Data and code for:
Trade and US Inequality in the Tokyo Round

Andrew Greenland*
North Carolina State University

James Lake†
University of Tennessee

John Lopresti‡
William & Mary

Contents

Overview **2**

Data Availability and Provenance Statements **2**

Statement about Rights 2

Summary of Availability 2

Details on each Data Source and Dataset List **2**

Computational requirements **7**

Software Requirements 7

Controlled Randomness 7

Memory, Runtime, Storage Requirements 8

Description of programs/code **8**

Instructions to Replicators **9**

List of tables and programs **10**

*Corresponding author e-mail: agreenl@ncsu.edu
†E-mail: jlake7@utk.edu
‡E-mail: jwloprest@wm.edu

Overview

This replication package uses STATA and Microsoft Excel. The [**code**] folder contains all code. The master do file **1_master_file.do** executes the dataset construction do files (with prefixes **2_**) from various data sources in the [**data**]\[**raw**] folder. It also executes the analysis do files (with prefixes **3_**) which save all figures and tables from the main text and the Online Appendix to the [**figures**] and [**tables**] folders. Upon executing the master do file, the code should take about 5 minutes to run.

Data Availability and Provenance Statements

Statement about Rights

- ☑ We certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ☑ We certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package.

Summary of Availability

- ☑ All data **are** publicly available.

Details on each Data Source and Dataset List

The master do file ([**code**]**1_master_file.do**) executes the dataset creation do files (with prefixes **2_**) that create the dta files used by the analysis do files (with prefixes **2_**). The following lists all of the raw datasets used by the dataset creation and analysis do files. All raw datasets are located in the [**data**]\[**raw**] folder.

[2a.] concordances.do

- **hs92_sitc2_concordance.xlsx**. Concordance between HS 1992 and SITC revision 2 from United Nations Statistics Division (https://unstats.un.org/unsd/classifications/Econ.)
- **hs_sic_naics_imports_89_121_20220627.dta**. Concordance between HS, NAICS and SIC from Pierce and Schott (2012).

- **sic_census_concordance_raw.xlsx**. Concordance between 1930s Census industries and 4-digit 1987 SIC industries. Created manually by authors.
- **cccn_hs_wits.csv**. Concordance between CCCN and HS from the World Bank's WITS platform (https://wits.worldbank.org/product_concordance.html).
- **canada_us_concordance.xlsx**. Concordance between Canadian and US SIC from Trefler (2004) (https://www-2.rotman.utoronto.ca/ dtrefler/files/Data.htm).
- **occ1990_xwalk.xls**. Concordance between IPUMS and Census occ1990 variables from Ruggles et al. (2024) (https://usa.ipums.org/usa/volii/occ_ind.shtml).
- **occ1990_occ1990dd.dta**. Concordance between Census occ1990 and David Dorn's occ1990dd variables from Autor and Dorn (2013) (http://www.ddorn.net/data.htm).
- **conc_sic72_sic87.dta**. Concordance between 1972 and 1987 SIC codes from Ruggles et al. (2024) (https://www.nber.org/research/data/nber-ces-manufacturing-industry-database).
- **concordance_1977.xlsx**. Concordance between BEA commodity codes and SIC codes used by BEA in their 1977 IO tables (https://www.bea.gov/industry/historical-benchmark-input-output-tables). This excel file was created by the authors from the raw downloaded file **1977 Sectoring Plan.doc**.
- **sic_1972_full.dta**. List of 1972 SIC industry codes and names from Ruggles et al. (2024) (https://www.nber.org/research/data/nber-ces-manufacturing-industry-database).

[2b.] clean_1930s_data.do

- **census1930_table5_gender.xlsx**. Digitized version of Table 1 from Chapter X of Volume I of the 1929 Census of Manufactures (https://www.census.gov/library/publications/1933/dec/1930f-vol-01-manufactures.html). Digitized by authors.
- **census1930_table5.xlsx**. Digitized version of Table 5 from Chapter II of

Volume I of the 1929 Census of Manufactures (https://www.census.gov/library/publications/1933/dec/1930f-vol-01-manufactures.html).

[2c.] reciporical_tariffs.do

- **xm_sic87_72_105_20120424.dta**. US import and export data for the 1972-2005 period at the 4-digit SIC level from Schott (2008).
- **eec_tokyo_cccn.xlsx**. Tokyo Round tariff schedule for the EEC, digitized by authors from document on the WTO website (https://www.wto.org/english/docs_e/gattbilaterals_e/indexbyround_e.htm).
- **japan_tokyo_cccn.xlsx**. Tokyo Round tariff schedule for Japan, digitized by authors from document on the WTO website (https://www.wto.org/english/docs_e/gattbilaterals_e/indexbyround_e.htm).
- **Trefler_Annual_Level_Transform.xls**. Tokyo Round tariff schedule for Canada from Trefler (2004) (https://www-2.rotman.utoronto.ca/ dtrefler/files/Data.htm).

[2d.] construct_sic_price_growth.do

- **sitc_prices_imports.dta**. SITC-level unit values and imports from Greenland et al. (2025).

[2e.] construct_mfa_shares.do

- **mfa_78.xlsx**. Digitized list of TSUSA tariff lines subject to Multi Fiber Agreement from pages 136-141 of the 1978 TSUSA (https://hdl.handle.net/2027/uiug.30112101582838).
- **TSUSA7_SITC4_crosswalk.dta**. Concordance between 7-digit TSUSA codes and 4-digit SITC codes from Greenland et al. (2025), which is based on the import data from Feenstra (1996) that contains booth 7-digit TSUSA and 4-digit SITC codes in the underlying import data.

[2f.] import_ipums.do

- **USA_00236.dat**. IPUMS data from Ruggles et al. (2024).

[2g.] construct_sic_routineness.do

- **occ1990dd_alm.dta**. Routineness variables from Autor and Dorn (2013) (http://www.ddorn.net/data.htm).
- **cw_ind1990_ind1990ddx.dta**. Concordance between ind1990 and ind1990ddx variables from Autor et al. (2019) (http://www.ddorn.net/data.htm).
- **cw_sic87_ind1990ddx.dta**. Concordance between ind1990ddx and SIC from Autor et al. (2019) (http://www.ddorn.net/data.htm).

[2h.] construct_sic_automation.do

- **task_content_bea_1947_1978.dta**. Automation data from Acemoglu and Restrepo (2020) (https://www.openicpsr.org/openicpsr/project/120425/version/V2/view).
- **industry_xwalk_census_bea.dta**. Concordance between BEA industry and Census ind1990 variables Acemoglu and Restrepo (2020) (https://www.openicpsr.org/openicpsr/project/120425/version/V2/view).
- **cw_ind1990_ind1990ddx.dta**. Concordance between ind1990 and ind1990ddx variables from Autor et al. (2019) (http://www.ddorn.net/data.htm).
- **cw_sic87_ind1990ddx.dta**. Concordance between ind1990ddx and SIC from Autor et al. (2019) (http://www.ddorn.net/data.htm).

[2i.] sic87_dataset_creation.do

- **xm_sic87_72_105_20120424.dta**. US import and export data for the 1972-2005 period at the 4-digit SIC level from Schott (2008).
- **nberces5818v1_s1987.dta**. NBER-CES Manufacturing Industry Database (Ruggles et al. (2024)).
- **ave_iv_ols.dta**. Relevant AVE and import data at SIC level from Greenland et al. (2025) for constructing dependent variable AVE measures and their instruments.
- **1977 Total Req Coeff 537-level Data.txt**. 1977 total requirement data from the US BEA IO tables (https://www.bea.gov/industry/historical-benchmark-input-output-tables).

[2j.] construct_women_shares.do

- **sic4_2_census90_v2.dta**. Concordance between IPUMS Census industry codes and SIC from Lake and Millimet (2016).

[2k.] **regression_dataset_creation.do**

- No raw data used.

[2l.] **ipums_data_manipulation.do**

- **occ1990dd_alm.dta**. Routine task intensity data from Autor and Dorn (2013).
- **sic4_2_census90_v2.dta**. Concordance between IPUMS Census industry codes and SIC from Lake and Millimet (2016).

As described above, the dataset creation do files (with prefixes **2_**) create the dta files needed for the analysis. The following lists all of the datasets used by the analysis do files (with prefixes **3_**). They are all contained in the [**data**]\[**analysis**] or [**data**]\[**raw**] folders.

[3a.]
- [**analysis**]**SIC87_regression_Dataset.dta**. Created in [**2k**].
- [**analysis**]**com_controls_clean.dta**. Created in [**2b**].
- [**raw**]**swiss_1930.dta**. At 4-digit SIC-level, digitized 1930 Smoot-Hawley tariffs from Greenland and Lopresti (2024) and inserted into Swiss IV.
- [**raw**]**xm_sic87_72_105_20120424.dta**. US import and export data for the 1972-2005 period at the 4-digit SIC level from Schott (2008).

[3b.]
- [**analysis**]**ipums_industry_data.dta**. Created in [**2l**].
- [**analysis**]**ind1990_changes_rout'r'.dta** for r=1,2,3. Created in [**2l**].
- [**analysis**]**ipums_1970_1990.dta**. Created in [**2f**].
- [**analysis**]**occ1990_occ1990dd_GLL.dta**. Created in [**2a**].
- [**analysis**]**occ1990_occ1990dd_weights.dta**. Created in [**2l**].
- [**raw**]**occ1990dd_alm.dta**. Routineness variables from Autor and Dorn (2013) (http://www.ddorn.net/data.htm).

[3c.]
- [**raw**]**sic87_3digit.xlsx**. 2-digit SIC industry names, slightly modified by authors from Census Bureau (https://dataweb.usitc.gov/classification/commodity-description/SIC/2).
- [**analysis**]**SIC87_regression_Dataset.dta**. Created in [**2k**].

- [**raw**]**xm_sic87_72_105_20120424.dta**. US import and export data for the 1972-2005 period at the 4-digit SIC level from Schott (2008).
- [**analysis**]**ipums_industry_data.dta**. Created in [**2l**].
- [**analysis**]**ind1990_changes_rout1.dta**. Created in [**2l**].

Computational requirements

Software Requirements

☑ The replication package contains one or more programs to install all dependencies and set up the necessary directory structure.

- Stata (code was last run with version 19.5)
  - mmerge (v2.5.0 as of 26fed2002)
  - unique (v1.2.4 as of Jun 17, 2020)
  - ereplace (v1.0.3)
  - estout (v3.31.0 as of 26apr2022)
  - ranktest (v2.0.04 as of 21sept2020)
  - quantiles (v1.1 as of June 2006)
  - require (v1.3.1 as of 19sept2023)
  - grc1leg (v1.0.5 as of 02jun2010)
  - ivreghdfe (v1.1.4 as of 29nov2025)
  - ivreg2 (v4.1.12 as of 14aug2024)
  - reghdfe (v6.13.0 as of 09jan2026)
  - ftools (v2.50.0 as of 09aug2026)
- The program **1_master_file.do** installs all dependencies locally and sets up paths for the remaining do files. Note that this do file uninstalls the packages ftools, reghdfe, ivreg2, and ivreghdfe and reinstalls them from particular sources to ensure correct functionality of ivreghdfe.

Controlled Randomness

☑ No pseudo random generator is used in the analysis described here.

Memory, Runtime, Storage Requirements

Summary Approximate time needed to reproduce the analyses on a standard 2026 desktop machine is:

☑ 5-10 minutes

Approximate storage space needed:

☑ 2 GB - 25 GB

Details The code was last run on a 7-core Intel-based desktop with Windows 11 version 23H2 with 64GB RAM and 1TB of free SSD hard drive space.

Description of programs/code

All do files are stored in the [**code**] folder. The master do file **1__master__file.do** first executes the dataset creation do files (which have prefixes **__2**) that create the analysis dta files and saves them in the [**data**]\[**analysis**] folder. In the process of creating the analysis files, dta files saved along the way are stored in [**data**]\[**processing**] which is an empty folder when first opening the replication package.

The master do file **1__master__file.do** then executes the analysis do files (which have prefixes **__3**) that use the dta files in the [**data**]\[**analysis**] folder and some files in the [**data**]\[**raw**] folder to create all tables and figures in the paper and in the Online Appendix. Figures are saved in the [**figures**] folder with filenames **fig__n.png**. Tables are saved in the [**tables**] folder in latex format with filenames **table__n.tex** and in the more easily readable log file format with file names **3a__sic__analysis__tables.log**, **3b__ipums__analysis__tables.log**, and **3c__additional__analysis__tables.log**.

The [**code**] folder contains the following do files:

1. **Master do file.** The master do file **1__master__file.do** begins by (i) setting paths as global macros and (ii) checking for the necessary STATA packages and their dependencies and installing them when not already installed on the computer. It then executes the dataset creation and analysis do files to generate all of the results, tables and figures for the main text and the Online Appendix.
2. **Dataset construction do files**
   - **2a__concordances.do** creates all necessary concordances.
   - **2b__1930s__data.do** creates the 1930s tariff and control variables.

- **2c_reciporical_tariffs.do** cleans the Tokyo Round tariff cuts of the EEC, Japan and Canada.
- **2d_construct_sic_price_growth.do** creates lagged price growth at 4-digit SIC level.
- **2e_construct_mfa_shares.do** creates share of TSUSA7 lines within 4-digit SIC covered by MFA.
- **2f_import_ipums.do** prepares IPUMS data.
- **2g_construct_routineness.do** constructs routineness measures at IPUMS ind1990 and 4-digit SIC levels.
- **2h_construct_automation.do** constructs automation measures at IPUMS ind1990 and 4-digit SIC levels.
- **2i_construct_women_shares.do** constructs women's share of employment at 4-digit SIC level.
- **2j_sic87_dataset_creation.do** constructs panel dataset at 4-digit SIC level with NBER outcomes, trade data outcomes, and tariff measures.
- **2k_sic_regression_dataset_creation.do** constructs the main dataset used for SIC-level analysis.
- **2l_ipums_regression_dataset_creation.do** constructs the main datasets used for IPUMS industry-level analysis.

3. **Analysis do files**
    - **3a_sic_analysis.do** carries out most of the SIC-level analysis
    - **3b_ipums_analysis.do** carries out most of the IPUMS industry-level analysis.
    - **3c_additional_analysis.do** carries out the remaining analysis.

Instructions to Replicators

Before executing the master do file [**code**]**1_master_file.do**, the user should open it to edit the global macro for the root path. This is done in line 21 and is the only place that the paths need to be updated when executing the replication package via the master file. When running the do files in isolation, order is important and should follow that as executed in the master do file.

It is recommended that the user clear all files from their [**personal**]\[**ado**] folder which is where STATA stores packages locally. The master do file will install all necessary packages.

The master file was last executed on January 12 2026.

List of tables and programs

The three analysis do files **3a_sic_analysis_tables.do**, **3b_ipums_analysis_tables.do**, and **3c_additional_analysis_tables.do** reproduce:

☑ All numbers provided in text in the paper
☑ All tables and figures in the paper
☐ Selected tables and figures in the paper, as explained and justified below.

For each table and figure in the main text and online appendix, the following lists the associated analysis do file and the line number that the code starts. All figures are stored in the [**figures**] folder with filenames **fig_n.png**. All tables are stored in the [**tables**] folder in both latex format with filename **table_n.tex** and the more user-friendly log file format with filenames corresponding to the name of the analysis do file (i.e. three log files for analysis do files 3a, 3b, and 3c).

- **Table 1.** Line 5 of do file **3a**.
- **Table 2.** Line 57 of do file **3a**.
- **Table 3.** Line 164 of do file **3a**.
- **Table 4.** Line 359 of do file **3a**.
- **Table 5.** Line 5 of do file **3b**.
- **Table 6.** Line 234 of do file **3b**.
- **Figure 1.** Line 5 of do file **3c**.
- **Figure 2.** Line 352 of do file **3c**.
- **Figure 3.** Line 543 of do file **3c**.
- **Figure 4.** Line 1247 of do file **3a**.
- **Online Appendix Table A1.** Line 243 of do file **3c**.
- **Online Appendix Table A2.** Line 595 of do file **3a**.
- **Online Appendix Table A3.** Line 57 of do file **3a**.
- **Online Appendix Table A4.** Line 641 of do file **3a**.
- **Online Appendix Table A5.** Line 742 of do file **3a**.

- **Online Appendix Table A6.** Line 856 of do file **3a**.
- **Online Appendix Table A7.** Line 945 of do file **3a**.
- **Online Appendix Table A8.** Line 1046 of do file **3a**.
- **Online Appendix Table A9.** Line 1152 of do file **3a**.
- **Online Appendix Table A10.** Line 164 of do file **3a**.
- **Online Appendix Table A11.** Line 5 of do file **3b**.
- **Online Appendix Table A12.** Line 536 of do file **3b**.
- **Online Appendix Table A13.** Line 750 of do file **3b**.
- **Online Appendix Table A14.** Line 865 of do file **3b**.
- **Online Appendix Table C1.** Line 352 of do file **3c**.
- **Online Appendix Figure A1.** Line 352 of do file **3c**.
- **Online Appendix Figure A2.** Line 1313 of do file **3a**.
- **Online Appendix Figure A3.** Line 1352 of do file **3a**.
- **Online Appendix Figure A4.** Line 1450 of do file **3a**.
- **Online Appendix Figure A5.** Line 601 of do file **3c**.
- **Online Appendix Figure A6.** Line 977 of do file **3b**.
- **Online Appendix Figure A7.** Line 1022 of do file **3b**.
- **Online Appendix Figure B1.** Line 1509 of do file **3a**.
- **Online Appendix Figure B2.** Line 1553 of do file **3a**.
- **Online Appendix Figure C1.** Line 352 of do file **3c**.
- **Online Appendix Figure D1.** Line 1582 of do file **3a**.
- **Online Appendix Figure D2.** Line 1582 of do file **3a**.

References

Acemoglu, D. and P. Restrepo (2020). Unpacking skill bias: Automation and new tasks. *AEA Papers and Proceedings 110*, 356–361.

Autor, D., D. Dorn, and G. Hanson (2019). When work disappears: Manufacturing decline and the falling marriage market value of young men. *American Economic Review: Insights 1*(2), 161–178.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review 103*(5), 1553–1597.

Feenstra, R. C. (1996). US imports, 1972-1994: Data and concordances. *NBER WP No. 5515*.

Greenland, A., J. Lake, and J. Lopresti (2025). The GATT vs Inflation: Tokyo Drift. *NBER Working Paper No. 34429*.

Greenland, A. and J. Lopresti (2024). Trade policy as an exogenous shock: Focusing on the specifics. *NBER Working Paper No. 33127*.

Lake, J. and D. L. Millimet (2016). Good jobs, bad jobs: What's trade got to do with it? *Mimeo*.

Pierce, J. R. and P. K. Schott (2012). A concordance between ten-digit US harmonized system codes and SIC/NAICS product classes and industries. *Journal of Economic and Social Measurement 37*(1-2), 61–96.

Ruggles, S., S. Flood, M. Sobek, D. Backman, A. Chen, G. Cooper, S. Richards, R. Rodgers, and M. Schouweiler (2024). IPUMS USA: Version 15.0.

Schott, P. K. (2008). The relative sophistication of Chinese exports. *Economic Policy 23*(53), 6–49.

Trefler, D. (2004). The long and short of the Canada-US free trade agreement. *American Economic Review 94*(4), 870–895.