Replication Files for "Trade, Migration and Productivity: A Quantitative Analysis of China"

The replication files include STATA do-files and MATLAB m-files to reproduce all main results. Specifically, Tables 4 and 6 are produced using STATA, and the included data files, while Tables 5, and 7-11 are produced using MATLAB, and the included data files.

Data Sets Included:

- migration_data.dta – includes the full province-sector matrix of migration shares in 2000 and 2005, the real income of each province-sector in 2000, the bilateral distance between provinces, and the three IVs used in estimating the income elasticity of migration. All variables are fully labelled.

- trade_data.dta – includes the bilateral flows in agriculture and non-agriculture between the 8 regions of China for 2002 and 2007. Also included are the bilateral trade shares, estimated trade costs, and exporter-specific costs used to construct the asymmetric trade costs used in the analysis. All variables are fully labelled.

- trade_ag.csv and trade_na.csv – the bilateral trade matrix for agriculture goods and for non-agricultural goods between provinces in 2002. This is used by MATLAB. Provinces are ordered according to the variable "region", which is from the employment_realGDP_data.csv file.

- mij2000.csv and mij2005.csv – the bilateral migration shares and non-migrant shares for each pair of province-sector (ag/nonag) pairs. The .csv is a vector. It is reshaped in the MATLAB code when relevant.

- tauhat.csv – the measured change in migration costs

- employment_realGDP_data.csv – provides provincial real GDP per worker in agriculture and nonagriculture in 2000 and 2005. Also provides employment in agriculture and nonagriculture in 2000 and 2005.

- dT.mat – the estimated productivity changes for the model to match observed real income changes for each province.

STATA and MATLAB Code

Replication files are provided for the main simulation and estimation results in the paper; that is, tables 4 through 11. Each table is a separate STATA or MATLAB file, depending. All are self-contained files that reproduce each table. For some tables, you must run the code twice with particular lines commented

out or not. For example, Table 8 reports our main estimates for the effect of lower trade costs with and without intermediate inputs. Run Table8.m once for the baseline estimates. Next, comment out the appropriate lines (as described in the file) to remove intermediate inputs from the model. Then run the code again.

In addition to files for each table, the following are also included:

- initial_eqm.m – sets up the initial equilibrium of the model. This file is called by all TableXX.m files.

- post_simulate.m – constructs a variety of results and summary statistics after simulating any given counterfactual.

- post_simulate_fixedlabor.m – as above, but with no-migration.

- main_simulate.m – the main system of equations to solve most counterfactuals.

- simulate_landreform.m – the system of equations to solve the land reform counterfactual.

- simulate_fixedlabor.m – the system of equations to solve the no-migration cases.

To solve the main growth accounting in Table 9, we provide two separate files: Table9a.m and Table9b.m. The former is for the top panel of the table and the latter is for the bottom panel. These can take considerable time to run, depending on your machine. We provided the output from each in decomposition_results.mat (for Table9a.m) and decomposition_results2.mat (for Table9b.m). These two .mat files are called by the bottom portion of code in Table9a.m and Table9b.m.