

Data Appendix for: “How Important is the New Goods Margin in International Trade?”
Timothy J. Kehoe and Kim J. Ruhl
April 2013

The DataAppendix folder contains the following subfolders

Data: Contains the data files discussed in sections 1 through 4 below

Programs: Contains Stata programs discussed in section 5

Output: Contains intermediate output created by the Stata programs in section 5

Figures: Contains the figures that appear in the paper, as created from the Stata programs in section 5

Tables: Contains the tables that appear in the paper in their unformatted version, as generated from the Stata programs and discussed in section 5.

1. Bilateral Trade Data (used in tables 3–7, figures 1, 3–13)

1.1 Original data

The data for Canada-Mexico, Canada-United States, China-United States, Germany-United States, Japan-United States, Mexico-United States, and United Kingdom-United States are from the United Nations COMTRADE database, accessed through World Integrated Trade Solution software.

The data for Chile-United States and Korea-United States are from the OECD at stats.oecd.org.

1.2 Processing

COMTRADE does not report a good that is never traded during the sample period. For example, good 01120 is never exported from Japan to the United States from 1989–99, so COMTRADE does not report good 01120 in the extracted data. We add these goods to the data set so that each country pair dataset lists 1,836 goods—the complete 5-digit set—even if some of those goods are never traded during the sample.

1.3 File list

The processed data are in the following 24 tab-delimited text files. The file name NNNMMMYYZZ.txt is made up of the exporting country (NNN) the importing country (MMM) the first year of the data (YY) and the ending year of the data (ZZ). Within the file, the first column lists the 5-digit SITC code (with leading zeros suppressed). The remaining columns list the value of trade by year, in sequential order, in thousands of U.S. dollars.

canmex8906.txt	canmex8999.txt	canusa8899.txt	chlusa7585.txt
chnusa8906.txt	chnusa9505.txt	gerusa8999.txt	jpnusa8999.txt
korusa7585.txt	mexcan8906.txt	mexcan8999.txt	mexusa8088.txt
mexusa8999.txt	ukkusa8999.txt	usacan8899.txt	usachl7585.txt
usachn8906.txt	usachn9505.txt	usager8999.txt	usajpn8999.txt
usakor7585.txt	usamex8088.txt	usamex8999.txt	usaukk8999.txt

2. World Trade Data (used in figures 12 and 13)

2.1 Original data

The data covering total world trade, by 5-digit SITC, for 1988–2006, are from COMTRADE.

2.2 Processing

These data have not been modified, besides importing them into a STATA data format.

2.3 File list

The first column of the file lists the 5-digit SITC code, with leading zeros suppressed. The remaining columns list the value of trade by year, sequentially, in thousands of U.S. dollars.

The file name is: total_by_good_year_8806.dta

3. Multicountry Panel Dataset (used in figure 2 and tables 1–2)

3.1 Original data

The original data are from the COMTRADE database. The data includes bilateral trade between all countries for which there is data availability for 1995 as well as non-zero reported trade. We do this separately for the SITC Rev. 2 and the HS 1992 concordances.

A complete list of data availability for each country by concordance system is available on COMTRADE: <http://comtrade.un.org/db/mr/daYearsResults.aspx>

3.2 Processing

For each country pair we perform the least traded exercise by sorting the goods from smallest to largest by the amount of trade over the first three years of the sample and then summing the goods up from smallest to largest until we reach 10% of total trade in the base year. We denote those goods as the least traded goods. For each year, we then compute the growth in total trade between the two countries and the growth in trade of least traded goods between the two countries.

The original data is not included in the appendix files due to size limitations, but is available from the authors by request.

3.3 File list

The data by SITC classification are in SITC_imports_least_1995.dta. The data by Harmonized System classification are in HS92_imports_least_1995.dta.

Each observation in the data set corresponds to an importer-exporter pair, where first two columns list the country codes of the importer country and exporter country, respectively. The columns total_X gives the total value of trade (imports by the importing country from the exporting country) in year X, where X ranges from 1995 to 2006, and the columns all_goods_X give the total number of goods that had positive trade reported in year X. The columns least_val_X gives the value of trade of least traded goods in year X, while the columns least_goods_X give the total number of least traded goods that reported positive trade in year X.

The columns growth_total_y1_yZ, where Z ranges from 2 to 12, lists the percent growth (1=100%) in total trade between the base year (y1=1995) and Z years after the base period (for example y12 = 2006). Similarly, the columns growth_least_y1_yZ similarly gives the percent growth in trade due to least traded products, while the columns growth_exm_y1_yZ gives the percent growth in the proportion of trade accounted for by least traded goods.

4. Miscellaneous Data

4.1 Real GDP per Capita (filename: rgdpl_WDI.dta)

Real GDP per capita for years 1995 and 2005 is from the World Bank's World Development Indicators database. (<http://databank.worldbank.org/ddp/home.do>, series name NY.GDP.PCAP.KD) The data are in constant 2000 U.S. dollars.

4.2 Trade Agreement Indicators (filenames: cepii_1995.dta, cepii_2005.dta)

The data are from The Cepii Gravity Dataset (<http://www.cepii.fr/anglaisgraph/bdd/gravity.asp>), for the years 1995 and 2005.

4.3 Country Names (filename: names.dta)

This file matches country names and abbreviations with the corresponding UN country number. This file also includes a hand-compiled indicator equal to 1 if the country is a member of the OECD.

4.4 SITC Classification (sitc_names.dta)

This file contains the description of each SITC 5-digit code.

4.5 Bilateral Trade Data for Canada and United States for Appendix

Data is from COMTRADE and gives bilateral trade between the United States and Canada as reported by Canada and as reported by the United States.

Filenames: can_to_usa_by_can.csv, can_to_usa_by_usa.csv, usa_to_can_by_can.csv, usa_to_can_by_usa.csv

4.6 Producer Price Index for Manufacturing in the United States (ppi.xlsx)

Data is from the BLS and gives the growth in the producer price index for manufacturing industries in the United States between 1995 and 2005.

5. Programs

The programs are titled according to which figures and tables they produce. Before running the programs change the Stata working directory to the base Data Appendix folder using the `cd` command. Alternatively, one may alter `data_dir` and `out_dir` in the programs to contain the full paths to the subfolders folders.

Note: Most programs must be run multiple times, moving the `/*` and `*/` after each run, to produce all output.

5.1 Table 1, Table 2, and Figure 2 (table_1_2_fig_2.do)

Uses the files from 3.3, 4.1, 4.2, 4.3 to create the output for table 1 (table1.txt), table 2 (reg_sitc_1995.txt and reg_hs_1995.txt), and figure 1 (fig1.eps).

5.2 Table 3, Table 4, and Table 5 (table_3_4_5.do)

Uses the files from 1.3 and 2.3 to create the output for tables 1, 2, and 3. For each file it inputs from 1.3 it outputs two text files, one with a _tss suffix and the other with a _bar suffix.

Therefore the output names NNNMMMYZZ_tss.txt and NNNMMMYZZ_bar.txt are made up of the exporting country (NNN) the importing country (MMM) the first year of the data (YY) and the ending year of the data (ZZ), where the list of NNNMMMYZZ used are given in the file list of 1.3 above.

Files that have the _tss suffix have the first column is the year and the second column is the share of exports from country NNN to country MMM accounted for by least traded goods.

Files that have the _bar suffix sort and separate trade into bins each accounting for 10% of trade in the base year (YY). The first column is the bin, the second column is the share of exports from country NNN to country MMM accounted for by goods in that bin in the end year (ZZ), and the third column is the number of goods in each bin.

The processing the program does to the data to create the output is the same as in 3.2, however the program is also able to do the initial sorting of goods from smallest to largest according to world trade in each good. This alternative sorting is enabled by changing the 0 next to alt_sort to 1; and creates identical files except with the additional suffix _alt. The output canmex8999_bar_alt.txt and mexcan8999_bar_alt.txt is required to run fig_12_13.do.

Note: this program must be run before running fig_1_3_5_6_9_10.do , fig_4_7_8_11.do, or fig_12_13.do.

Figures 1, 3, 5, 6, 9, and 10 (fig_1_3_5_6_9_10.do)

Uses the output from table_3_4_5.do to produce figures 1, 3, 5, 6, 9, and 10. Filenames: fig1.eps, fig3.eps, fig5.eps, fig6.eps, fig9.eps, and fig10.eps.

Figures 4, 7, 8, and 11 (fig_4_7_8_11.do)

Uses the output from table_3_4_5.do to produce figures 4, 7, 8, and 11. Filenames: fig4.eps, fig7.eps, fig8.eps, and fig11.eps.

Figure 12 and Figure 13 (fig_12_13.do)

Uses the output from table_3_4_5.do to produce figures 12 and 13. Filenames: fig12.eps, fig13.eps.

Table 6 (table_6.do)

Uses the data from 1.3 to produce the output for table 6. Results are only printed on the screen and not saved to file.

Table 7 (table_7.do)

Uses the data from 1.3 to produce the output for table 7. Results are only printed on the screen and not saved to file.

Table A1 and Figure A1 (appendix.do)

Uses the data from 4.5 to produce the output for table A1 and figure A1 (figA1.eps). The results for table A1 are saved to the files: app_can_to_usa_by_can.txt, app_can_to_usa_by_usa.txt, app_usa_to_can_by_can.txt, and app_usa_to_can_by_usa.txt; where the first column is the year, the second column is the number of code births in that year, and the third column is the number of code deaths in that year.