Per Capita Income and the Demand for Skills
- Readme for replication package -

Justin Caron, Thibault Fally and James Markusen*

December 2019

Abstract

This document describes the content of the replication package attached to "Per Capita Income and the Demand for Skills" to be published in the Journal of International Economics. The estimation of model parameters, as well as model calibration and simulation are all implemented using the GAMS modelling language. Compilation of results and graph-making are done in STATA. This note describes how to replicate all results and robustness checks.

*Justin Caron: Department of Applied Economics, HEC Montreal. Thibault Fally: Department of Agricultural and Resource Economics, University of California Berkeley; James Markusen: Shandong University and University of Colorado-Boulder.

Introduction

Estimation and simulation are implemented in the GAMS (general algebraic modeling system) modeling language. The GAMS programs and STATA do-files included in the package allow the reader to replicate the results of the paper. Result generation is mostly automated except for manually copying results from GAMS into STATA. The user can use the **runall.bat** batch file to follow the workflow.

The most important files are **demand_gtap5.gms**, which describes the estimation of demand parameters, and **CF_skill.gms**, which describes the general equilibrium model for simulation.

In the following, we describe the list of GAMS programs, do-files and data files contained in the replication folder.

Data

All of the data used in this paper are freely available.

GTAP: The main estimation and simulation exercise is based on the GTAP (Global trade analysis project) version 5 dataset (with robustness checks using GTAP8). GTAP5 and GTAP8 are freely available from the GTAP website at www.gtap.agecon.purdue.edu.

Extraction of the raw GTAP5 data and translation into GAMS format relies on the GTAPinGAMS code developed by Thomas Rutherford (University of Wisconsin), available at www.mpsge.org/gtap5/. The following describes the steps required for the creation of the GAMS-ready GDX file (`GAMS/data/gtap5.gdx`) used for analysis.

- Download the GTAP data from the GTAP website at www.gtap.agecon.purdue.edu/databases/v5/default.asp Chose "Download GTAPAgg 5.4". Also download the appropriate GTAP licence file at "GTAPAgg License" and copy the `gtapagg.lic` file to the `GTAPagg` folder.
- Run `gtapagg.exe` to extract the whole dataset. We don't aggregate any of the dimensions here. Thus click on "View/change regional aggregation" and chose the "1 to 1" option. Do the same for "sectoral aggregation" and "factor aggregation". Create the aggregated database.
- `gtapagg.exe` creates a ZIP folder including a number of datafiles in the *HAR* format. Copy this ZIP file to the `GAMS/data` folder
- Then run the `make-GTAPAGG.bat` batch file from the */GAMS/data/GTAP data extraction* folder. This successively runs the `ZIP2GDX.gms` file (extracts data from the ZIP folder into a GAMS-readable *GDX* file) and the `FILTER.GMS` file which filters out extremely small values in the GTAP data (this step isn't strictly necessary).
- Output is the GAMS-useable `GTAP5.gdx` file stored in the `/GAMS/Data` folder.

Other data sources:

- Bilateral trade cost determinants. From CEPII. Imported into the `GAMS/data/tradecostimp.gdx` file.
- Rates of GDP and export growth. From the Penn Tables (PWT9). Imported into the `GAMS/targets.gms` file
- Historical changes in the skill premium. From WIOD. Extracted with the `STATA/WIOD/WIOD_skill premium extraction.do` do-file.
- Historical changes in years of schooling (proxy for skill endowment) from Barro and Lee (2005). Imported into the `STATA/Barro lee 25_yr schooling.dta` file.

All of the above data sources are freely available and we have left them in the package to facilitate replication.

GAMS files for estimations and simulation

Running GAMS All GAMS programs have been used and tested with GAMS version 24.2.3 on a PC computer. We solve the estimation and simulation models with the following solvers:

- CONOPT non-linear solver for Maximum Likelihood estimation of gravity and demand equations. Other non-linear solvers should work. Alternatively NLLS estimation of gravity could be done using CPLEX or any solver accepting quadratic optimization.
- PATH non-linear solver. Used to solve the simulation model, which is a formulated as a mixed-complementarity (MCP) model. Other non-linear solvers may work.

You will need the appropriate licences.

The following batch file can be used to execute all GAMS files in the correct order:

- **Runall.bat** - a windows batch file (text editable) which runs all GAMS files necessary for estimation and simulation. The content of this file describes the whole workflow, up to the creation of GDX files to be copied to STATA. Executing this batch file in one go should replicate all estimation and simulation results but will take a long time. Alternatively, it can be used to follow the workflow piece-by-piece.

To run the batch file you will need to add GAMS to your computer's PATH file.

Data preparation and estimation

The following describes the files used for the estimation of gravity and demand parameters as well as the preparation of the parameters subsequently used to calibrate the simulation model.

These programs should be executed in the following order:

- **gravity_with_atc.gms** - Estimates the gravity parameters and implied Φ parameters (with asymmetric trade costs). The file estimates gravity Poisson ML but can also be configured to estimate an OLS model.
  - Output: `gravityestimates_gtap5_atc.gdx`
- **demand_gtap5.gms** - Estimates the CRIE demand system with different specifications for θ. Also inverts the IO matrix to compute total factor intensities and prepares all sector-level reporting parameters to be exported to STATA. The file must be run with the following options for the *SPEC* (specification) global variable (see batch file):
  - *tc*: the theta parameter is estimated from the Φ parameter (derived from trade costs). Benchmark specification.
  - *theta4*: θ is fixed at 4 in all sectors
  - *notc*: no trade costs ($\Phi = 0$ in all sectors)
  - *nobc*: no budget constraint
  - *reducedform*: a reduced form estimation with lambda replaced by a linear function of log per capita income.
  - Output: `estimates_gtap5_logweighted_%spec%_rall.gdx`
- **demand_gtap5_IV.gms** - Estimates CRIE demand system with Φ parameters instrumented using 'international only' Φ parameters, using 2-stage least-squares.
  - Output: `estimates_gtap5_logweighted_tc_rall_PHI_IV`
- **demand_clm_t4_gtap5.gms** - estimates non-homothetic CES preferences as in Comin et al. 2016. The file assumes $\theta = 4$ for all sectors as default.
  - Output: `estimates_gtap5_CLM.gdx`
- **datapreparation_GTAP5.gms** - prepares IO coefficients, trade shares, and other parameters to be subsequently passed on to the simulation simulation programs below. The file must be run for the following demand specifications, i.e. options for the *DEMANDEST* variable: *tc*, *theta4* and *tccsbeta*. The first two are with average input shares, the later with country-specific input shares.
  - Output: `DATA_gtap5_tc.gdx`, `DATA_gtap5_theta4.gdx` and `DATA_gtap5_tccsbeta.gdx`

General equilibrium simulation

- **Targets.gms** - Short file including the GDP and export/GDP growth rates used for targeting.
- **CF_skill.gms** - This is the main file containing the general equilibrium GE model which is used to simulate counterfactual changes in the skill premium based on CRIE preferences. This file also computes the partial equilibrium approximations of the skill premium and the decompositions of the growth and trade counterfactuals.

  The following global variables must be defined:

  - *SPEC*, defines demand specifications: *NH* (non-homothetic demand) or *H* (homothetic demand)
  - *DEMANDEST*, defines the demand estimation of θ: *TC* (default), *THETA4* or *TCCSBETA*
  - *TRADETARG* and *PRODTARG*, define trade and productivity targeting with options *NO* (no targeting), *YES* (endogenous targeting) and *LOAD* (loads productivity/trade adjustment parameters from unified counterfactual).

  The file must be run with the following options for these global variables (see batch file):

  - To run the UNIFIED counterfactual, set *TRADETARG* and *PRODTARG* to *YES* (targets GDP and export growth). This counterfactual will take several hours to solve.
  - To run the TRADE-ONLY counterfactual, set *TRADETARG* to *LOAD* and *PRODTARG* to *NO*
  - To run the GROWTH-ONLY counterfactual, set *PRODTARG* to *LOAD* and *TRADETARG* to *NO*

  For the robustness checks, the *DEMANDEST* global variable can be switched to *THETA4* and *TCCSBETA*.

  Results are exported to two places. Results used in the paper are exported to the `results/SPESTIMATES_TOMER` subfolder (respectively theta4 and theta4_clm). All the files in these folders can be merged to a merged.gdx file to facilitate copy-pasting into STATA. Merging is done in the runall.bat file and can also be done by typing "gdxmerge *.gdx" in the command line.

  All other results and parameters are exported to the `results/TC subfolder` (respectively theta4 and theta4_clm).

- **CF_skill_CLM.gms** - file including the same GE model as above, but with Comin et al. (2016) (CLM) preferences.

STATA do-files for graph and table generation

Two do-files are required to generate the paper's tables and graphs. These can be found in the STATA subfolder. The generated graphs are saved in the `STATA/Figures` sub-folder. Tables must be copied to excel to be formatted in order to obtain the exact tables reported in the paper.

- **Sector_correlations.do** - Compiles all 'sector-level' results:
  - Figures 1, A1, A2, A3
  - Tables 1, A4
  - Various correlation and regression coefficients discussed in the text.

  This file requires copying data from GAMS: manually copy the COEFFS parameter compiled in `data_preparation_gtap5.gms` and unloaded onto: `DATA_GTAP5_TC.GDX`. Copy with sectors as rows and parameters as columns. For alternative specifications (robustness checks), it is also necessary to copy income elasticity coefficients from the COEFFS parameter in the *IV*, *CLM*, *nobc*, *notc*, *theta4* and *reducedform* results files.

- **country_skill_premium_chg.do** - Compiles all 'sector-level' results:
  - Figures 2, 3, 4, 5, 6, 7, A4, A5, A8
  - Tables 2, A3 , A5
  - Comparison to WIOD estimates.
  - Various correlation and regression coefficients discussed in the text.

  This file requires manually copying data from GAMS (see the top of the file for full instructions):
  - Create a file with log pci, population and shocks from reporting from the file `RESULTS/TC/CF_Skill_prod_` Copy the REPORTING parameter, chosing the FINAL iteration (row = countries, cols = params). Save as `reporting_NH.dta`.
  - For the GROWTH-ONLY decompositions, create file with PE approximations. Copy the DECOMPOSITION parameter from the `RESULTS/TC/CF_Skill_prod_load_trade_no_NH.GDX` file (row = countries, cols = params). Save as `growthapprox_NH.dta`.
  - For the TRADE-ONLY decompositions, create file with decompositions. Copy the `LOG_CHG_SKILLPREM_AI` parameter from the `RESULTS/TC/CF_Skill_prod_no_trade_load_NH.GDX` file (row = countries, cols = params) (chose "T1" or any other). Save as `tradedecomp_NH.dta`.
  - Then do the same for homothetic preferences. Copy the `LOG_CHG_SKILLPREM_APPROX` parameter from the `RESULTS/TC/CF_Skill_prod_no_trade_load_H.GDX` file (row = countries, cols = params) (chose "T1" or any other). Save as `tradedecomp_H.dta`. clear

- For robustness checks, copy theta4 results. From `RESULTS/SPESTIMATES_TOMERGE_theta4/MERGED.GDX` file, copy the SKILLPREMIUM parameter (rows = countries, cols = specifications). Save as `skillpremium_t4.dta`. clear
- Copy Comin et al 2016 CLM results. From `RESULTS/SPESTIMATES_TOMERGE_theta4/MERGED.GDX` file, copy the SKILLPREMIUM parameter (rows = countries, cols = specifications). Save as `skillpremium_clm.dta`.
- Copy country-specific input shares (CSbeta) results. From `RESULTS/SPESTIMATES_TOMERGE_csbeta/MERG` file, copy the SKILLPREMIUM parameter (rows = countries, cols = specifications). Save as `skillpremium_csbeta.dta`.
- Finally create the main file in which all results are merged. From the `RESULTS/SPESTIMATES_TOMERGE_TC/` file. Copy the SKILLPREMIUM parameter (rows = countries, cols = specifications).

This files also loads historical skill premium data from WIOD and skill endowment data from Barro and Lee (2005).