6_type.dta	
Variable Name in Stata	Explanation/note
men	Dummy for products for men
women	Dummy for products for women
cotton	Dummy for cotton products
Wool	Dummy for wool products
manmade	Dummy for synthetic material products

chinatousexports19982011.dta	
Variable Name in Stata	Explanation/note
commoditycode	Commodity Code
period	Year

tradevalue	Trade value

mextousexports19982011.dta	
Variable Name in Stata	Explanation/note
tradevalue	Trade value
period	Year
commoditycode	Commodity code

quota_mex_weight_hs6.dta	
Variable Name in Stata	Explanation/note
iq_mex_bind80	Discrete binding measure, using imports from Mexico as the weight, which is our main "Binding" measure.
bind80_mex	Continuous binding measure, using imports from Mexico as the weight, for the analysis of Table A.19 and A.20.
iq_china_bind80	Discrete binding measure, using imports from China as the weight, for the analysis of Table A.21 and A.22.

4 "List of Tables and Programs"

Figure/Table #	Program	Line Number	Output file
Figure 1	analyze_mfa_effects_aggregate.do	all	mfa_effects_on_exports.png
Figure 2	theory figure		
Figure 3	analyze_partner_change.do	731-841	partner_change_intial_ranking_apparel.png
Figure 4	analyze_new_old_rank_correlation.do	43-131	new_old_rank_correlation_US.png new_old_rank_correlation_mexico.png

Table 1	analyze_summary_statistics.do	22-51	summarystats0407_productlevel.dta summarystats0407_firmlevel.dta
Table 2	analyze_export_change.do	11-101	aggregate_decomposition.dta partner_margin_coef.tex
Table 3	analyze_export_change.do	105-197	aggregate_decomposition_main_partner.dta main_partner_margin_coef.tex
Table 4 Panel A	analyze_firm_exit.do	11-420	mexican_exit.tex
Table 4 Panel B	analyze_number_of_partners.do	9-89	change_number_of_partners_0407.tex
Figure A1	check_sample_selection.do	50-63	distribution_kept_maquila.png distribution_kept_nonmaquila.png
Figure A2	theory figure		
Figure A3	theory figure		
Figure A4	analyze_partner_change.do	844-951	partner_change_intial_ranking_non_apparel.png
Figure A5	analyze_new_old_rank_correlation.do	43-131	new_old_rank_correlation_US_binding.png new_old_rank_correlation_mexico_binding.png
Table A1	analyze_partner_change.do	105-156	partner_change_0407_weighted.tex
Table A2	analyze_import_growth.do	all	import_growth.tex
Table A3	analyze_number_of_partners.do	94-138	ndam_alternative_matching_def.tex
Table A4	analyze_summary_statistics.do	22-51	summarystats0506_productlevel.dta summarystats0506_firmlevel.dta
Table A5	analyze_summary_statistics.do	54-58	summarystats_regsample_product_04050607.dta

Table A6	analyze_summary_statistics.do	59-63	summarystats_regsample_firm_04050607.dta
Table A7	analyze_export_change.do	200-236	decomposition_extensive.dta
Table A8	analyze_partner_change.do	160-205	partner_change_0407_probit.tex
Table A9	analyze_summary_statistics.do	66-106	import_product_level_characteristics_obs.dta product_level_characteristics_controls.csv product_level_characteristics_diff.tex
Table A10	analyze_summary_statistics.do	109-148	import_product_level_characteristics_obs.dta import_product_level_characteristics_control.csv import_product_level_characteristics_diff.tex
Table A11	analyze_summary_statistics.do	151-188	export_product_level_characteristics_obs.dta export_product_level_characteristics_control.csv export_product_level_characteristics_diff.tex
Table A12	analyze_partner_change.do	209-334	partner_change_0407_us_controls.tex
Table A13	analyze_partner_change.do	209-334	partner_change_0407_mex_controls.tex
Table A14	analyze_partner_change.do	338-383	partner_change_0407_multiproduct.tex
Table A15	analyze_partner_change.do	386-401	partner_change_beyond_within.tex
Table A16	analyze_partner_change.do	404-432	partner_change_multipartner_us.tex
Table A17	analyze_partner_change.do	435-461	partner_change_multipartner_mex.tex

Table A18	analyze_partner_change.do	465-507	partner_change_bin5.tex partner_change_bin10.tex partner_change_bin20.tex
Table A19	analyze_partner_change.do	511-549	partner_change_cont_binding_measure.tex
Table A20	analyze_firm_exit.do	424-625	mexican_exit_cont_binding_measure.tex
Table A21	analyze_partner_change.do	553-591	partner_change_alternative_binding_measure.tex
Table A22	analyze_firm_exit.do	630-829	mexican_exit_alternative_binding_measure.tex
Table A23	analyze_partner_change.do	595-638	gradual_partener_change.tex
Table A24 Panel A	analyze_partner_change.do	642-726	partner_change_alternative_ranking.tex
Table A24 Panel B	analyze_new_old_rank_correlation.do	135-201	rank_correlation_alternative_rank_measures.tex

Reference

Brambilla, Irene, Amit K. Khandelwal, and Peter K. Schott. 2010. China's Experience under the Multi-fiber Arrangement (MFA) and the Agreement on Textiles and Clothing (ATC). Robert C. Feenstra and Shang-Jin Wei eds., *China's Growing Role in World Trade*. University of Chicago Press.

Schott, Peter K. 2008. The Relative Sophistication of Chinese Exports. *Economic Policy* 53:5-49.

Acknowledgment

Hiroshi Kumanomido and Yuri Sugiyama provided excellent research assistance in preparing this documentation. The authors are solely responsible for the contents.