Elasticity Optimism: Read me file on additional material

Jean Imbs * Isabelle Méjean†

June 2014

1 Estimating Trade Elasticities: Caliendo and Parro (2012)

1.1 Source data

- The datasets used in the estimations are called temp_bil.dta (sector and pooled microeconomic estimates) and temp_agbil.dta (aggregate estimates)
- The dataset providing the weights involved into the calculation of aggregate trade elasticities is called weight.dta.
- The raw data used in the estimation come from ComTrade (bilateral trade flows in thousands of US dollars) and UNCTAD-TRAINS (tariff data). They have been downloaded from WITS (www.wits.worldbank.org).
- Data coverage:
  - The trade data are for 1993.
  - The tariff data are also for 1993. When 1993 data are not available, however, we use the closest available year going backwards and forwards on a maximum time span between 1989 and 1996.

*Paris School of Economics, CNRS, and CEPR. Corresponding author: Paris School of Economics, 106 Boulevard de l'Hopital, 75013 Paris, France. jean.imbs@psemail.eu. www.jeanimbs.com.

†Ecole Polytechnique, CREST, and CEPR. Département d'Economie, Ecole Polytechnique, 91128 Palaiseau Cedex - France. isabelle.mejean@polytechnique.edu. www.isabellemejean.com.

- They are defined at the ISIC (Revision 3) level. The raw data are disaggregated at the 6-digit level of the HS nomenclature before being transformed into 21 product categories: Agriculture and Fishing (Isic 01-05), Mining and Quarrying (Isic 10-14), Food and tobacco (Isic 15-16), Textiles (Isic 17-19), Wood (Isic 20), Paper and publishing (Isic 21-22), Refined petroleum (Isic 23), Chemicals (Isic 24), Plastic products (Isic 25), Non-metallic mineral products (Isic 26), Iron and steel (Isic 271), Precious and non-ferrous metals (Isic 272), Metal products (Isic 28), Machinery and equipment (Isic 29), Office and computing machinery (Isic 30), Electrical machinery (Isic 31),Radio, TV, Communication (Isic 32), Medical and optical instruments (Isic 33), Motor vehicules (Isic 34), Other transport equipment (Isic 35), Furniture (Isic 36). The correspondence between sector codes and sector labels is provided in ISICRev3_cp_name.txt.
- The country sample contains 16 countries: Argentina, Australia, Canada, Chili, China, Colombia, the European Union, India, Indonesia, Japan, Korea, New Zealand, Norway, Switzerland, Thailand and the United States.
- The only difference in comparison with Caliendo and Parro (2012) is that we retain only those bilateral flows that involve the European Union, either as an exporter or as an importer.

- The tariff data are effectively applied rates aggregated across tariff lines within an ISIC category using simple average (although the weighted means are also available in the databases). The tariff variable is equal to one plus the tariff rate and is set to one whenever recorded tariff rates are negative.

- temp_bil.dta contains the following variable:
  - "isic3d" the product category (based on the ISIC Rev3 nomenclature)
  - "lntrade" the log of the ratio of reciprocal trade flows ($\ln \tilde{s}_l^k \equiv \ln \frac{s_{ij}^k s_{jh}^k s_{hi}^k}{s_{ji}^k s_{ih}^k s_{hj}^k}$ in the paper)
  - "lntariff" the log of the ratio of reciprocal tariff rates ($\ln \tilde{\tau}_l^k \equiv \ln \frac{\tau_{ij}^k \tau_{jh}^k \tau_{hi}^k}{\tau_{ji}^k \tau_{ih}^k \tau_{hj}^k}$ in the paper) where tariff rates are simple averages of effectively applied rates over tariff lines within a sector
  - "lnwtariff" the log of the ratio of reciprocal tariff rates ($\ln \tilde{\tau}_l^k \equiv \ln \frac{\tau_{ij}^k \tau_{jh}^k \tau_{hi}^k}{\tau_{ji}^k \tau_{ih}^k \tau_{hj}^k}$ in

the paper) where tariff rates are weighted averages of effectively applied rates over tariff lines within a sector

  - "nsec" a numerical sector identifier
  - "triplet" the identifier of the country triplet

- The aggregation of the data from the product to the aggregate level is done as follows:
  - Bilateral trade flows are summed over all product categories
  - Bilateral tariffs are averaged across product categories using simple averages (although the corresponding variable using a weighted mean of tariffs is also available in the database)

- temp_agbil.dta contains the following variable:
  - "lntrade" the log of the ratio of reciprocal aggregate trade flows ($\ln \tilde{s}_l \equiv \ln \frac{s_{ij}s_{jh}s_{hi}}{s_{ji}s_{ih}s_{hj}}$ in the paper)
  - "lntariff" the log of the ratio of reciprocal aggregate tariff rates ($\ln \tilde{\tau}_l \equiv \ln \frac{\tau_{ij}\tau_{jh}\tau_{hi}}{\tau_{ji}\tau_{ih}\tau_{hj}}$ in the paper) where tariff rates are simple averages of effectively applied rates over tariff lines
  - "lnwtariff" the log of the ratio of reciprocal aggregate tariff rates ($\ln \tilde{\tau}_l \equiv \ln \frac{\tau_{ij}\tau_{jh}\tau_{hi}}{\tau_{ji}\tau_{ih}\tau_{hj}}$ in the paper) where tariff rates are weighted averages of effectively applied rates over tariff lines
  - "triplet" the identifier of the country triplet

- The data in weight.dta combine information from the US input-output tables and from STAN. Namely,
  - The share of imports in expenditures (λ_j^k called "wkm_y_io" in weight.dta) is obtained from the US IO tables and is defined as the 1997 ratio of imports over domestic gross output.
  - The sectoral shares (m_j^k) are computed using information from STAN on the share of each sector in value added (n_j^k called "wk_vaid_stan" in weight.dta) (1997 data) and the share of imports in each sector's expenditures (λ_j^k). Using

these calculated import shares instead of the ones directly observed from trade data limits the influence of imported intermediate inputs, that are not taken into account in the definition of the aggregate trade elasticity (η_j).

1.2 Sector estimates

- Estimated in AEJMacro-2013-0231_SectorEstimates.do. The program can be run once the directories are defined (ie change the definition of the "datapath", "dopath" and "outputpath" global variables).

- Input to this program: temp_bil.dta, weight.dta and ISICRev3_cp_name.txt plus two Stata programs, namely bysector_reg_constant.do (sector-by-sector estimation) and unconstrained.do (aggregation of sectoral elasticities into an aggregate trade elasticity). The .dta and .txt files must be saved in $datapath and the .do files in $dopath.

- In bysector_reg_constant.do, the estimated equation corresponds to equation (4) in the paper.

- The aggregation exercise in unconstrained.do corresponds to equation (17) in the paper

- The output is saved in $outputpath into two datasets: regsec_wc_lntradelntariff.dta (sectoral estimates) and unc_wc_lntradelntariff.dta (Aggregate trade elasticity). The graph corresponding to Figure 1 in the paper is also saved as graphcp_lntradelntariff _EUref.eps

1.3 Pooled microeconomic estimates

- Estimated in AEJMacro-2013-0231_PooledMicroeconomicEstimates.do. The program can be run once the directories are defined (ie change the definition of the "datapath", "dopath" and "outputpath" global variables).

- Input to this program: temp_bil.dta, weight.dta and regsec_wc_lntradelntariff.dta (output of FinalAEJ_SectorEstimates.do) plus a Stata program, namely constrained_wc.do (pooled microeconomic estimation and calculation of the aggregate trade elasticity). The .dta files must be saved in $datapath and the .do file in $dopath.

- In constrained_wc.do, the estimated equation corresponds to equation (5) in the paper and the aggregation exercise follows equation (17).

- The output is saved in $outputpath into con_wc_lntradelntariff.dta (Aggregate trade elasticity + estimated coefficient)

1.4 Aggregate estimates

- Estimated in AEJMacro-2013-0231_AggregateEstimates.do. The program can be run once the directories are defined (ie change the definition of the "datapath", "dopath" and "outputpath" global variables).

- Input to this program: temp_agbil.dta, weight.dta and regsec_wc_lntradelntariff.dta (output of FinalAEJ_SectorEstimates.do) plus one Stata programs, namely aggregate_noc.do (Estimation on aggregated data and calculation of the aggregate trade elasticity). The .dta files must be saved in $datapath and the .do file in $dopath.

- In aggregate_noc.do, the estimated equation corresponds to equation (7) in the paper and the aggregation exercise follows equation (17).

- The output is saved in $outputpath into ag_noc_lntradelnmtariff.dta (Aggregate trade elasticity + estimated coefficient)

2 Estimating Trade Elasticities: Feenstra (1994)

2.1 Source data

- The datasets used to estimate elasticities using Feenstra's (1994) methodology are called temp_bil.sas7bdat (Sector and pooled microeconomic estimates) and temp_agbil.sas7bdat (aggregate estimates)

- They are obtained using as input the CEPII-BACI dataset (www.cepii.fr). See Gaulier and Zignago (2010) for a description of the database.[1]

[1]Gaulier, Guillaume, and Soledad Zignago. 2010. BACI: International Trade Database at the Product-Level. The 1994-2007 Version." CEPII Working Paper No. 2010-23.

- The raw data are defined at the 6-digit product level of the HS nomenclature. The sectoral estimation is performed sector-by-sector, a sector being defined as a 3-digit level of the ISIC (revision 3) nomenclature.

- An outlier treatment is performed on the raw data. The procedure:

  - Drops annual changes in bilateral unit values that are larger or lower than 5 times the median change in the same product category
  - Drops annual changes in bilateral market shares that are larger or lower than 5 times the median change in the same product category
  - Keeps only those ISIC sectors in which there are at least 20 exporters present in the US market over the whole 1996-2005 period.

- Variables in temp_bil.sas7bdat:

  - "t" year
  - "i" numerical code for the exporting country
  - "hs6" hs6-product code
  - "isic3" ISIC (Rev3) sector code
  - "dlp" First difference of the log of unit values (referred to $d \ln P_{it}^k$ in the paper). The estimated equation is based on the supply-demand system expressed in difference with respect to a reference country. "dlp_gap" ($= d \ln P_{it}^k - d \ln P_{rt}^k$ in the paper) corresponds to this difference. For each HS6-product category, the reference country is chosen among the set of exporters serving the US market over the whole 1996-2005 period. When there are several of those continuing exporters, the first one is arbitrarily chosen.
  - "dls" First difference of the log of market shares (referred to $d \ln s_{it}^k$ in the paper). The estimated equation is based on the supply-demand system expressed in difference with respect to a reference country. "dls_gap" ($= d \ln s_{it}^k - d \ln s_{rt}^k$ in the paper) corresponds to this difference. The reference country is the same as the one chosen to construct the double price differences.
  - "dlp2" corresponds to $Y_{it}^k = (d \ln P_{it}^k - d \ln P_{rt}^k)^2$ in the paper
  - "dls2" corresponds to $X_{1it}^k = (d \ln s_{it}^k - d \ln s_{rt}^k)^2$ in the paper

  - "dlps" corresponds to $X_{2it}^k = (d\ln s_{it}^k - d\ln s_{rt}^k)(d\ln P_{it}^k - d\ln P_{rt}^k)$ in the paper

- The aggregation of product-level variables at the country-pair aggregate level is performed as follows:

  - Aggregate price changes are defined as a weighted average of product-level bilateral unit value changes. The weights are either those of year $t-1$ (Laspeyres price index) or those observed in 1996 data (constant weight price index). The benchmark estimation in the paper uses the constant weight price index.

  - Aggregate market share changes are defined as a weighted average of product-level bilateral market share changes. The weights are either those of year $t-1$ (Laspeyres price index) or those observed in 1996 data (constant weight price index). The benchmark estimation in the paper uses the constant weight price index.

- Variables in temp_agbil.sas7bdat:

  - "t" year

  - "i" numerical code for the exporting country

  - "dlp" First difference of the log of aggregate unit values (referred to $d\ln P_{it} = \sum_{k=1}^K m_{it-1}^k d\ln P_{it}^k$ in the paper). The estimated equation is based on the supply-demand system expressed in difference with respect to a reference country. "dlp_gap" ($= d\ln P_{it} - d\ln P_{rt}$ in the paper) corresponds to this difference. For aggregate data, the reference country is Canada.

  - "dlp0" First difference of the log of aggregate unit values when constant weights are used in the aggregation ($d\ln P_{it}^{cst} = \sum_{k=1}^K m_{i1996}^k d\ln P_{it}^k$). "dlp0_gap" is the corresponding aggregate price variable in deviation from the reference country ($= d\ln P_{it}^{cst} - d\ln P_{rt}^{cst}$)

  - "dls" First difference of the log of aggregated market shares (referred to $d\ln s_{it} = \sum_{k=1}^K m_{it-1}^k d\ln s_{it}^k$ in the paper). The corresponding variable expressed in deviation with respect to the reference country is denoted "dls_gap" ($= d\ln s_{it} - d\ln s_{rt}$ in the paper).

  - "dls0" First difference of the log of aggregate market shares when constant weights are used in the aggregation ($d\ln s_{it}^{cst} = \sum_{k=1}^K m_{i1996}^k d\ln s_{it}^k$). "dls0_gap"

is the corresponding aggregate price variable in deviation from the reference country ($= d\ln s_{it}^{cst} - d\ln s_{rt}^{cst}$)

- "dlp2" corresponds to $Y_{it} = (d\ln P_{it} - d\ln P_{rt})^2$ in the paper. "dlp20" is the corresponding variable that uses a constant-weight aggregation scheme ($=Y_{it}^{cst} = (d\ln P_{it}^{cst} - d\ln P_{rt}^{cst})^2$)
- "dls2" corresponds to $X_{1it} = (d\ln s_{it} - d\ln s_{rt})^2$ in the paper. "dls20" is the corresponding variable that uses a constant-weight aggregation scheme ($=X_{1it}^{cst} = (d\ln s_{it}^{cst} - d\ln s_{rt}^{cst})^2$)
- "dlps" corresponds to $X_{2it} = (d\ln s_{it} - d\ln s_{rt})(d\ln P_{it} - d\ln P_{rt})$ in the paper. "dlps0" is the corresponding variable that uses a constant-weight aggregation scheme ($=X_{2it}^{cst} = (d\ln s_{it}^{cst} - d\ln s_{rt}^{cst})(d\ln P_{it}^{cst} - d\ln P_{rt}^{cst})$)

- The aggregation of sectoral estimates into an aggregate trade elasticity is performed using a vector of weights provided in weight_IO_MO_wkva.sas7bdat. This dataset contains the following variables:
  - "isic3" ISIC (Rev3) sector code
  - "nK" the share of each sector in the total value of US imports (m_j^k in the paper). Source: 1997 Input-Output tables
  - "wkm" the sector-specific share of imports in domestic gross output (λ_j^k). Source: 1997 Input-Output Tables.
  - "wK" the share of each sector in the aggregate interior demand (n_j^k). Source: STAN, 1997 data. As explained above, the benchmark results use these sectoral shares combined with the shares of imports in each sector's expenditures (λ_j^k) to calculate an alternative measure of import shares (m_j^k) that better reflects the contribution of each sector to the demand of foreign consumption goods.

2.2 Sector estimates

- Estimated in AEJMacro-2013-0231_Feenstra_SectorEstimates.sas. The program can be run in SAS once the directories are defined and the input datasets/programs are saved accordingly (input data in "&data" and programs in "&prog").

- Input to this program: temp_bil.sas7bdat and weight_IO_MO_wkva.sas7bdat plus three sas programs, namely panel_sitc_unc_nohetero_hs6FE.sas (IV estimation without the heteroscedasticity correction), panel_sitc_unc_hetero_FE.sas (IV estimation with the heteroscedasticity correction)and GridSearch.sas (Grid Search algorithm)

- The estimated equation corresponds to equation (10) in the paper and the aggregation exercise follows equation (17).

- The output is saved in "&res" into sigma.sas7bdat (product-level elasticities) and sol_u.sas7bdat (Unconstrained trade elasticity)

2.3 Pooled microeconomic estimates

- Estimated in AEJMacro-2013-0231_Feenstra_PooledMicroEstimates.sas. The program can be run in SAS once the directories are defined and the input datasets/programs are saved accordingly (input data in "&data" and programs in "&prog").

- Input to this program: temp_bil.sas7bdat and weight_IO_MO_wkva.sas7bdat plus two sas programs, namely panel_hs6_c_nohetero.sas (constrained IV estimation without the heteroscedasticity correction) and panel_hs6_c_hetero.sas (constrained IV estimation with the heteroscedasticity correction)

- The estimated equation corresponds to equation (13) in the paper and the aggregation exercise follows equation (17).

- The output is saved in "&res" into sol_c.sas7bdat (Constrained trade elasticity)

- The variant that uses instruments defined at the country and sector-level instead of the country-level is estimated in AEJMacro-2013-0231_Feenstra_PooledMicroEstimates _SectorIV.sas (see footnote 30 in the paper).

2.4 Aggregate estimates

- Estimated in AEJMacro-2013-0231_Feenstra_AggregateEstimates.sas. The program can be run in SAS once the directories are defined and the input datasets/programs are saved accordingly (input data in "&data" and programs in "&prog").

- Input to this program: temp_agbil.sas7bdat and weight_IO_MO_wkva.sas7bdat

- The estimated equation corresponds to equation (14) in the paper and the aggregation exercise follows equation (17).
- The output is saved in "&res" into sol_a.sas7bdat (Aggregate trade elasticity)
- The variant that uses an aggregation scheme using lagged weights instead of 1996 ones is estimated in AEJMacro-2013-0231_Feenstra_AggregateEstimates_LaspeyresW.sas (see footnote 30 in the paper).