

An Anatomy of International Trade: Evidence from French Firms

Guide to Data and Programs

7 December 2010

This document describes the programs and datafiles used in the paper, including:

- SAS programs (.sas) with log files (.log),
- STATA programs (.do) with data files (.dta),
- GAUSS programs (.prg) with data in matrices (.fmt) or datasets (.dat),
- Excel spreadsheets (.xls) or (.csv), or
- ASCII files with data (.raw).

All of the SAS programming was done by Francis Kramarz in France. These programs extract the firm-level data and compute the statistics used in our paper. Francis then typically put the results into STATA files, but sent the original SAS files (and/or log files) for documentation. The data were written to ASCII files for uploading into GAUSS, which was used in estimation and simulation. All of the figures in the paper were created with STATA. Tables were done in Excel.

Firm-Level Data

The firm-level data come from Francis organized in three distinct ways:

1. By export destination of French firms, we have the 1986 data in `paysd.dta` (the 1992 data is `paysf.dta`). The associated SAS log file is `moment_pay_082007_86.log`. The data to be used in GAUSS are extracted and written to `paysd.raw` using `paysd.do`.

2. By strings of export destinations, we have the 1986 data in seqd.dta (the 1992 data are seqf.dta). The associated SAS log file is moment_pay_082007_86_92.log. The data to be used in GAUSS are extracted and written to seqd.raw using seqd.do.
3. By the number of export destinations per firm, we have the 1986 data in statsexplsd.dta (the 1992 data are in statsexplsf.dta). The associated SAS log file is moment_pay_082007_ssexplus_86.log. These data are used for figures but are not to construct the moments for estimation.

There are two other firm-level data files that are used in the estimation:

1. For counting up the total number of French exporters (and for constructing N_{nF} and \bar{X}_{nF} for the estimation) we include firms that appear not to sell in France (i.e. their exports exceed their total sales). These firms are in statneg.dta, with associated SAS log file moment_pay_072008_statneg.log. The data to be used in GAUSS are extracted and written to paysneg.raw using paysneg.do.
2. For calculating the variance-covariance matrix of the moments used in estimation and for calculating standard errors of our estimates, we create 2000 datasets by resampling from the original data. The SAS file to do the resampling is moment_boot_new102008.sas. The resamples of paysd.dta are stacked in payfin102008a.dta, payfin102008b.dta, payfin102008c.dta, and payfin102008d.dta. These data are combined and written to payfin2000p.raw (you need to recreate this .raw file, as it is quite large) using payfin102008.do. The resamples of seqd.dta are in seqfin092008.dta. It is written to seqfin2000p.raw (you need to recreate this .raw file, as it is quite large) using seqfin102008.do. The variable itir indicates the resample number, from 1, ..., 2000.

Other Data Files

Country names are contained in country.dta. Data on manufacturing absorption, used in some figures, is in aerdata.dta. Data on bilateral trade, trade deficits, and GDP, used for the counterfactuals, are in xnidata.raw, surpdata.raw, and gdpdata.raw.

Estimation

The parameter estimation is done by the method of simulated moments in `gmmsavc1.prg`. It takes data inputs: `paysd.raw`, `paysneg.raw`, and `seqd.raw`. Two programs must be run first:

1. The program `stem7mc8.prg` creates the underlying stochastic draws used in the estimation. These draws are stored in GAUSS datasets, and are held fixed for a single run of the estimation program.
2. The program `bootcov2.prg` constructs the weighting matrix for estimation, `ivcmat.fmt`, based on the variance covariance matrix of the moments. It takes the resampled moments, `payfin2000p.raw` and `seqfin2000p.raw` as input.

Robustness

There are a set of programs set up to redo the estimation using different sets of moments (as reported in Appendix D of the paper): `gmm-savc_big.prg`, `gmmsavc_25.prg`, and `gmmsavc3.prg`. The second require data on the 25th percentiles, which can be obtained by running `paysd_25.do` to construct `paysd_25.raw`. Also, to run the 2nd or 3rd of these programs one must reconstruct `ivcmat.fmt` using `bootcov_25.prg` or `bootcov3.prg`. The first of these requires `payfin2000p_25.raw` as input which in turn requires running `payfin102008_25.do` (you need to recreate this `.raw` file, as it is quite large).

Standard Errors

We estimate the standard errors of the parameter estimates by rerunning the estimation program using resampled data (and new simulation draws). This estimation on the resampled data is done in `gmmboot1.prg`. It is nearly identical to the main estimation program `gmmsavc1.prg` except that it takes the moments from `payfin2000p.raw` (instead of `paysd.raw`) and `seqfin2000p.raw` (instead of `seqd.raw`). Each time you run it, you must choose a new value of `itir` (which indicates a new resampling of the data) and you must rerun `stem7mc8.prg` (to generate different simulation error).

Monte Carlo Check

Starting with some parameter vector, suppose we simulate firms from the model and then proceed as if they were the actual data. We want to be sure that our estimation routine will deliver parameter estimates not far from

those we started with. Results of this exercise are reported in Appendix C of the paper). To perform this exercise we require a number of programs:

1. We simulate artificial firms and store the results in GAUSS datasets using `moc7f1r.prg`.
2. The program `moc7tf2r.prg` reads in the simulated firms and creates moments that are the analogs of those we use in the estimation.
3. The program `moc7gmm1.prg` performs the estimation. As in the actual estimation we must first run `stem7mc8.prg` to get simulation draws and `ivcmat` must have already been created (we don't reconstruct it from simulated firms).

Counterfactuals

There are two steps in performing counterfactuals. The general equilibrium calculations are done in `cfagg11.prg`. The implications at the firm-level are calculated in `cfmoc11.prg`.

Decompositions

The entry decompositions based on the estimated parameters are done in `entdec3c.prg`. The sales decompositions are done in `saldec3c.prg`.

Figures

Figures 1, 3, and 5 are drawn in `cnty10_new.do`, taking as input `country.dta`, `paysd.dta`, `statsexplused.dta`, and `aerdata.dta`.

Figure 2 is drawn with `distributions_new.do`. It uses the ranked sizes of individual firms according to their sales in each market. The data on sales in the French market is in `binsballsfr.dta` (with SAS log file `binsballsfr.log`) while in export markets it is `binsballs.dta` (with SAS log file `binsballs.log`). Due to confidentiality concerns, we cannot make these data available in their raw form. Contact the authors if you would like a version that is slightly aggregated.

Figure 5 is drawn in `moc7tf20_pay_new.do`. This last figure requires predicted values for firm-level moments, based on the estimated parameters. These predicted moments are calculated in `moc7tf20.prg`, with `moc7tf20_pay.csv` created from the GAUSS output file.

Although not reported, one can plot Figures 1, 3, and 5 separately for consumer and investment goods industries by running `cnty10_con.do` and `cnty10_inv.do`.

Tables

The tables are in a spreadsheets: `tables_I_and_II.xls`, `table_III.xls`, and `tables_IV_V_and_VI.xls`.

Regressions

The reported regression slopes are generally estimated in `cnty10_new.do`. The regression table in Appendix B is done in `appendixb.do` which uses: `niels.dta`, `gdp_capita.csv`, `paysf.dta`, `paysd.dta`, `aerdata.dta`, `exports_uruguay9.dta`, `gtrade92.dta`, and `concordances_uruguay_france.csv`.