No 2023-01 – January

Working Paper

CEPII

The CEPII Trade and Production Database

Thierry Mayer, Gianluca Santoni & Vincent Vicard

Highlights

- This paper documents the CEPII Trade and Production Database (TradeProd), which provides data on international and domestic trade flows and trade protection.
- The database covers 162 countries and 9 industrial sectors over the period 1966-2018.
- TradeProd is intended for econometric estimation of the gravity equation and also includes a yearly balanced dataset necessary for counterfactual exercises using new quantitative trade models.

Abstract

This paper documents the CEPII Trade and Production Database (TradeProd), which provides data on international and domestic trade flows and trade protection. The database covers 162 countries and 9 industrial sectors over the period 1966-2018. TradeProd is intended for econometric estimation of the gravity equation and also includes a yearly balanced dataset necessary for counterfactual exercises using new quantitative trade models.

Keywords

International Trade, Gravity Equation, Tariffs.

JEL

F14, F13.

Working Paper
CEPII

CEPII (Centre d'Etudes Prospectives et d'Informations Internationales) is a French institute dedicated to producing independent, policy-oriented economic research helpful to understand the international economic environment and challenges in the areas of trade policy, competitiveness, macroeconomics, international finance and growth.

CEPII Working Paper
Contributing to research in international economics

Editorial Director:
Antoine Bouët

Production: Laure Boivin

Published on 14.02.23

No ISSN: 1293-2574

CEPII
20, avenue de Ségur
TSA 10726
75334 Paris Cedex 07

contact@cepii.fr
Press contact: presse@cepii.fr
www.cepii.fr – @CEPII_Paris

RESEARCH AND EXPERTISE
ON THE WORLD ECONOMY

The CEPII Trade and Production Database

Thierry Mayer [*] Gianluca Santoni [†] Vincent Vicard [‡]

1 Introduction

The CEPII Trade and Production database (TradeProd) provides data on international and domestic trade flows and trade protection at the bilateral level for 162 countries and 9 industrial sectors over the period 1966-2018. It is intended for econometric estimation of the workhorse gravity model of trade, and also provides a yearly balanced (including domestic trade) dataset necessary for counterfactual exercises using new quantitative trade models.

TradeProd combines trade data from Comtrade, production data from UNIDO and tariffs data from WITS. We ensure an easy match with CEPII-Gravity dataset for other gravity variables (Conte, Cotterlaz & Mayer 2022). It is based on an earlier version presented in de Sousa, Mayer & Zignago (2012). The database will be updated yearly and made available on CEPII's website: `http://www.cepii.fr/`.

Data are aggregated into 9 industrial sectors in order to ensure a sectoral disaggregation consistent over a long time period for a maximum of countries. The TradeProd database covers the 1966-2018 period. Such long time span provides a large sample of historical episodes of trade policies and enables estimating their long term impact for which having a long panel has been proven to be important, e.g. the impact of the European Union (Mayer, Vicard & Zignago 2019, Head & Mayer 2021) or regional trade agreements more generally (Limão 2016).

Given its temporal depth and balanced version, TradeProd complements other existing trade and production data. WIOD provides information on both trade in goods and services for 43 countries and 56 sectors over 2000-2014 but includes estimated data and is not intended for econometrics analysis (Timmer, Dietzenbacher, Los, Stehrer & Vries 2015). The International Trade and Production Database provides information for a larger set of countries and industries – 243 countries and 170 industries – but over a short time span of 17 years (Borchert, Larch, Shikher & Yotov 2021).

This short paper presents the data source and construction of the database in Section 2 and descriptive statistics in Section 3 . In Section 4, we present benchmark estimations of the tariff elasticity, the border effect and the distance effect using TradeProd.

[*]Science-Po, CEPII, CEPR. Email: thierry.mayer@sciencespo.fr
[†]CEPII. Email: gianluca.santoni@cepii.fr
[‡]CEPII. Email: vincent.vicard@cepii.fr

2 Data source and construction

The TradeProd database is constructed combining international trade data from the UN Commodity Trade Statistics Database (COMTRADE) with production data from the UNIDO Industrial Statistics database (INDSTAT). The data cover a long period of time from 1966 to 2018 for 9 industrial sectors based on 2-digit ISIC (International Standard Industrial Classification, Rev. 3): Food (ISIC rev.3 15t16), Textiles (17t19), Wood-Paper (20t22), Chemicals (23t25), Minerals (26), Metals (27t28), Machines (29t33), Vehicles (34t35), Other (36). The number of industries is mandated by the fact that in the INDSTAT dataset a non-negligible number of countries declare a combination of two or more 2-digit ISIC categories.

International trade information is originally provided in Standard International Trade Classification (SITC) then converted into ISIC using available concordance tables and aggregated to match the 9-industry output grouping. Gross production minus total exports gives self-imports by country, industry and year. In the cases where self-imports are negative, we set them to missing. The final database provides both raw and extrapolated self-imports series (in two separate files). The extrapolation proceeds as follow: we start by computing the gross output to total export ratio for each country-industry-year cell then we fill the missing values using the nearest observation adjusted for the average growth rate of $GDP_{ot}/Export_{ot}$ and $Prod_{kt}/Export_{kt}$ to factor in country and industry level globalization trends.[1] Finally we recover gross output multiplying the extrapolated output to export ratio by the observed total exports for a given country-industry-year. International trade data, on the other hand, are never imputed or extrapolated.

Raw data on tariffs are sourced from the World Integrated Trade Solutions (WITS) of the World Bank. We start from the bulk database on both preferential and most-favored-nation (MFN) tariff rates at 6-digit HS level. Raw tariff data are treated as follow: i) we harmonize the different HS classifications using UN conversion tables to HS 1996; ii) we combine all the available years form 1988 to 2018 in a single dataset and squared it; iii) we carry forward the tariff rates to fill gaps in the data by origin-destination-year-product. Finally, we average the 6-digit HS rates at the ISIC 9-industry classification to be compatible with the rest of the data.[2]

In order to ensure the conformability of TradeProd with the CEPII-Gravity database (Conte et al. 2022), we opt for the same treatment for the territorial changes that occurred over the database time window. The final dataset ensures a perfect match with CEPII-Gravity, Annex A includes a short guide on how to combine TradeProd with CEPII-Gravity.[3]

We provide the dataset in two versions $TPe\_V202201$ and $TPc\_V202201$. Tariffs are included in both datasets. The name of each file contains the version identifier: $V202201$ refers to 2022 TradeProd release, version 1.

The TPe version is intended for estimation purposes, the dataset is not squared and the domestic production

[1] Practically, a missing observation in the output to export ratio is extrapolated as follow:

$$\widehat{Prod_{okt}/Export_{okt}} = Prod_{okt-1}/Export_{okt-1} \times [0.5 * (\Delta(GDP_{ot}/Export_{ot}) + \Delta(Prod_{kt}/Export_{kt}))]$$

with Δ being the growth rate. For all the series used in the extrapolation procedure we adopt a simple outlier selection rule based on the interquartile range (IQR). Namely we exclude observations higher than $75^{pc} + (1.5 * IQR)$ and lower than $25^{pc} - (1.5 * IQR)$. The outliers are replaced by linear interpolation.

[2] Notice that we also include the ad-valorem equivalent tariff rates as made available on the WITS platform. When tariffs and ad-valorem equivalent are available for a given country-pair and HS 6-digit product we retain the minimum value between the two.

[3] As an example, the country identifier for West Germany is coded as DEU1 and has 1990 as last year; starting from 1991 (first year the country code 276 appears in ComTrade) the unified Germany is coded as DEU2.

is not extrapolated.[4] The TPe database includes three different series for trade: i) $trade\_i$ based on importing country declarations (trade flow from origin o to destination d as reported by country d); ii) $trade\_e$ based on exporting country declarations (trade flow from o to d as reported by country o); and iii) $trade\_comb$ which combine import declaration flows with export declarations whenever the import one are missing (combined trade series ensure broader coverage). To ensure that the import and export series are consistent, we use the median transportation cost by sector to deflate the import series. The transport costs come from the OECD's Maritime Transport Cost Database.[5]

The TPc version is instead intended for counterfactual exercises using new quantitative trade models. The simulation oriented database includes two different series for trade based on $trade\_comb$: i) $trade\_sq$ which is squared by industry and year, i.e. it is non-missing for the same set of origin and destinations countries by industry and year; ii) $trade\_sq\_yr$ which is squared by year, i.e. it is non-missing for the same set of origin and destinations countries on all 9 industries in a year. The TPc version also includes a rest of the World aggregate, "ROW". Gross production for the "ROW" is extrapolated using the average gross output to total export ratio in a given industry and year. The same average ratio is used to fill the remaining missing values at the country level to ensure the squared by year trade series.[6]

3 Descriptive statistics

Table 1 reports the list of variables in the final dataset. Trade flows are identified by country of origin ($iso3\_o$) country of destination ($iso3\_d$), industry and year. Whenever self-imports are unavailable or negatives they are set to missing in the TPe version while they are extrapolated in the TPc version. In the TPc version the dummy variable $flag\_extra\_cty$ identifies the domestic sales observations based on extrapolated gross output using the adjusted country specific output to export ratios; whereas the dummy variable $flag\_extra\_avg$ identifies the domestic sales observations extrapolated using industry averages, as for the ROW aggregate. Finally, in 8.27 percent of the observations (year-country-industry) the production reported in INDSTAT results in negative domestic sales, which are then set as missing and extrapolated. These cases are indicated with the $flag\_extra\_neg$ variable.[7]

Figure 1 reports a summary of the coverage of the database comparing the total manufacturing output in TradeProd, computed as yearly $\sum_{odk} trade_{odk,t}$, with the aggregated figures reported in INDSTAT (ISIC D aggregate). Overall, TradeProd ensure a great coverage: over the period 2010-2018 TradeProd traces around 91.4 percent of world manufacturing production in the TPe version and 93.1 percent in the TPc version.

[4] We just linearly interpolate domestic production by country and industry to fill gaps in the data.

[5] OECD Maritime transport costs database contains data from 1991 to 2007 of bilateral maritime transport costs. Transport costs are available for 43 importing countries (including EU15 member states as a customs union) from 218 countries of origin and organized by 6-digit product according to the 1988 Harmonized System (we dropped 9.25 percent of the observations as reported at 2-digit). The raw data are available at `https://stats.oecd.org/Index.aspx?DataSetCode=MTC`, while a detailed description of the sources and methodology can be found at Korinek & Sourdin (2009). Since the original data use HS classification, while the sectors in Trade Prod are based on ISIC, we use the World Bank's World Integrated Trade Solutions (WITS) conversion tables to harmonize them. Table A1 in Appendix B provides the median and the mean transport costs for the TP industries.

[6] The list of the countries in the ROW aggregate is provided in a separate file available on the CEPII-TradeProd website. Only those countries for which there is no information in INDSTAT are included in the ROW aggregate.

[7] Notice that the "flag" dummy variables are expanded to identify the country with extrapolated domestic sales across all its bilateral observations, hence a simple *drop(keep) if* condition selects the relevant sample.

Table 2 reports the coverage of the squared data by decade and industries. Over time the coverage of the data, when filtered on self-import availability, increases steadily. Extrapolating self-imports improves substantially the overall coverage of the database. We assess the quality of the extrapolated series by industry in the following benchmark section.

The variable $tariff$ (included in both versions) combines MFN and preferential tariff rates. Starting from the 6-digit HS version $tariff$ retains the $min(tariff_{MFN}, tariff_{pref})$ when exporting country is a WTO member state (and $tariff_{pref}$ otherwise), the combined tariff rate is then averaged to the 9-industry ISIC classification used in TradeProd. MFN and $Pref$ tariff schedules are also provided.

Table 1: List of variables included in TradeProd

version	Variable	type	Description	Note
Common to TPe and TPc	year	int		1966-2018
	industry	str3	Based on 2-digit ISIC Rev. 3	9-industry aggregates
	$iso3\_o$	str4	origin country	ISO3 alphabetic code,
	$iso3\_d$	str4	destination country	territorial changes conform to CEPII-gravity
	$iso3num_o$	str4		ISO3 numeric code conform to CEPII-gravity
	$iso3num_d$	str4		
	$tariff_{MFN}$	double	MFN tariff rate	simple average WITS HS 6-digit, starts in 1988
	$tariff_{pref}$	double	Preferential tariff rate	simple average WITS HS 6-digit, starts in 1988,
	$tariff$	double	combines MNF & Pref rate	Min($tariff_MFN$, $tariff_pref$)
TPe	$trade\_i$	double	value of trade (Mln US $)	trade flow from o to d as reported by country d
	$trade\_e$	double	value of trade (Mln US $)	trade flow from o to d as reported by country o
	$trade\_comb$	double	value of trade (Mln US $)	combines $trade\_i$ with $trade\_e$
TPc	$trade\_sq$	double	value of trade (Mln US $)	squared by industry and year
	$trade\_sq\_yr$	double	value of trade (Mln US $)	squared by year
	$flag\_extra\_neg$	double	= 1 extrapolated negative domestic sales use country export to output ratios	
	$flag\_extra\_cty$	double	= 1 extrapolated domestic sales use country export to output ratios	
	$flag\_extra\_avg$	double	= 1 extrapolated domestic sales use average export to output ratios	

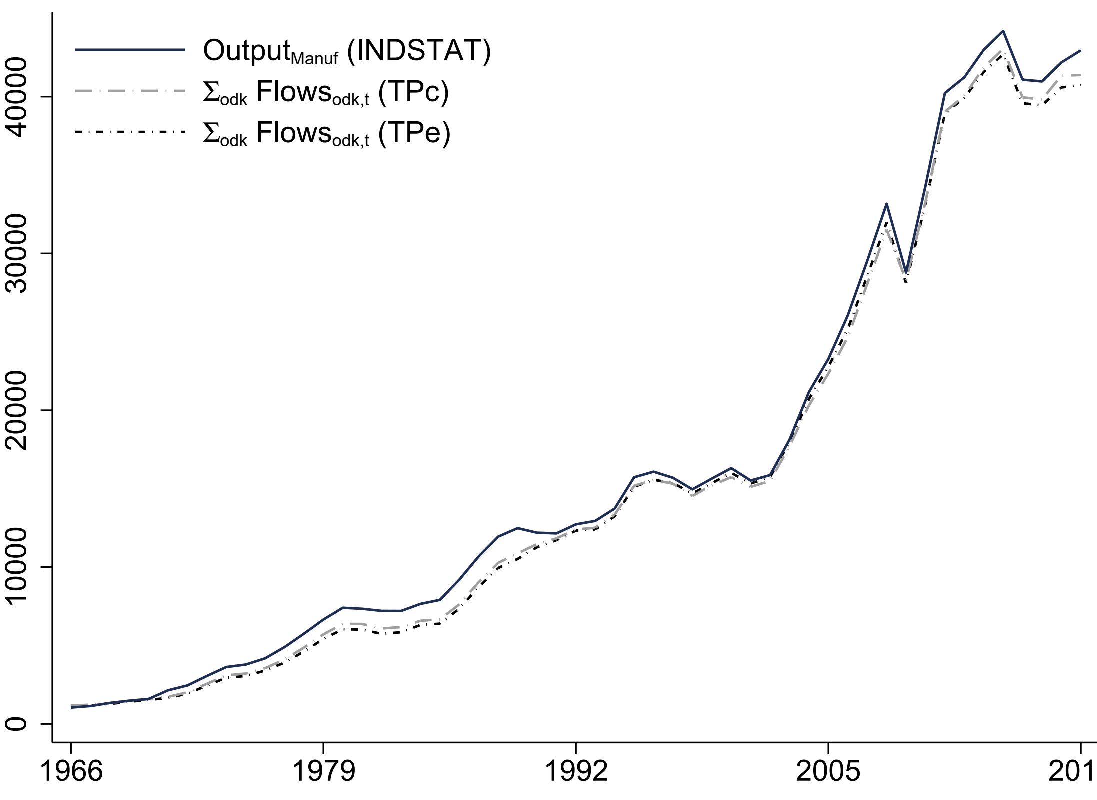
Note: $tariff_{MFN}$ and $tariff_{pref}$ are computed starting from HS 6-digit from the World Bank World Integrated Trade Solution (WITS) database. 6-digit values are aggregated to match the 9-industry grouping taking simple averages.

Table 2: Coverage by decade, countries with non missing domestic sales

Decade	Version	# Origin and Destinations, industry−by−year		
		Average	Max (industry)	Min (industry)
1966-1979		73	92 (15t16)	48 (34t35)
1980-1989		86	99 (15t16)	73 (34t35)
1990-1999	*TPe*	96	117 (15t16)	71 (29t33)
2000-2009	*trade_comb*	101	119 (15t16)	71 (29t33)
2010-2018		93	117 (15t16)	57 (17t18)
1966-1979		93	112 (23t25)	67 (26)
1980-1989		107	116 (27t28)	89 (34t35)
1990-1999	*TPc*	116	142 (15t16)	93 (34t35)
2000-2009	*trade_sq*	129	150 (15t16)	103 (29t33)
2010-2018		125	147 (15t16)	99 (29t33)
			Max (year)	Min (year)
1966-1979		104	118 (1979)	88 (1966)
1980-1989	*TPc*	119	121 (1989)	118 (1981)
1990-1999	*trade_sq_yr*	133	145 (1999)	120 (1991)
2000-2009		153	154 (2006)	152 (2005)
2010-2018		147	152 (2010)	134 (2018)

Note: the Max and Min columns also indicate the industry with the narrower and broader coverage by decades average, or the year with the narrower/broader coverage for all industries.

Figure 1: World Manufacturing Output



Note: The graph reports the total manufacturing output in TradeProd and INDSTAT. Total output in TradeProd is computed as the yearly $\sum Flows_{odk,t}$; whereas INDSTAT total manufacturing output refers to the ISIC D aggregate.

4 Benchmark Estimates

In this section, we report a simple benchmark exercise on : i) the $\log(1 + tariff)$; ii) the border effect; iii) the log(distance). We estimate a standard gravity equation using a PPML estimator as follows:

$$T_{odkt} = \exp(\beta_1 \log(1 + t_{odkt}) + \beta_2 B_{dk} + \beta_3 \log(d_{od}) + \beta_4 X_{od} + \omega_{okt} + \omega_{dkt}) + \epsilon_{odkt}. \quad (1)$$

We present both the results using the dependent variable T_{odkt} in level (T_{odkt} represent exports from country o to country d in sector k and year t) and in share of destination country absorption (T_{odkt} is exports divided by total imports of country d in sector k and year t), where the latter is meant to mitigate the difference in the penalization of large and small trade flows, see Eaton, Kortum & Sotelo (2013) and Head & Mayer (2014).

t_{odkt} is the import tariff rate, B_{dk} a border effect dummy equal to one when $o \neq d$, and d_{od} is weighted distance. X_{od} include the usual dyadic trade cost components: common language, contiguity and colonial ties. Finally, ω_{okt} and ω_{dkt} are fixed effects by country-industry-year that control for Multilateral Resistance Terms. Standard errors are two-way clustered at the origin country and destination country levels.

We report point estimates and their 95% confidence intervals by sector k (plus the pooled point estimate) in Figures 2, 3 and 4. Left panels report results based on trade flows in level and the right panels report results using trade as a share of destination country absorption. To assess the quality of extrapolated series, each figure reports estimations based on the TPe version of the dataset (in grey) and the TPc version (dark blue).

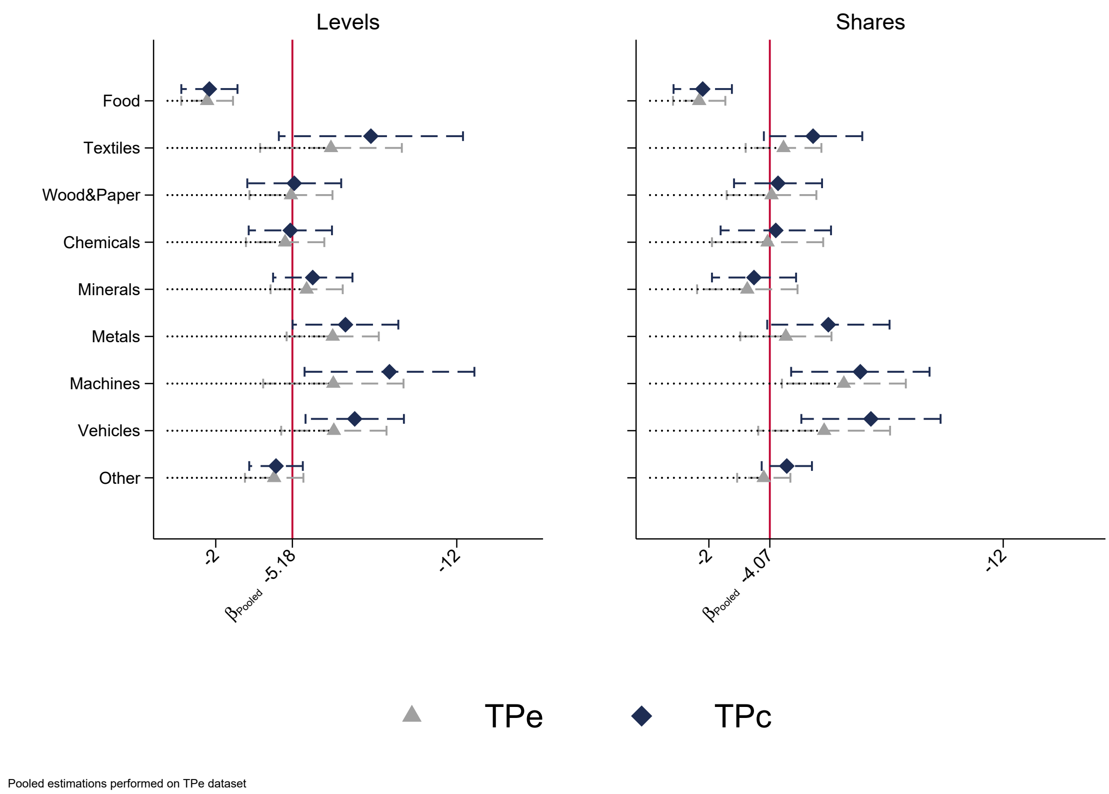
Figure 2 presents the estimated elasticity of substitutions, σ, based on $\log(1 + tariff)$. The estimated coefficient in levels, -5.18 is aligned with the most recent estimates (Fontagné, Guimbard & Orefice 2022). At the sectoral level the estimated elasticities are fairly centered around the average values, with the lower bound being Food products (ISIC rev.3 15t16) and the upper bound being Machines (29t33).

In Figure 3 we report the estimated border effects by industry, as well as with the one for manufactures. On average, crossing an international border represents a major impediment to trade flows. Taking our estimated $\sigma = 5.18$ (PPML-levels) and $\sigma = 4.07$ (PPML-shares) the tariff equivalent of the border effects is between 78.5 percent (levels) and 141.6 percent (shares), in line with previous estimates, see for example Head & Mayer (2014) and Yotov, Piermartini, Monteiro & Larch (2017).

Finally, in Figure 4 we report the estimated effect of distance on trade flows. Considering the pooled estimation the implied tariff equivalent range between 14.0 percent (levels) and 27.2 percent (shares).[8]

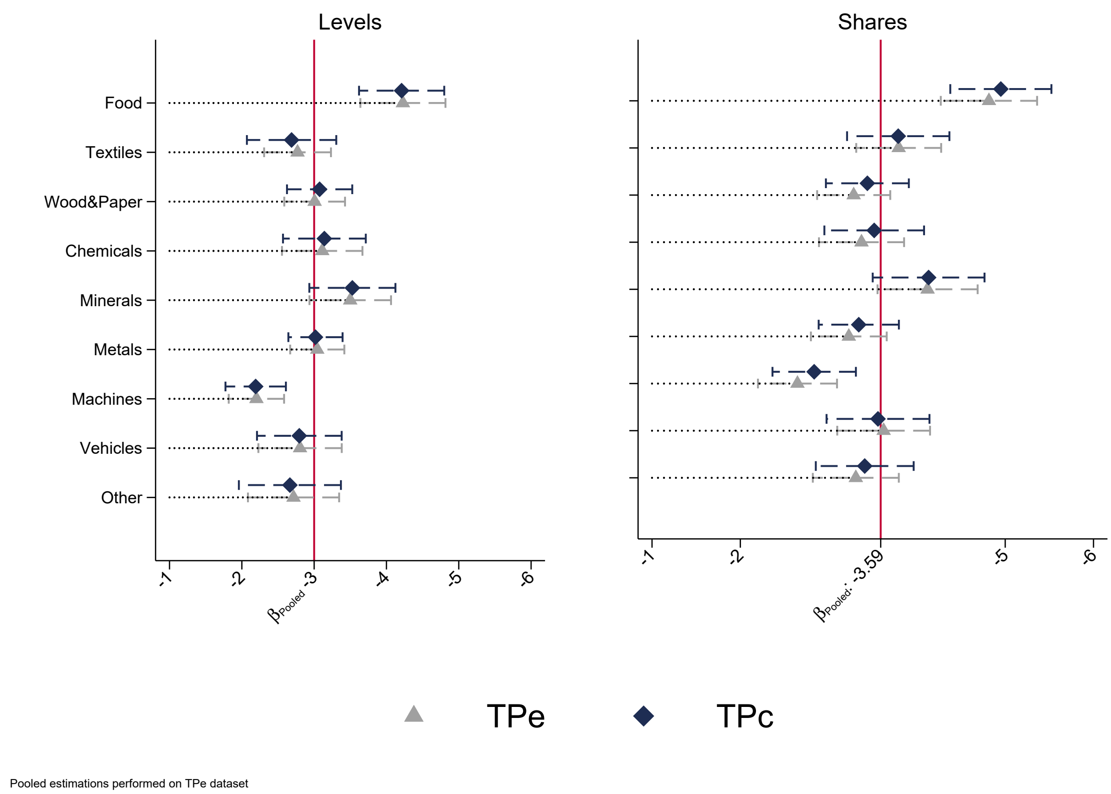
[8] The tariff equivalent effects are computed as: $\rho = [exp(\beta / - \sigma) - 1] * 100$

Figure 2: Tariff point estimates by industry



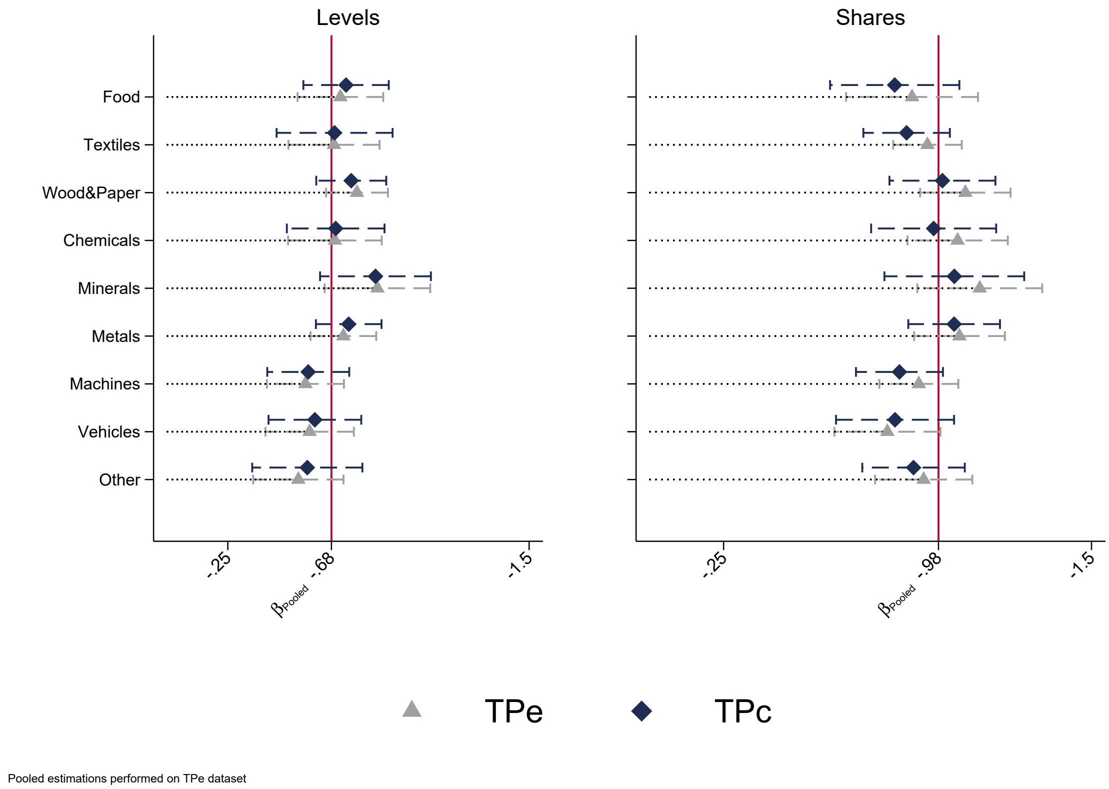
Note: The graph reports the estimated coefficients for $log(1 + tariff)$ from industry specific regressions on the period 1988-2018 controlling for both outward and inward multilateral resistance terms, i.e. with origin-by-year and destination-by-year fixed effects, as well as dyadic fixed effects (origin-by-destination). The left panel plots coefficients from a PPML regression in levels, while the right panel reports the estimated coefficients from PPML regression in shares of destination absorption. Whiskers display 95% confidence intervals ($\pm 1.96 * SE$), where standard errors, SE, are two-way clustered at the origin and destination level. β Pooled refers to the estimated effects in the industry pooled sample with origin-industry-year, destination-industry-year and origin-destination-industry fixed effects.

Figure 3: International Border point estimates by industry



Note: The graph reports the estimated international border coefficients from industry specific regressions on the period 1966-2018 controlling for both outward and inward multilateral resistance terms, i.e. with origin-by-year and destination-by-year fixed effects. The estimated equation also control for common language, contiguity and colonial ties. The left panel plots coefficients from a PPML regression in levels, while the right panel reports the estimated coefficients from PPML regression in shares of destination absorption. Whiskers display 95% confidence intervals ($\pm 1.96 * SE$), where standard errors, SE, are two-way clustered at the origin and destination level. β Pooled refers to the estimated effects in the industry pooled sample with origin-industry-year and destination-industry-year fixed effects.

Figure 4: Distance point estimates by industry



Note: The graph reports the estimated coefficients for $log(distance)$ from industry specific regressions on the period 1966-2018 controlling for both outward and inward multilateral resistance terms, i.e. with origin-by-year and destination-by-year fixed effects. The estimated equation also control for common language, contiguity and colonial ties. The left panel plots coefficients from a PPML regression in levels, while the right panel reports the estimated coefficients from PPML regression in shares of destination absorption. Whiskers display 95% confidence intervals ($\pm 1.96 * SE$), where standard errors, SE, are two-way clustered at the origin and destination level. β Pooled refers to the estimated effects in the industry pooled sample with origin-industry-year and destination-industry-year fixed effects.

Bibliography

Borchert, I., Larch, M., Shikher, S. & Yotov, Y. V. (2021), 'The International Trade and Production Database for Estimation (ITPD-E)', *International Economics* **166**(C), 140–166.

Conte, M., Cotterlaz, P. & Mayer, T. (2022), The CEPII Gravity Databa, Working Paper 2022-99999, CEPII research center.

de Sousa, J., Mayer, T. & Zignago, S. (2012), 'Market access in global and regional trade', *Regional Science and Urban Economics* **42**(6), 1037–1052.

Eaton, J., Kortum, S. & Sotelo, S. (2013), *International Trade: Linking Micro and Macro*, Vol. 2 of *Econometric Society Monographs*, Cambridge University Press, p. 329–370.

Fontagné, L., Guimbard, H. & Orefice, G. (2022), 'Tariff-based product-level trade elasticities', *Journal of International Economics* **137**, 103593.

Head, K. & Mayer, T. (2014), *Gravity Equations: Workhorse, Toolkit, and Cookbook*, Vol. 4 of *Handbook of International Economics*, Elsevier, chapter 3, pp. 131–195.

Head, K. & Mayer, T. (2021), 'The united states of Europe: A gravity model evaluation of the four freedoms', *Journal of Economic Perspectives* **35**(2), 23–48.

Korinek, J. & Sourdin, P. (2009), 'Clarifying trade costs: Maritime transport and its effect on agricultural trade', *OECD, Trade Directorate, OECD Trade Policy Working Papers* **32**.

Limão, N. (2016), Preferential trade agreements, *in* 'Handbook of Commercial Policy', Vol. 1, Elsevier, pp. 279–367.

Mayer, T., Vicard, V. & Zignago, S. (2019), 'The cost of non-Europe, revisited', *Economic Policy* **34**(98), 145–199.

Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R. & Vries, G. J. (2015), 'An Illustrated User Guide to the World Input–Output Database: the Case of Global Automotive Production', *Review of International Economics* **23**(3), 575–605.

Yotov, Y. V., Piermartini, R., Monteiro, J.-A. & Larch, M. (2017), *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*, United Nations publication, World Trade Organization.

Appendix

A How to merge with CEPII-Gravity Database

The following Stata code shows how to combine TradeProd ($TPe_V202201$) with the CEPII-Gravity Database ($Gravity\_V202102$). The same procedures applies to the $TPc_V202201$, with the GE version users need also to match the ROW aggregate as reported below.

```
. use "$CEPII\Gravity_V202102.dta", clear

. keep if year >= 1966         /*** first year in TPe_V202201: 1966 ***/
(1,107,072 observations deleted)

. keep if year <= 2018         /*** last  year in TPe_V202201: 2018 ***/
(61,504 observations deleted)

. keep iso3num_o iso3num_d year dist contig comlang_off comcol rta eu_d eu_o wto_o wto_d gatt_o gatt_d

. merge 1:m iso3num_o iso3num_d year using "$trade_prod\TPe_V202201.dta"

    Result                      Number of obs
    -----------------------------------------

    Not matched                     2,152,137
        from master                 2,152,137  (_merge==1)
        from using                          0  (_merge==2)

    Matched                         9,968,175  (_merge==3)
    -----------------------------------------

. keep if _m == 3
(2,152,137 observations deleted)

. drop _m

.
```

```
. use "$CEPII\Gravity_V202102.dta", clear

.
. cap drop if iso3num_o == .

. cap drop if iso3num_d == .

. keep if year >= 1966         /*** first year in TPc_V202201: 1966 ***/
(1,098,162 observations deleted)

. keep if year <= 2018         /*** last  year in TPc_V202201: 2018 ***/
(61,009 observations deleted)

. keep iso3num_o iso3num_d year dist contig comlang_off comcol rta eu_d eu_o wto_o wto_d gatt_o gatt_d

.
.
.
. *******************************************************************************
. * ROW
. merge m:1 iso3num_o using "$trade_prod\cty_list_ROW", keepusing(iso3num_o)

    Result                      Number of obs
    -----------------------------------------
    Not matched                     2,210,897
        from master                 2,210,897  (_merge==1)
        from using                          0  (_merge==2)

    Matched                         1,022,580  (_merge==3)
    -----------------------------------------

. replace iso3num_o =999 if _m ==3
(1,022,580 real changes made)

.
. drop _m

.
.
. merge m:1 iso3num_d using "$trade_prod\cty_list_ROW", keepusing(iso3num_d)
(variable iso3num_d was int, now float to accommodate using data´s values)
(label _merge already defined)

    Result                      Number of obs
    -----------------------------------------
    Not matched                     2,210,897
        from master                 2,210,897  (_merge==1)
        from using                          0  (_merge==2)

    Matched                         1,022,580  (_merge==3)
    -----------------------------------------

. replace iso3num_d =999 if _m == 3
(1,022,580 real changes made)

. drop _m

.
```

```
. ********************************************************************************
. * simple collapse (mean) gravity controls
.
. collapse (mean) dist,  by(iso3num_o iso3num_d year)

.
.
. merge 1:m iso3num_o iso3num_d year using "$trade_prod\TPc_V202201.dta"

    Result                      Number of obs
    -----------------------------------------

    Not matched                       604,267
        from master                   604,267  (_merge==1)
        from using                          0  (_merge==2)

    Matched                         8,328,591  (_merge==3)
    -----------------------------------------

. keep if _m ==3
(604,267 observations deleted)

. drop _m

.
```

B Transport cost statistics

Table A1 shows the mean and median values of transportation costs calculated from the OECD Maritime Transport Costs database. The median values are those used as the deflator for the value of imports when combining them with export declarations.

Table A1: List of variables included in TradeProd

TP Industries	ISIC	Ad Valorem Transport Costs	
		Median	Mean
Food	15t16	0.079	0.102
Textiles	17t19	0.044	0.065
Wood-Paper	20t22	0.083	0.108
Chemicals	23t25	0.060	0.076
Minerals	26	0.096	0.124
Metals	27t28	0.055	0.070
Machines	29t33	0.034	0.055
Vehicles	34t35	0.050	0.078
Other	36	0.064	0.092

Note: raw data from OECD Maritime Transport cost database.