Replication Materials for:
"Tariff Passthrough at the Border and at the Store: Evidence from US Trade Policy."

By ALBERTO CAVALLO, GITA GOPINATH, BRENT NEIMAN, AND JENNY TANG

Available at Harvard Dataverse: https://doi.org/10.7910/DVN/JV7FCH

This replication package contains the retail price data, some processed statistics of our computations based on both our data and confidential datasets, and the program files to replicate most of the Figures and Tables.

The full replication of results presented in our paper requires access to confidential data from the US Bureau of Labor Statistics and Datamyne LLC, as detailed in Table A1 in the Online Appendix. It is possible to gain access to these data by filing a request under the for the BLS restricted micro data available at www.bls.gov/rda/. To access to the Datamyne database, see www.datamyne.com.

Our retail data combines product information collected by the Billion Prices Project (www.thebillionpricesproject.com) and by PriceStats LLC (www.pricestats.com), as detailed in the Online Appendix. All data from these sources are included in the replication package. Another subset of the retail data containing the country of origin and HS codes for some individual goods in Retailer 2 is proprietary information provided by that retailer and cannot be shared.

This replication package contains the following files:

RETAIL

The following files are inside the RETAIL subfolder:

Stata do-files:

The Stata do-files have been tested on Stata versions 15. Some user-written commands may need to be manually installed.

MAIN_RETAIL.do	Main replication file for the retail results. This is the only file that needs to be executed. All other do files are automatically executed from the main file.
The main do-file contains explanations for the other do files in the CODE subfolder.	

List of Figures/Tables and the program code that generates them

Figure or Table	Do-file
Figure 2 and A4	selected_goods.do
Figure 3b and A8	international_identical_prices.do
Figure A9	international_manyretailers.do

Figure A7	international_cpi.do
Figure 4 and A10	customs.do
Table A7	Main.do
Figure 3a and A6	main_china_index_month
Table A6	main_stats.do main_stats_subsets.do
Column 7 in Table 1 and Column 3 in Table A8	main_china_preg_rate.do
Columns 5 in Table A8	main_china_preg_rate_ALLMC.do
Columns 6 in Table A8	main_china_preg_rate_DR2.do
Column 2 in Table A8	main_china_preg_rate_ck.do
Column 4 in Table A8	main_china_preg_rate_convex.do
Table A9	main_china_preg_rate_coicop.do

Stata data-files:

These dta files are in the MICRODATA subfolder:

ALL_alldata_monthly.dta	Combined micro data from the 2 large retailers, with country of origin and HS code information.
R1_alldata_monthly.dta	Micro data from Retailer 1
R2_alldata_monthly.dta	Micro data from Retailer 2
DR2_alldata_monthly.dta	Micro data from Retailer 2, subset with direct imports
ALLMC_alldata_monthly.dta	Subset of all manually-classified goods in ALL.
DR2ALLMC_alldata_monthly.dta	Combines DR2 and ALLMC datasets
bicycle_all.dta	Micro data for bicycles from multiple retailers. No country of origin or HS code information.
handbag_all.dta	Micro data for handbags from multiple retailers. No country of origin or HS code information.
refrigerator_all.dta	Micro data for refrigerators from multiple retailers. No country of origin or HS code information.
tire_all.dta	Micro data for tires from multiple retailers. No country of origin or HS code information.

wash_all.dta	Micro data for washing machines from multiple retailers. No country of origin or HS code information.
matched_ids_allprices.dta	Micro data for US and Canada prices for identical goods matched across countries in the same retailer.

Variable Description:

All the data files with names ending in "alldata_monthly" contain the following variables:

- date : date in stata %td format
- ym: year-month in Stata %tm format
- price: product's price in USD
- miss: =1 if the price was missing that day. =0 otherwise
- coicop: coicop category, 3-digit level. See UN (2018)
- china: =1 if the country of origin is China.
- Imported: =1 if the product is imported
- HS6: Global product classification. HS code at the 6-digit level. See WCO (2017)
- hs_rate_eff: effective additional tariff rate for a given HS6 code at that point in time.
- hs_tariff: =1 if the HS6 category to which the product belong is affected by additional tariffs at that point in time (if so, this is =1 even if the product is not from China)
- wave: tariff wave. Values represent the ym date the additional tariff was first imposed.
- hs_impacted: =1 if the HS6 category to which the product belong is affected by additional tariffs at any point in time (if so, this is =1 even if the product is not from China)
- tariff: =1 if hs_tariff=1 and china==1
- direct_import: =1 if the product is imported directly by the retailer (no intermediaries)
- iso2: is the country of origin of the product, using iso2 country codes
- hs_manual: =1 if the product was manually classified into the HS code
- retailer: 1 or 2 , anonymized
- id: id of the product. Anonymized (do not match the original ids in the retailers)

All the data files with names ending in "_all" contain the following variables:

- day : day of the week
- month : month
- year: year
- price : price of the product in USD
- url_id : id that identifies the url of the product.
- id: id of the product. Anonymized (do not match the original ids in the retailers)
- date: date in Stata %td format
- retailer: 1 or 2, anonymized

The file matched_ids_allprices.dta contains the following variables:

- id: id of the product. Anonymized (do not match the original ids in the retailers)
- ym: year-month in Stata %tm format
- date : date in stata %td format
- coicop: coicop category, 3-digit level. See UN (2018)

- uprice: price of the product in the US, in USD
- umiss: =1 if the price was missing that day in the US. =0 otherwise
- cprice: price of the product in Canada, in Canadian Dollars
- cmiss : =1 if the price was missing that day in Canada. =0 otherwise

Instructions for retail replication

1. Download and extract the replication package to "c:\TPBS\REPLICATION"
2. Run MAIN_RETAIL.do in the RETAIL subfolder.
3. The code should run and save the output in the existing subfolders.

***Figures that cannot be reproduced without access to the confidential / proprietary data**

- Figure 4: Front Running and Trade Diversion

Reason: We are not share the proprietary data from Datamyne. The code can be found in our replication packet.

- Figure A5: Retail Washing machine Prices, Variation across Brands

Reason: We are not able to show individual product descriptions because they would identify the retailers. The brands are extracted from the product descriptions. The code can be found in our replication packet.

- Table A6: The country and HS6 count that can be generated with this replication code will not be identical to the one reported in the paper because we are not able to share country of origin or HS codes for direct imports from Retailer 2.

BLS

The following files are inside the BLS subfolder:

Stata do-files:

The Stata do-files have been tested on Stata versions 15.

MAIN_BLS.do	Main replication file for the BLS IPP results. This is the only file that needs to be executed. All other do files are automatically executed from the main file.
The main do-file contains explanations for the other do files in the CODE subfolder.	

Stata data-files:

The code calls the following dta files that are based on the BLS IPP restricted access data and are not included in this package.

IPP_data_raw.dta	A file that merges the BLS "Price" and "Item master" files.

Item_dynamic.dta	The BLS "Item dynamic" file.
Mapyear_tree_timeline.dta	A Stata version of the "Methodology Timeline.xls" file available on the BLS data server which gives map years for each data/price year.
Map_unique.dta	A subset of the BLS "Map" file keeping only the relevant "trees" from the "Methodology Timeline" for each map year to get a unique mapping between each item_code and corresponding classification group and lower stratum.
Weight_classif_unique.dta	A subset of the "Weight classif" file keeping only the relevant "tree groups" from the "Methodology Timeline" for each map year to get a unique mapping between classification groups and weights.

These dta files are in the DATA subfolder:

countrynamemapping.dta	Correspondence between country codes available in the IPP data and country names/classifications.
macro_all.dta	Macro data on USD exchange rates and aggregate price indices (PPI if available, CPI otherwise).
imports_HS6*yyyy*.dta exports_HS6*yyyy*.dta	Additional import or export duties at the HS6-country level for HS classification system of various vintages.
imports_HS6*yyyy*_effdate.dta exports_HS6*yyyy*_effdate.dta	Effective dates of additional import or export duties at the HS6-country level for HS classification system of various vintages. Used to construct static tariff indicators.
imports_HS6*yyyy*_effdate_anycntry.dta	Effective dates of additional import duties at the HS6 level for HS classification system of various vintages. Used to construct static tariff indicators at the item level indicating the first instance of an additional duty for any country under various tariff programs.
FATUS_HS6*yyyy*.dta	Agricultural classification for imports and exports collected from the USDA Foreign Agricultural Service's Global Agricultural Trade System. Classifications are available at the HS10 level and are aggregated to the HS6 level based on whether 50% or more of the underlying HS10 codes (unweighted) are classified as agriculture. Almost all HS6 codes have either all or no underlying HS10 codes classified as agricultural.
rauch_HS5*yyyy*.dta	Rauch (1999) classifications of goods into those traded on an exchange (organized), had prices listed in trade publications (reference), or were brand name products (differentiated). We map the classifications originally

	made based on SITC2 (Rev 2) codes to HS5 codes for use with our data.
macro_sources.dta	Lists all the sources and data series identifiers, for either Haver or the BIS database, for the exchange rate and aggregate price index data included in the macro_all.dta file.

Instructions for BLS replication

1. Download and extract the replication package to a machine that contains BLS IPP data.
2. Construct BLS IPP data files required as described above and save in DATA/IPP subfolder.
3. Change line 15 of MAIN_BLS.do to reflect the relevant local computer directories in which the replication package is saved.
4. Run MAIN_BLS.do.
5. Figures will be displayed and regression output results will be printed in log_run_BLS.txt.

***Results for border prices cannot be reproduced without access to the restricted access BLS IPP data. This includes Figures 1, A1, A2, and A3 as well as Tables 1 (columns 1-6), A2, A3, A4, and A5.**