README

Scott Orr* Mokhtar Tabari†

May 23, 2025

Overview

This replication package accompanies the paper *Decomposing the Within-Firm Productivity Gains from Trade: Evidence from India*. It includes all the raw data (except Prowess), cleaned datasets, Stata and R scripts, and auxiliary files necessary to fully replicate the analysis and results presented in the paper from the original data.

Folder Structure

- `01_Master_Code.do`: Main script that orchestrates the entire replication pipeline.
- `code/`: Contains all data preparation and estimation scripts.
  - `cleaning/`: Scripts for cleaning raw datasets and constructing variables.
  - `estimation/`: Scripts for running the empirical analysis and generating results.
- `data/`:
  - `raw/`: Contains the original raw data and documents.

*Sauder School of Business, University of British Columbia: scott.orr@sauder.ubc.ca

†University Canada West: mokhtar.tabari@ucanwest.ca

  - `clean/`: Final cleaned datasets used for analysis.
    * Created when you run `01_Master_Code.do`
  - `working/`: Intermediate and estimation-ready datasets generated during replication.
    * Created when you run `01_Master_Code.do`
- `figures/`: Any figures generated in the paper
  - Created when you run `01_Master_Code.do`
- `tables/`: Contains LaTeX tables used in the manuscript
  - Created when you run `01_Master_Code.do`

How to Replicate

1. Open `01_Master_Code_2024.do` in Stata.
2. Adjust any file paths at the top of the script.
3. Run the script sequentially. It will:
   (a) Clean and merge raw datasets.
   (b) Generate analysis variables.
   (c) Estimate regressions and produce tables and figures.
4. Some cleaning scripts use R (e.g., `"06_clean_prowess_wits.R"` or `"create_figb1.R"`). These must be run manually or integrated via Stata's `shell` command if desired. Also, ensure to adjust the paths inside these R files.

Individual Figures and Tables

Figures are produced in the "`figures/`" folder. Below, we reproduce the names of the figure(s) corresponding to each of numbered figures in the main text and online appendix.

- **Figure 1** SVA_decomp_y_2O1I.png, SVA_decomp_yn_2O1I.png, SVA_decomp_ycn_2O1I.png, Input_access_focus.png
- **Figure B.1** figb1a.png, figb1b.png
- **Figure F.1** SVA_percent_percent_effects.png
- **Figure G.1** GO_decomp_y_2O1I.png, GO_decomp_yn_2O1I.png, GO_decomp_ycn_2O1I.png
- **Figure G.2** GO_decomp_rshare_y_2O1I.png, GO_decomp_rshare_yn_2O1I.png, GO_decomp_rshare_ycn_2O1I.png
- **Figure G.3** Input_access_comparison.png
- **Figure G.4** SVA_decomp_ad_robust_b2_add.png, SVA_decomp_ad_robust_b2_drop.png, SVA_decomp_ad_robust_b3_add.png, SVA_decomp_ad_robust_b3_drop.png, SVA_decomp_ad_robust_add_I_b2.png, SVA_decomp_ad_robust_drop_I_b2.png
- **Figure G.5** SVA_decomp_interactions_y_2O1I_size1_nobs.png, SVA_decomp_interactions_y_2O1I_size2_nobs.png, SVA_decomp_interactions_yn_2O1I_size1_nobs.png, SVA_decomp_interactions_yn_2O1I_size2_nobs.png, SVA_decomp_interactions_ycn_2O1I_size1_nobs.png, SVA_decomp_interactions_ycn_2O1I_size2_nobs.png
- **Figure G.6** SVA_decomp_l_interactions_y_2O1I_size1_nobs.png, SVA_decomp_l_interactions_y_2O1I_size2_nobs.png, SVA_decomp_l_interactions_yn_2O1I_size1_nobs.png,

SVA_decomp_l_interactions_yn_2O1I_size2_nobs.png,
SVA_decomp_l_interactions_ycn_2O1I_size1_nobs.png,
SVA_decomp_l_interactions_ycn_2O1I_size2_nobs.png

- **Figure G.7** robustness_SVA_y2.png, robustness_SVA_y3.png, robustness_SVA_yn2.png, robustness_SVA_yn3.png, robustness_SVA_ycn2.png, robustness_SVA_ycn3.png
- **Figure G.8** robustness_GO_y2.png, robustness_GO_y3.png, robustness_GO_yn2.png, robustness_GO_yn3.png, robustness_GO_ycn2.png, robustness_GO_ycn3.png
- **Figure G.9** robustness_SVA_Iy2.png, robustness_SVA_Iyn2.png, robustness_GO_Iy2.png, robustness_GO_Iyn2.png

Similarly, Tables are produced in the "`tables/`" folder. Below, we reproduce the names of the tables(s) corresponding to each of numbered figures in the main text and online appendix.

- **Table 1** sumstat.tex
- **Table 2** firmtable_SVA_decomp.tex
- **Table B.1** sample_products.tex
- **Table B.2** sample_comparisons.tex
- **Table C.1** production_parameter_new_kl.tex
- **Table D.1** production_parameter_new_klme.tex
- **Table D.2** TFPQ_IV_Demand.tex
- **Table D.3** TFPQ_IV_Demand_FS.tex
- **Table D.4** demand_parameters.tex
- **Table E.1** firmtable_SVA_all.tex

- **Table F.1** FS_SVA.tex
- **Table G.1** P_Regs.tex
- **Table G.2** Q_Regs.tex
- **Table G.3** SumAddDrop.tex
- **Table G.4** welfare_regs.tex
- **Table G.5** welfare_CF.tex
- **Table G.6** VarianceDecomp.tex
- **Table G.7** WithinVariance_Regs.tex.tex
- **Table H.1** output_elasticities_sva.tex
- **Table H.2** output_elasticities_go.tex
- **Table H.3** elasticities_by_ind_thetas_R_kl.tex
- **Table H.4** elasticities_by_ind_thetas_R_kl_3lags.tex
- **Table H.5** elasticities_by_ind_thetas_Q_kl.tex
- **Table H.6** elasticities_by_ind_thetas_CS_kl_nic4.tex
- **Table H.7** elasticities_by_ind_thetas_CS_kl_year_nic2.tex
- **Table H.8** elasticities_by_ind_thetas_Q_klme.tex
- **Table H.9** elasticities_by_ind_thetas_Q_klme_2lags.tex
- **Table H.10** elasticities_by_ind_thetas_CS_klme_nic4.tex
- **Table H.11** elasticities_by_ind_thetas_CS_klme_year_nic2.tex
- **Table H.12** elasticities_by_ind_thetas_Q_klm.tex

Software Requirements

- **Stata 15 or later** with the following user-written packages:

  `rsource, reclink2, sencode, texsave, egenmore, unique, ftools, reghdfe, ivreg2, ranktest, ivreghdfe, estout`

- **R** (for specific cleaning steps and to create figure B1).

Data Access

All raw data in `data/raw/` are either:

- Publicly available (e.g., Comtrade, WITS, emission factor PDFs, and IO tables),
- Purchased/licensed datasets (e.g., CMIE Prowess), or
- Hand-coded files included for transparency.

Due to licensing restrictions, we cannot redistribute any of the purchased or licensed datasets used in this study. Researchers wishing to replicate our results can obtain their own access to these datasets through the appropriate licensing agreements. In particular, access to the Prowess database can be arranged directly with the Centre for Monitoring Indian Economy (https://prowess.cmie.com/). We use data from the following Prowess modules in this study: Identity, Standalone Annual Financial Statements, Products Produced Traded, Productwise Energy Consumption, Raw Materials Consumed, and Energy Consumption.[1] These files need to be moved to the `data/raw/prowess/` folder before

[1]The Standalone Annual Financial Statements have too many variables to be extracted in one data request. We extracted all variables from these datasets in nine separate data requests, generating nine datasets with up to 500 variables each; we name these datasets:

1. "Standalone Annual Financial Statements_1_500.dta"
2. "Standalone Annual Financial Statements_501_1000.dta"

replication.

Acknowledgements

Please cite this replication package if using it for academic or teaching purposes. For questions, contact scott.orr@sauder.ubc.ca.

3. “Standalone Annual Financial Statements_1001_1500.dta”
4. “Standalone Annual Financial Statements_1501_2000.dta”
5. “Standalone Annual Financial Statements_2001_2500.dta”
6. “Standalone Annual Financial Statements_2501_3000.dta”
7. “Standalone Annual Financial Statements_3001_3500.dta”
8. “Standalone Annual Financial Statements_3501_4000.dta”
9. “Standalone Annual Financial Statements_4000_4269.dta”

To run our master code without further modification, extract all variables from The Standalone Annual Financial Statements sequentially in groups of 500 and use the same final dataset names enumerated above.