README: "Trade with Correlation" Replication

Overview

The code in this replication package produces all of the results and figures in the paper and online appendix. The file `code/replication.jl` generates the 11 figures and 12 tables by running scripts in `code/`. The replicator should expect the code to run for about an hour if the model is not re-estimated. A full re-estimation requires repeatedly running the estimation algorithm from a large number of random initial conditions, as described below, with each repetition taking about 3-4 hours.

Data Availability and Provenance Statements

Statement about Rights

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

Summary of Availability

- ☑ All data **are** publicly available.
- ☐ Some data **cannot be made** publicly available.
- ☐ **No data can be made** publicly available.

Details on each Data Source

Data.Name	Data.Files	Location	Provided	Citation
"World Input-Output Database Release 2013"	wiot_full.dta	data/WIOT/	Yes	Timmer et al. [2015]
"COMTRADE"			No	UN [2023]
"World Tariff Database"	$year.csv for year = 1999, . . . , 2007	data/tariffs/	Yes	Caliendo et al. [2023]
"Gravity Database"	cepii.csv	data/cepii/	Yes	Conte et al. [2022]
"Harmonized System to ISIC Revision 3 Concordance"	HS_to_I3.csv	data/sectors/	Yes	World Bank [2023]
"Harmonized System to SITC Revision 2 Concordance"	HS_to_I2.csv	data/sectors/	Yes	World Bank [2023]
"SITC Nomenclature"	S2_descriptions.csv	data/sectors/	Yes	World Bank [2023]

We use the 2013 release of the World Input Output Database (WIOD) data [Timmer et al., 2015], which is available at https://www.rug.nl/ggdc/valuechain/wiod/wiod-2013-release. The file `data/WIOT/wiot_full.dta` contains the full input-output tables. The data are in the public domain.

We construct 4-digit SITC revision 2 trade flow data from underlying COMTRADE data [UN, 2023]. For each year from 1984 to 2007, the file `data/comtrade/$year.csv` corresponds to the merge of "dot=1" (importer reporting) and "dot=2" (exporter reporting) data sets underlying the World Trade Flow (WTF) database (https://cid.ucdavis.edu/world-trade-flows), which are constructed using

the procedure available at http://www.robertcfeenstra.com/data.html. The data are in the public domain.

We use an early version of the Worldwide Tariff Database (WTD) [Caliendo et al., 2023]. The file `data/tariff/$year.csv` for year = 1999,...,2007 contains this data. This preliminary version has since been updated to version 1, which is publicly available at https://rcfeenstra.github.io/CFRT/Tariff-Data.html.

Sector concordances and descriptions come from the World Bank's World Integrate Trade Solution (WITS) website. Concordances are available at https://wits.worldbank.org/product_concordance.html and descriptions (nomenclatures) are available at https://wits.worldbank.org/referencedata.html. The data are in the public domain.

Dataset list

Data file (in `data/`)	Source	Notes	Provided
`WIOT/wiot_full.dta`	WIOD		Yes
`comtrade/$year.csv`	COMTRADE	year = 1984,...,2007	Yes
`tariff/$year.dta`	WTD version 0	year = 1999,...,2007	Yes
`cepii.csv`	CEPII		Yes
`sectors/HS_to_I3.csv`	WITS		Yes
`sectors/HS_to_S2.csv`	WITS		Yes
`sectors/S2_descriptions.csv`	WITS		Yes

All other files in `data/` are generated from these files as described below.

Computational requirements

Software Requirements

- Julia 1.4.2
  - Available at https://julialang.org/downloads/oldreleases/.
  - The files `Manifest.toml` and `Project.toml` provide all information necessary for replication. They get called on line 13 of `code/replication.jl` to setup the environment with all required packages.

Controlled Randomness

The estimation routines called by `estimateLFM` and `estimateSGM` on lines 60 and 61 of `code/replication.jl` use random initial conditions. Since the estimation algorithm has no guarantee of global convergence, we repeated the estimation from many random initial conditions (called on lines 28 to 31 and lines 62 to 65 of `code/steps/C1_estimateLFM.jl`). The directory `results/estimates/LFM/` includes subdirectories for each specification we estimate (e.g. `results/estimates/LFM/1/` corresponds to estimates of the LFM with 1 latent factor), and within each of these subfolders the folder `bestrep` has the results for the best repetition that we found. To rerun the estimation, change `false` on line 59 of `code/replication.jl` to `true`. Running this code will add additional repetitions within the directory `results/estimates/LFM/`. After adding additional repetitions, the call to `selectOrderAndSave` on line 68 of `code/replication.jl` will find the best repetition contained in `results/estimates/LFM/` and update `results/estimates/LFM/bestReps.csv`.

Memory and Runtime Requirements

Summary Approximate time needed to reproduce the analyses on a standard (2023) desktop machine **without re-running LFM estimation** is 30-40 minutes. Approximate time to perform one repetition of the LFM estimation algorithm for all specification is 3 to 4 hours.

Details The code was last run on a **desktop running Windows 11 with a 12-core 2.10 GHz Intel processor and 32 GB of RAM**.

Description of programs/code

In addition to `Manifest.toml` and `Project.toml`, which contain information on the Julia environment and required packages, and LICENSE.txt, which contains information on data licenses, the replication package consists of three directories containing code (`code/`), data (`data/`), and results (`results/`).

- `code/replication.jl` setups the Julia environment and runs the full replication by calling the other scripts within `code/`.
- The directory `code/steps/` contains a sequence of scripts, ordered via prefixes as follows. The initial letter in the file name corresponds to a group of files from a particular stage of the analysis (A = data preparation, B = reduced form evidence, C = estimation and specification testing, D = summary statistics of estimation results, and E = counterfactuals). Within each stage, the files are ordered by an integer indicating the order in which they should be run. Finally, following the letter indicating the stage of analysis and this integer, each file has a brief name.

  Description of files in `code/steps/`:

  1. `A1_prepareData.jl` defines several functions and a type for preparing and organizing data. The function `prepareDataFiles`, called on line 46 of `code/replication.jl`, loads the 4-digit trade and tariff data and the WIOT data, and then creates the files `data/secData.csv`, `data/sitcData.csv`, `data/sitcToSec.csv`, and `data/sitcCodes.csv`. The callable type `ProjData`, called on line 47 of `code/replication.jl`, is used for organizing the data.
  2. `A2_prepareUseShares.jl` defines the function `prepareUseShares`, called on line 51 of `code/replication.jl`. This function loads `data/WIOT/wiot_full.dta`, calculates final and intermediate use shares, and then saves them to `data/finalUseShares.csv` and `data/intermediateUseShare.csv`.
  3. `B1_reducedFormEvidence.jl` defines the function `reducedFormEvidence` and calling this function creates reduced form results in the Online Appendix as well as estimates CES and SGM specifications via PPML. This function generates Appendix Figure O.1 (saving the files `FigureO1a.png`, `FigureO1b.png`, `FigureO1c.png`, and `FigureO1d.png` to `results/onlineAppendix/`). It then saves estimates of the CES model to `results/estimates/CES.csv` and the SGM to `estimates/SGM.csv`. Next, it creates Appendix Table O.1, saving to `results/onlineAppendix/TableO1.csv` and `results/onlineAppendix/TableO1.tex`. Finally, it saves the indices of third party exposure to tariffs used in this analysis to `data/indices.csv`.
  4. `C1_estimateLFM.jl` defines the functions `estimateLFM` and `estimateSGM` which run the estimation algorithm for the LFM and LFM with SGM restrictions. *WARNING*: the estimation is numerically intensive and may take many hours to run. Lines 59 to 62 of `code/replication.jl` performs a single run of the estimation algorithm per specification. The results are saved in `results/estimates/LFM/`. This folder contains a subfolder for each specification, and within each of these subfolders it generates a subfolder for separate repetitions of the algorithm. For each repetition, 5 files are saved: `eta.csv` contains estimates of factor-level elasticities; `history.csv` contains a history of the algorithm's

iterations including data on the poisson deviance, and the R squared; `Lambda.csv` contains estimates of the factor weights; `mValues.csv` contains a key to the ordering of the latent factor estimates across origins, destinations, and years; and `Phi.csv` contains latent factor estimates. Note that within this folder we provide the best (lowest deviance) repetition that we found for each specification. As a result, it is not necessary to run lines 59 to 62 in `code/replication.jl` before running the rest of the file. Since running these lines is computationally intensive, the default is to skip this step. To run a single repetition of the estimation, change `false` on line 59 of `code/replication.jl` to `true`.

5. `C2_selectOrderAndSave.jl` defines the function `selectOrderAndSave` which uses the LFM estimation outputs to find the best repetition of the algorithm for each specification and saves the name of this repetition in `results/LFM/bestReps.csv`. It then performs sequential likelihood ratio tests to determine the number of factors and produces `paper/Table1.csv`, `paper/Table1.tex`, `onlineAppendix/TableO2.csv`, and `onlineAppendix/TableO2.tex`. Finally, it saves the implied elasticities and factor weights for the 14-factor model with weights restricted to correspond to the SGM model at the WIOT level, and the implied elasticities, factor weights and factor-level expenditure of the preferred (7-factor) LFM specification.
6. `C3_standardErrors.jl` defines the function `standardErrors` which creates the CSV file `results/estimates/LFM/vcovMatLFM.csv` containing an estimate of the variance-covariance matrix of the elasticities of the 7-factor LFM, and `results/estimates/LFM/elasticitiesWithSEs.csv` containing the elasticity estimates along with their standard errors.
7. `C4_secondStep.jl` defines the function `estimateSecondStep` which provides alternative estimates of the theta parameter. It saves these results for the 7-factor LFM to `results/onlineAppendix/TableO4.csv` and `results/onlineAppendix/TableO4.csv`. Additionally, it outputs to the REPL analogous results for the SGM.
8. `D1_results.jl` defines the function `computeResults` which creates tables and figures reporting estimates and summary statistics (see below).
9. `D2_goodnessOfFit.jl` defines the function `goodnessOfFit` which creates tables and figures related to the model's goodness of fit (see below).
10. `D3_elasticities.jl` defines the function `computeElasticities` which computes substitution elasticities for the LFM and SGM models and saves them to the CSV file `results/estimates/substitutionElasticities.csv`. Then, it creates figures related to these elasticities (see below).
11. `E1_counterfactuals.jl` defines the function `computeCounterfactuals`, which creates tables and figures related to counterfactual results (see below).
12. `E2_gainsFromTrade.jl` defines the function `computeGainsFromTrade` which creates tables and figures for results on the gains from trade.

- The directory `code/tools/` contains contains scripts that define functions that are called throughout the analysis.
- The directory `data/` contains the following files and folders:
  1. `cepii/cepii_data.dta` contains gravity covariates from CEPII.
  2. `comtrade/` contains CSV files with trade flow data across 4-digit SITC revision 2 sectors, importers, and exporters by year.
  3. `sectors/` contains concordances between levels of sectoral aggregation along with descriptions.
  4. `tariff/` contains CSV files with tariff data across 4-digit SITC revision 2 sectors, importers, and exporters by year.
  5. `WIOT/wiot_full.dta` contains the full tables from the World Input-Output Database.
  6. `countryOrder.csv` contains an ordering of countries in terms of 1999 GDP per capita from lowest to highest. GDP per capita data comes from `cepii_data.dta`. It is generated

by `prepareUseShares` called on line 51 of `code/replication.jl`.

7. `finalUseShares.csv` contains final use shares for aggregated WIOT sectors by country and year. It is generated by `prepareUseShares` called on line 51 of `code/replication.jl`.
8. `gravityData.csv` contains a combination of trade flow, tariff, and gravity covariate data for gravity regressions. It is generated by calling `prepareGravityData` defined in `code/steps/A1_prepareData.jl` and called by `prepareDataFiles` on line 46 of `code/replication.jl`.
9. `indices.csv` contains indices of third party tariff exposure. It is generated by calling the function `reducedFormEvidence` called on line 55 of `code/replication.jl`.
10. `intermediateUseShares.csv` contains intermediate use shares for pairs of aggregated WIOT sectors by country and year. It is generated by `prepareUseShares` called on line 51 of `code/replication.jl`.
11. `secData.csv` contains trade data at the WIOT sector level across origins, destinations, and years. It is generate by calling `prepareDataFiles` on line 46 of `code/replication.jl`.
12. `secNames.csv` contains names of WIOT sectors. Generated by `code/tools/secNames.jl`, which is called on line 42 of `code/replication.jl`.
13. `sitcCodes.csv` contains codes from 4-digit SITC sectors (including residuals). It is generate by calling `prepareDataFiles` on line 46 of `code/replication.jl`.
14. `sitcData.csv` contains trade data at the 4-digit SITC sector level across origins, destinations, and years. The ordering for sectors matches the ordering in `sitcCodes.csv`. It is generate by calling `prepareDataFiles` on line 46 of `code/replication.jl`.
15. `sitcToSec.csv` provides a correspondence between 4-digit SITC and WIOT sectors. It is generate by calling `prepareDataFiles` on line 46 of `code/replication.jl`.
16. `wiodToSec.csv` provides a correspondence between raw WIOT data and the level of aggregation used in the other files in this folder. It is generated by calling `prepareDataFiles` on line 46 of `code/replication.jl`.

Instructions to Replicators

1. Edit line 10 of `code/replication.jl` to adjust the current directory to the location of the replication folder on your computer.
2. If re-estimating the LFM, replace `false` on line 59 of `code/replication.jl` with `true`. Doing so will increase computation time by approximately 3 to 4 hours on a standard 2023 desktop computer, produce one additional repetition of the estimation algorithm for all specifications, and save the results to `results/estimates/LFM/`. To do, say, 10 additional repetitions replace all of line 59 in `code/replication.jl` with `for rep = 1:10`.
3. Run `code/replication.jl` to run all steps in sequence.

List of tables and programs

The provided code reproduces:

☐ 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.

Figure/Table #	File (in `code/steps/`)	Line #	Output File (in `results/paper/`)
Figure 1.A.	D1_results.jl	277	`Figure1a.pdf`
Figure 1.B.	D1_results.jl	282	`Figure1b.pdf`
Figure 1.C.	D1_results.jl	327	`Figure1c.pdf`

Figure/Table #	File (in `code/steps/`)	Line #	Output File (in `results/paper/`)
Figure 1.D.	D1_results.jl	329	`Figure1d.pdf`
Figure 2.A.	E2_gainsFromTrade.jl	99	`Figure2a.pdf`
Figure 2.B.	E2_gainsFromTrade.jl	119	`Figure2b.pdf`
Figure 3.A.	D3_elasticities.jl	67	`Figure3a.pdf`
Figure 3.B.	D3_elasticities.jl	73	`Figure3b.pdf`
Figure 4.A.	D3_elasticities.jl	84	`Figure4a.pdf`
Figure 4.B.	D3_elasticities.jl	95	`Figure4b.pdf`
Figure 5.A.	E2_gainsFromTrade.jl	130	`Figure5a.pdf`
Figure 5.B.	E2_gainsFromTrade.jl	141	`Figure5b.pdf`
Figure 5.C.	E2_gainsFromTrade.jl	179	`Figure5c.pdf`
Figure 5.D.	E2_gainsFromTrade.jl	190	`Figure5d.pdf`
Figure 6.A.LFM	E1_counterfactuals.jl	280	`Figure6A_LFM.pdf`
Figure 6.A.SGM	E1_counterfactuals.jl	288	`Figure6A_SGM.pdf`
Figure 6.B.LFM	E1_counterfactuals.jl	303	`Figure6B_LFM.pdf`
Figure 6.B.SGM	E1_counterfactuals.jl	314	`Figure6B_SGM.pdf`
Table 1.	C2_selectOrderAndSave.jl	170	`Table1.tex`
Table 2.	D1_results.jl	226	`Table2.tex`
Table 3.	D1_results.jl	243	`Table3.tex`

Figure/Table #	File (in `code/steps/`)	Line #	Output File (in `results/onlineAppendix/`)
Figure O.1.A.	B1_reducedFormEvidence.jl	73	`FigureO1a.png`
Figure O.1.B.	B1_reducedFormEvidence.jl	78	`FigureO1b.png`
Figure O.1.C.	B1_reducedFormEvidence.jl	87	`FigureO1c.png`
Figure O.1.D.	B1_reducedFormEvidence.jl	98	`FigureO1d.png`
Figure O.2.	D1_results.jl	266	`FigureO2.pdf`
Figure O.3.	D1_results.jl	343	`FigureO3.pdf`
Figure O.4.A.	D3_elasticities.jl	116	`FigureO4a.png`
Figure O.4.B.	D3_elasticities.jl	119	`FigureO4b.png`
Figure O.5.	E1_counterfactuals.jl	167	`FigureO5.pdf`
Table O.1.	B1_reducedFormEvidence.jl	177	`TableO1.tex`
Table O.2.	C2_selectOrderAndSave.jl	161	`TableO2.tex`
Table O.3.	D2_goodnessOfFit.jl	220	`TableO3.tex`
Table O.4.	C4_secondStep.jl	57	`TableO4.tex`
Table O.5.	D1_results.jl	364	`TableO5.tex`
Table O.6.	D1_results.jl	323	`TableO6.tex`
Table O.7.	D1_results.jl	381	`TableO7.tex`
Table O.8.	E2_gainsFromTrade.jl	151	`TableO8.tex`
Table O.9.	E1_counterfactuals.jl	171	`TableO9.tex`

References

Lorenzo Caliendo, Robert C Feenstra, John Romalis, and Alan M Taylor. Tariff reductions, heterogeneous firms, and welfare: Theory and evidence for 1990–2010. *IMF Economic Review*, 71(4): 817–851, 2023. URL https://rcfeenstra.github.io/CFRT/Tariff-Data.html.

Maddalena Conte, Pierre Cotterlaz, Thierry Mayer, et al. The cepii gravity database. (Working Paper 2022-05), 2022. URL http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=8.

Marcel P Timmer, Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J De Vries. An illustrated user guide to the world input–output database: the case of global automotive production. *Review of International Economics*, 23(3):575–605, 2015. URL https://www.rug.nl/ggdc/valuechain/wiod/wiod-2013-release.

UN. Comtrade database. 2023. URL https://comtradeplus.un.org/.

World Bank. World integrated trade solution. 2023. URL https://wits.worldbank.org.