Description of data and code files for "Bilateral Economies of Scope"
Authors: Yao Amber Li, Stephen Yeaple, Sichuang Xu and Tengyu Zhao
Forthcoming at the Review of Economics and Statistics

Our analysis exploits confidential matched data from Chinese Customs Trade Statistics and China's Annual Survey of Industrial Enterprises. While we are unable to make these confidential data publicly available, we include all our regression codes and simulation programs in the replication package, which other researchers with access to the data could use to replicate our work.

The **Data** folder contains the public data used in the analysis, while the **Program_codes** folder contains the scripts used to perform the analysis. The data and replication instructions below provide a step-by-step guide to reproduce the results of the original study.

Data Sources

The main firm-level data used in this paper is **confidential and proprietary** data. The original data on manufacturing survey and the trade data from Chinese Customs are obtained in collaboration with NBSC (National Bureau of Statistics of China) through their authorized vendor:
Beijing IVG Information Consulting Co., Ltd.
Website: www.china-ivg.com
Tel: 8610-63806335; 8610-83509152.
Fax: 8610-83509185.
Contact: Ms. SONG Jia. Email: marketing@china-ivg.com
Address: CUV International Building, Wanliu Qiao, Fengtai District, Beijing, China.
Postal Code: 100070

Programs and Codes

Empirical analysis:
+ Software Required: Stata/MP (17.0 or later)
+ Operating System: Windows 10
+ Raw data:
  + the Chinese Customs Trade Statistics (2000-2015);
  + the Annual Survey of Industrial Enterprise (1998-2009);
  + the CEPII Gravity Database.
+ Stata do files (in order):
  + **0_data_preparation.do**: Aggregate customs data to firm-country level, using the raw data sets;
  + **1_fact1.do**: Calculate the conditional ratios in Fact 1.
  + **2_reduced_form_fact2.do**: Perform sample construction and conduct reduced-form tests related to Fact 2.
  + **3_quantitative_prep.do**: Generate targeted and non-targeted moments file: moments.xls (or renamed as momentsTop30New1.xls), for quantitative analysis.
    + **getGravity.do**: Extract gravity variables from CEPII file: grav_200.dta, and generate a

consistent country set in year 2007;

+ **getMoments.do**: Calculate targeted and non-targeted moments in Table 9 and generate bootstrap moments;

+ **getSourcingExportingPotential.do**: Estimate sales potentials and sourcing potentials in Figure 4;

+ **getSigmaTheta.do**: Estimate parameters, sigma and theta, in Table 7;

+ **getCFs.do**: Calculate counterfactual changes in sales and sourcing potentials for Section 6;

+ **getHierarchy.do**: Calculate the two-sided hierarchy structure in Table A12.

+ **4_rank_rank_fact3.do**: Generate rank-rank correlation in Fact 3.

Quantitative analysis:

+ Software Required: MATLAB (R2024a or later), gcc, openmp
+ Operating System: CentOS 7
+ Scripts:
  + **aglobalConfig.m**: This script sets the global configuration for the model.
  + **asolveAllModels.m**: This script solves the model for all parameter sets.
    + set default options
    + set parameters
    + get baseline
    + get exp induce imp model
    + get imp induce exp model
    + get restricted model
    + trade liberalization (for all four types of models)
    + simulated method of moments
    + save model results
  + **atableFigureAllModels.m**
    + table & figures for the main text
      + 01. table: baseline parameter estimates: atabFigTableParBaseline.m
      + 02. table: baseline model fit: atabFigTableMomentFit.m
      + 03. figure: jacobian matrix: atabFigFigureSMMHessian.m
      + 04. table: parameter assignments and moments for the alternative/restricted models: **atabFigTableParAlternative.m**
      + 05. figure: rank+rank figure: atabFigFigureRankRank.m
      + 06. table: extensive margin of trade liberalization: **atabFigTableDecompositionOfTradeShock.m**
      + 07. table: firms response to trade liberalization: atabFigFigureDecomposition.m
      + 08. figure: state+space trade policies: atabFigFigurePunchFixed.m
      + 09.figure: state+space trade policies (use baseline deviation): **atabFigFigurePunchFixedDevBaseline.m**
    + table & figures for the appendix
      + 10. table: monte carlo: atabFigTableMonteCarlo.m
      + 11: figure: rank+rank for overall firms: atabFigFigureRankRank.m
      + 12: table: correlation of fixed cost draws: atabFigTableCorrelation.m

+ 13: table: hierachy entry patterns: ataFigTableAFTHir.m

+ 14: table: alternative sourcing & sales potential: drop commodity and homogeneos goods: **atabFigTableAlterSourcingPotent.m**

+ 15: table: pareto shape: atabFigTableParetoShape.m

+ 16: figure: residual plot: atabFigFigureRankRankResidual.m

+ 17: table: same \\alpha: atabFigTableSameAlpha.m

+ 18: table: full moments: atabFigTableParAlternativeFull.m

+ 19: figure: rank+rank for full moments: atabFigFigureRankRankFull.m

+ Helper Functions:

+ **fimsearchbnd.m**: This function is used to perform the optimization with parameter bounds.

+ **extStat**.m: This function is used to calculate the moments of the data (or other by group statistics)

+ **extWtmean.m**: This function is used to calculate the weighted mean of the data

+ **extWtPrctile.m**: This function is used to calculate the weighted percentile of the data

+ **extPrint.m**: This function is used to print the results of the model

+ Replication Steps:

+ Step 1: Set the global configuration for the model by running the script aglobalConfig.m.

+ Step 2: Solve the model for all parameter sets by running the script asolveAllModels.m.

+ Step 3: Generate tables and figures for the main text and appendix by running the script **atableFigureAllModels.m**.

+ Note: The code was run on an HPC cluster at the Chinese University of Hong Kong (Shenzhen) using MATLAB (R2024a) and C (gcc+openmp) with 40 logical cores and 512GB of memory. On this platform, Step1 to Step3 took approximately 42 hours to complete, and the author does not guarantee the performance of the code on other platforms.