Leibniz Institute for
EAST AND SOUTHEAST
EUROPEAN STUDIES

Arbeitsbereich Ökonomie

IOS Working Papers

No. 399 June 2023

Time-varying trade cost (terms) and the distance puzzle

Richard Frensch* and Michael Rindler **

* Corresponding author, Email: frensch@ios-regensburg.de; IOS Regensburg, University of Regensburg, ZU Friedrichshafen. ** ZU Friedrichshafen, ifo Institute München.

Leibniz Institute for
EAST AND SOUTHEAST
EUROPEAN STUDIES
Landshuter Straße 4
D-93047 Regensburg

Telefon: (09 41) 943 54-10
Telefax: (09 41) 943 54-27
E-Mail: info@ios-regensburg.de
Internet: www.leibniz-ios.de
ISSN: 2199-9465

Contents

List of Tables

List of Figures

Abstract

We present a novel two-stage gravity specification with period-varying bilateral trade cost terms. We test our specification to confirm the benchmark result on declining international distance elasticities over time, using two new data sets. Analyzing period-varying bilateral trade cost derived from our specification offers additional insights: first, globalization has erased more than a third of the effect of distance on trade cost, mostly until the mid-nineties. Second, identifying period-varying bilateral trade cost separately for domestic vs. international trade offers a natural illustration to globalization – international trade cost are less persistent than domestic trade cost. Finally, reflecting the importance of general equilibrium adjustment, total bilateral trade cost – relating partial bilateral trade cost to multilateral resistances – are more appropriate to reflect globalization than partial bilateral trade cost.

JEL-Classification: C23, F15, F40, O18

Keywords: Gravity, geography, panel models

1 Introduction

The gravity literature has moved towards identifying trade cost within structural approaches, explicitly derived from general equilibrium consistent models (for a survey, see Yotov, 2022). Neglecting time, theoretical demand side derivation of structural gravity, based on identical individual CES preferences, results in the following expression to govern nominal trade flows X from country o to d, when normalized by world income Y (as in Allen et al., 2020),

$$\mathrm{x_{od}} = \mathrm{X_{od}}/\mathrm{Y} = y_o e_d \left(\frac{t_{od}}{\Pi_o P_d}\right)^{1-\sigma}, \tag{1}$$

where $\sigma > 1$ is the elasticity of substitution between any pair of goods. y_o and e_d are production and expenditure shares in the world for origin and destination countries, respectively; $t_{od} \geq 1$ is partial bilateral trade cost between trading partners o and d, routinely defined as iceberg costs. Outward multilateral resistance, $\Pi_o^{1-\sigma}$, is a weighted-average aggregate of all partial bilateral trade costs facing the producers of country o,

$$\Pi_o{}^{1-\sigma} = \sum_d \left(\frac{t_{od}}{P_d}\right)^{1-\sigma} \frac{E_d}{Y}. \tag{2}$$

$P_d^{1-\sigma}$ is inward multilateral resistance of destination country d, a weighted-average aggregate of all bilateral trade costs facing the consumers in country d,

$$P_d{}^{1-\sigma} = \sum_o \left(\frac{t_{od}}{\Pi_o}\right)^{1-\sigma} \frac{Y_o}{Y}. \tag{3}$$

Multilateral resistance is a general equilibrium concept, and the intuitive consequence of rising multilateral resistance is that the higher the trade barriers of a country with the world for fixed trade barriers with a specific country, the more the country will be driven to trade with this specific country rather than with the rest of the world (Anderson and van Wincoop, 2003). Thus, decomposing the right-hand side of equation (1) into a size term, $y_o e_d$, and a trade cost term, $(t_{od}/(\Pi_o P_d))^{1-\sigma}$, is intuitively instructive: the size term describes frictionless trade. The trade cost term summarizes the deviation of actual from frictionless trade due to partial bilateral trade costs relative to multilateral trade resistances: $t_{od}/(\Pi_o P_d)$ represents total bilateral trade cost between countries o and d.

Empirical structural gravity estimation specifications typically include time-varying directional country fixed effects and time-invariant country-pair fixed effects to be estimated on panels of worldwide trade data, including countries' domestic trade. We deal with two aspects of trade cost in these specifications: (1) Time (in)variance; (2) partial *versus* general equilibrium scope, i.e., we differentiate between partial *versus* total bilateral trade cost. We do so against the background of the well-known distance puzzle.

2 Gravity specifications for solving the distance puzzle

While globalization has induced falling distance-related costs (Coe et al., 2002), traditional gravity failed to find declining distance elasticities over time (Disdier and Head, 2008). Solutions to this distance puzzle were worked out in two steps. Yotov (2012) includes both international and domestic trade into a sequence of otherwise traditional cross-section gravity estimations to document declining differences over time between international and domestic distance elasticity estimates. Bergstrand et al. (2015) use a structural gravity specification with time-varying direction-specific country fixed effects on both domestic and international panel trade data, removing remaining bias in (change of) distance elasticity estimates by accounting for both time-invariant unobserved heterogeneity across country pairs and time-variant trade policies.

Our benchmark specification, oriented at Bergstrand et al. (2015), includes time-varying directional country fixed effects, $\eta_{o,t}$ and $\theta_{d,t}$, and time-invariant country-pair fixed effects, γ_{od},

$$x_{od,t} = \exp\left(\sum_{p} \beta_{1p} \ln IntDist_{od,p} + \sum_{l=0}^{4} \beta_l \, RTA_{od,t-l} + \gamma_{od} + \eta_{o,t} + \theta_{d,t}\right) \times \varepsilon_{od,t} \tag{4}$$

where $x_{od,t}$ is nominal exports normalized by world income from country of origin, o, to country of destination, d, in year t. Time-varying regional trade agreements (RTA) include four lags. We estimate international distance ($IntDist$) elasticities as varying by period p, each period covering three non-overlapping consecutive years. Consequently, β_{1p} estimates are to be interpreted as absolute changes relative to a base period. Unlike Bergstrand et al. (2015), our base period is our final period such that we expect β_{1p} to be negative and decreasing in absolute value.

Our second specification rests on a two-stage procedure. The first stage consists of a general equilibrium constrained unsaturated but very high goodness of fit decomposition, including RTA-effects, of observed international and domestic annual trade into time-varying directional country-specific and country-pair fixed effects, letting country-pair effects vary by three-year period p,[1]

[1] For doing so on an annual basis, there are not enough degrees of freedom: N-country and T-year T×N2 panels of trade observations, including countries' domestic trade, cannot be decomposed into 2×T×N time-varying directional country effects plus T×N2 time-variant country-pair effects. For that lack of sufficient degrees of freedom, fully saturated decompositions as in Egger and Nigai (2015) require normalizing domestic trade costs. Moreover, full decompositions suffer from concerns about singletons due to frequent zeros in the trade matrix (Correia, 2015; Correia et al., 2019).

$$x_{od,t} = \exp\left(\sum_{l} \beta_l \, RTA_{od,t-l} + \sum_{p} \gamma_{od,p} + \eta_{o,t} + \theta_{d,t}\right) \times \varepsilon_{od,t} \tag{5a}$$

From this first-stage decomposition, we recover fitted values of all bilateral, i.e. all international and domestic exponentiated period-varying country-pair fixed effects, $\exp(\hat{\gamma}_{od,p})$, and period averages of the exponentiated time-varying directional country fixed effects, $\exp(\hat{\eta}_{o,t})$, $\exp(\hat{\theta}_{d,t})$. We test our novel gravity specification against the Eq. (4) benchmark on the basis of predictions from the recovered fixed effects, $\exp(\hat{z}_{od,p}) = \exp(\hat{\gamma}_{od,p}) \cdot \sum_{t \in p} \exp(\hat{\eta}_{o,t} + \hat{\theta}_{d,t}) \cdot \hat{\varepsilon}_{od,t}$.[2] We then apply a period-specific variant of Eq. (4) on the $\exp(\hat{z}_{od,p})$ in a second stage,

$$\exp(\hat{z}_{od,p}) = \exp\left(\sum_{p} \beta_{2p} \ln IntDist_{od,p} + \delta_{od} + \varphi_{o,p} + \xi_{d,p}\right) \times \varepsilon_{od,p} \tag{5b}$$

As in Eq. (4), our β_{2p} estimates are to be interpreted as absolute changes, relative to the final period.

Estimations are executed in stata, relying on recent advances in high-dimension fixed effects estimation techniques, using `ppmlhdfe` (Correia et al., 2020), followed by `ppml_fe_bias` (Weidner and Zylkin, 2021), to implement an analytical asymptotic incidental parameter bias correction for PPML regressions with two-way (as in Eq. (5a), when only standard errors are biased) or three-way fixed effects (as in Eq. (4)).

[2] We attribute first-stage residual terms to the fixed effects, as recommended for the case of time-invariant country-pair effects in Spornberger (2022). Honoré and Kesina (2017) demonstrate that this recommendation hinges on making random effects-type assumptions on the time-invariant explanatory variables behind the pair effects.

3 Data

We use two manufacturing trade data panels, each including domestic trade. International trade data are from the 2021 version of CEPII's BACI and from CEPII's TradeProd data set. Domestic manufacturing trade is constructed using gross production data in manufacturing sectors from UNIDO's Indstat 2 database, complemented by CEPII's TradeProd data (for details, see the appendix). Our long panel A covers 55 major trading countries between 1983 and 2018. We apply the dynamic country code identifier of the USITC Dynamic Gravity Dataset. In this sense, panel A is fully balanced for 49 countries. Our shorter but broad panel B covers 94 countries between 1995 and 2018, coming close to covering worldwide international trade in manufacturing goods. The balanced version of panel B still covers 71 countries, capturing about 75 percent of worldwide international trade in manufacturing goods. RTA information is from USITC DGD (Gurevich and Herman, 2018), as is country pair information on (population-weighted) distance.

4 Distance effects

In Figure 1, we present international distance effects estimated according to Eqs. (4) and (5b) on the balanced versions of our samples A and B. For both panels, we find the same pattern: estimated distance elasticities monotonically decline over time. Across panels A and B, no correlation coefficient between respective Eqs. (4) or (5b) distance elasticity point estimates is below .998. With our higher frequency of estimates (for three-year periods), the length of our panel A and the breadth of our panel B, we take this result as confirming and extending Bergstrand et al. (2015). Within panels A and B, respectively, no correlation coefficient between Eqs. (4) and (5b) distance elasticity point estimates is below .999. We take this to reflect the success of our two-stage procedure to reproduce the Eq. (4) benchmark precisely. First, this is in spite of the lower frequency of our second stage three-year period data. Second, calculating an imputed R^2 as the squared correlation between outcome and fitted values (Egger and Staub, 2016) reveals a very high fit for the Eq. (5a) decomposition (see the appendix), which is in fact "almost saturated," i.e., it accounts for almost all of the variation of observed trade. Therefore, attributing or not first-stage residual terms when recovering fixed effects from the first for the second stage estimations does not have any impact on our results. Accordingly, we take Eqs. (5a) and (5b) to represent our concept of time-varying trade cost terms.

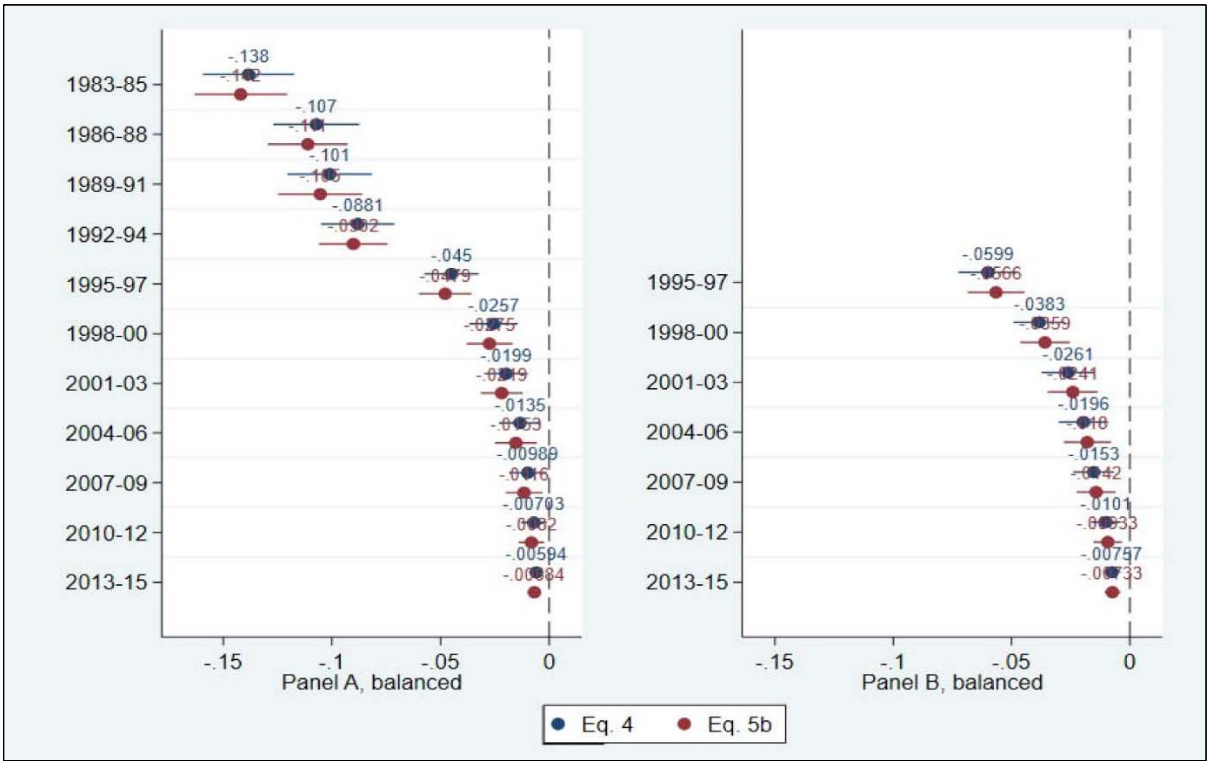

Figure 1: Time-varying international distance effects on trade over time

	Panel A, balanced		Panel B, balanced	
	Eq. (4)	Eq. (5b)	Eq. (4)	Eq. (5b)
Observations	86,436	28,165	115,328	37,334
Clusters	2,401	2,401	4,875	4,842
Pseudo-R^2	0.506	0.748	0.508	0.760
Imputed R^2	0.999	0.999	0.999	0.999

Notes: For all other coefficient estimates, see the appendix. For computing parameter estimates and standard errors (clustered on country pairs), we use `ppmlhdfe` (Correia et al., 2020), followed by a local de-biasing adjustment according to Weidner and Zylkin (2021). Imputed R^2 is the squared correlation between outcome and fitted values (Egger and Staub, 2016).

5 Trade cost

The strict monotonicity of the decline in international distance effects across two estimation approaches and two data sets is noteworthy: it is the result of the combined effects of declining distance-related costs on trade costs and their subsequent effects on trade, spread out over time. To isolate the first from the second, in order to concentrate on the trade cost side of globalization, with Eq. (1) we transform our first stage Eq. (5a) period-varying fitted values of all domestic and international trade cost term estimates into period-varying *partial bilateral trade cost* estimates, according to $PBTC_{od,p} = \hat{t}_{od,p} = \left(\exp(\hat{\gamma}_{od,p})\right)^{\frac{1}{1-\sigma}} \cdot \hat{\varepsilon}_{od,t}$, choosing σ = 7 for the elasticity of substitution between any pair of goods, as is standard in the literature (Head and Mayer, 2014).

In addition to accounting for multilateral resistances, the first stage Eq. (5a) time-varying directional country-specific fixed effects also absorb country-specific y_o and e_d, the respective production and expenditure shares in the world for origin and destination countries. We extract period-specific estimates for outward and inward multilateral resistances according to $\widehat{\Pi}_{o,p} = \left(\sum_{t \,\epsilon\, p} \frac{y_{o,t}}{\exp(\hat{\eta}_{o,t})}\right)^{\frac{1}{1-\sigma}}$ and $\hat{P}_{d,p} = \left(\sum_{t \,\epsilon\, p} \frac{e_{d,t}}{\exp(\hat{\theta}_{d,t})}\right)^{\frac{1}{1-\sigma}}$. This finally gives us period-varying *total bilateral trade cost* estimates, including total domestic trade cost estimates, $TBTC_{od,p} = \hat{t}_{od,p}/(\widehat{\Pi}_{o,p} \cdot \hat{P}_{d,p})$.

We then analyze our panels of both partial and total bilateral trade cost estimates similarly to Jacks et al. (2008), i.e., using a specification with period-fixed and country-pair fixed effects, but preferring PPML to their log-linear OLS estimation on only international trade cost,

$$BTC_{od,p} = \exp\left(\sum_{p} \beta_{cost,p} \ln IntDist_{od,p} + \delta_{od} + \varphi_p\right) \times \varepsilon_{od,p} \tag{6}$$

where BTC is either partial or total bilateral trade cost, and $\beta_{cost,p}$ estimates are again to be interpreted as absolute deviations, relative to the final 2016–18 period.

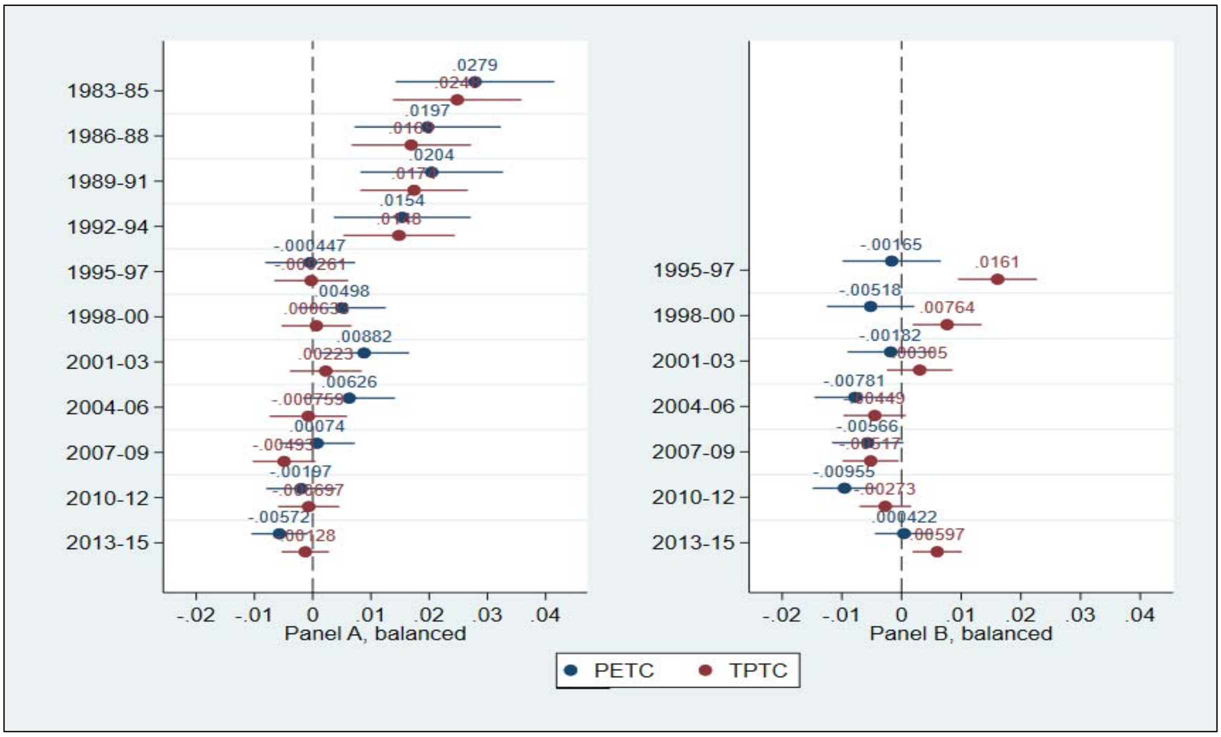

Figure 2: Time-varying international distance effects on partial *vs*. total bilateral trade cost: Eq. (6)

	Panel A, balanced		Panel B, balanced	
	PBTC	TBTC	PBTC	TBTC
Observations	28,165	28,165	37,334	37,334
Clusters	2,401	2,401	4,842	4,842
Pseudo-R^2	0.129	0.039	0.118	0.042
Imputed R^2	0.752	0.706	0.792	0.773

Notes: See Figure 1.

Globalization has erased a large part of the effect of distance on trade costs: Jacks et al. (2008) find a level distance effect of 0.08 on partial bilateral trade costs. According to our Figure 2, this coefficient would have declined by .0279, i.e., by more than a third since the mid-eighties, mostly already by the mid-nineties. Different from the distance effect on trade, the change in the impact of distance on both partial and total bilateral trade costs has not been monotonic, and has come to a standstill since the early 2000s at the latest. In as much as globalization goes beyond declining distance effects, this result suggests that other globalization-related adjustment processes of reducing trade costs may have gained importance since the mid-1990s. Also, according to Figure 2, there is no visible difference in the patterns of distance effects upon partial *versus* total bilateral trade costs.

6 Trade cost persistence

Anderson and Yotov (2020) introduce *short-run gravity* to reflect their model of globalization: a smooth, speed-of-adjustment-parameter governed process of reducing short-run capacity constraints to bilateral trade, through costly reallocation of relation-specific capital. This can be incorporated in a single-stage dynamic approach, where lagged trade – meant to proxy the presence of capacity constraints – is added to the benchmark specification in Eq. (4). Analogously, we now extend our Eq. (6) specification towards a dynamic approach. However, we include both international and domestic versions of our lagged dependent variables, partial *versus* total bilateral trade costs, to allow for smooth adjustment of time-varying but persistent trade costs over time, whatever the reason behind the persistence may be. Such a dynamic specification may be problematic when unobserved effects are part of the composed error term and thus – by construction – correlated with the lagged dependent variable. As in Anderson and Yotov (2020), our straightforward solution to this endogeneity issue is to keep Eq. (6) country-pair fixed effects,

$$\begin{aligned} BTC_{od,p} &= \exp\left(\sum_{p} \beta_{cost,p} \ln IntDist_{od,p} + \delta_{od} + \varphi_p\right) \times \\ &\times \exp\left(\sum_{p} \beta_{dyn,dom} \ln\left(BTC_{od,p-1}, dom\right) + \beta_{dyn,int} \ln\left(BTC_{od,p-1}, int\right)\right) \\ &\times \varepsilon_{od,p} \end{aligned} \tag{7}$$

Figure 3 shows that both partial and total bilateral trade cost developments over time can well be represented by speed-of-adjustment-parameter governed processes as in Anderson and Yotov (2020): in the presence of trade cost persistence, the change of the distance effect on trade costs does not any more exhibit a perceptible trend. In line with Anderson and Yotov (2020), Figure 3 distance effects are smaller than in Figure 2, and are generally higher for the lower trade cost persistence sample B.

Most importantly, identifying period-varying bilateral trade costs separately for domestic *vs*. international trade offers a natural illustration to globalization – international trade costs are less persistent than domestic trade costs. This difference, however, is both economically and statistically significant only for total rather than for partial bilateral trade costs, reflecting the importance of general equilibrium adjustment during globalization. In this sense, total bilateral trade costs is a more appropriate concept to reflect globalization than partial bilateral trade costs. As trade cost persistence – both domestically and internationally – is lower in our broader but shorter sample,[3] we may also conclude that trade cost persistence has declined during globalization.

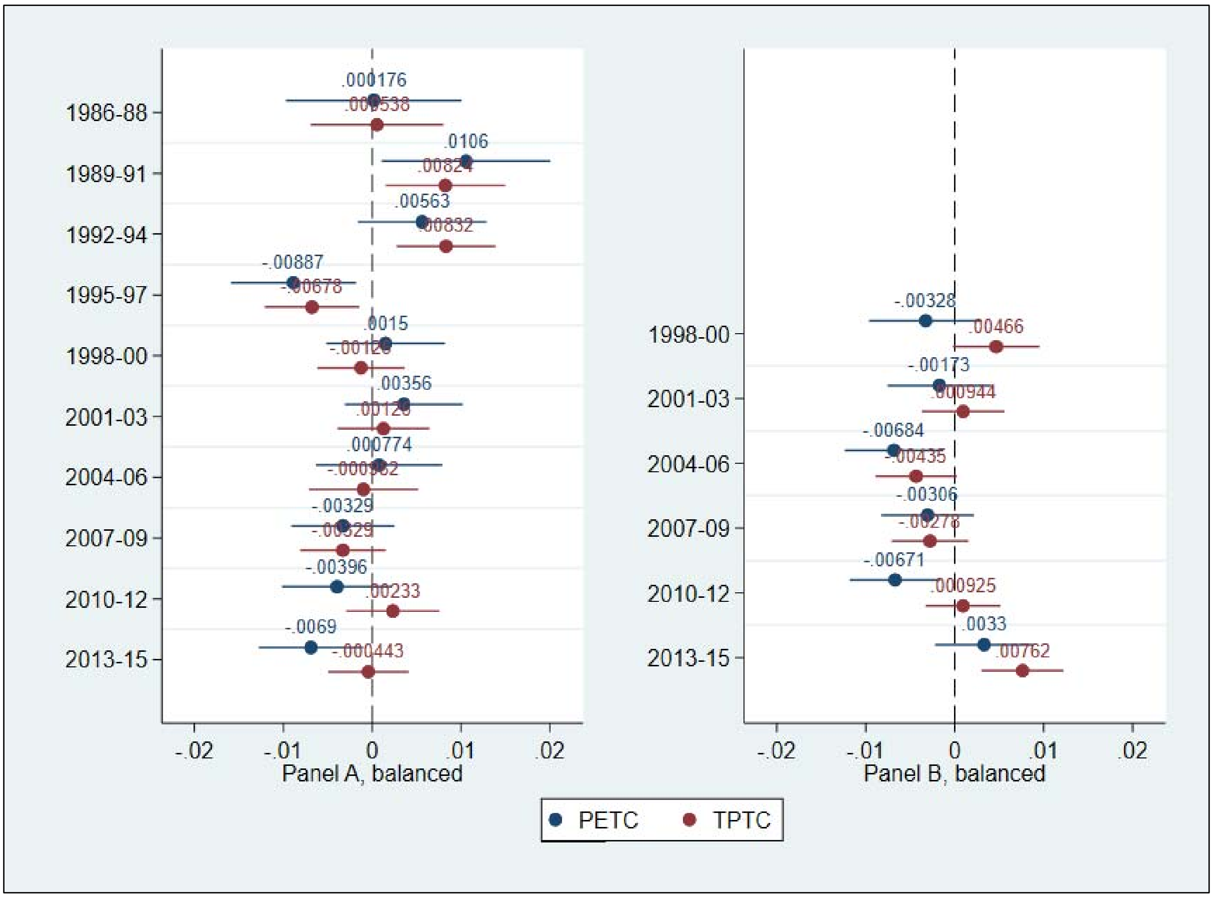

Figure 3: Time-varying international distance effects on partial *vs*. total bilateral trade cost, with lagged trade cost: Eq. (7)

	Panel A, balanced		Panel B, balanced	
	PBTC	TBTC	PBTC	TBTC
$BTC_{od,p-1}(dom)$	0.504 (0.100)***	0.939 (0.082)***	0.474 (0.104)***	0.740 (0.087)***
$BTC_{od,p-1}(int)$	0.475 (0.016)***	0.406 (0.019)***	0.337 (0.011)***	0.240 (0.012)***
Observations	25,669	25,669	32,195	32,195
Clusters	2,400	2,400	4,782	4,782
Pseudo-R^2	0.120	0.037	0.114	0.040
Imputed R^2	0.823	0.786	0.835	0.808

Notes: See Figure 1. Standard errors (clustered on country pairs) are bootstrapped based on 1,000 replications, to account for the presence of generated regressors, $BTC_{od,p-1}$(dom) and $BTC_{od,p-1}$(int).

[3] That difference is not due to panel breadth: we receive very results similar to panel B when we estimate panel A at panel B length.

These results are qualitatively robust to using all unbalanced data (see the appendix), to extending the policy variables in decomposition (5a) to WTO or EU membership, to decomposing only symmetric trade cost data in (5a), to varying the elasticity of substitution when transforming bilateral trade cost terms into partial bilateral trade cost estimates, and to estimating Eqs. (6) and (7) with a log-linear OLS specification.[4]

[4] All available upon request.

References

Allen, T., Arkolakis, C., and Y. Takahashi (2020). Universal gravity. *Journal of Political Economy 128*(2), 393–433

Anderson, J.E., and E. van Wincoop (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review 93*(1), 170–192

Anderson, J.E., and Y.V. Yotov (2020). Short run gravity. *Journal of International Economics 126*, 103341

Bergstrand, J., M. Larch, and Y.V. Yotov (2015). Economic integration agreements, border effects, and distance elasticities in the gravity equation. *European Economic Review 78*(C), 307–327

Carrère, C., M. Mrázová, and J.P. Neary (2020). Gravity without apology: The science of elasticities, distance and trade. *The Economic Journal 130*, 880–910

Coe, D., A, Subramanian, and N. Tamirisa (2007). The missing globalization puzzle: Evidence of declining importance of distance. *IMF Staff Papers 54*(1), 34–58

Correia, S. (2015). Singletons, cluster-robust standard errors and fixed effects: A bad mix. Mimeo

Correia, S., Guimarães, P., and Zylkin, T. (2019). Verifying the existence of maximum likelihood estimates for generalized linear models. Mimeo, revised June 2021. doi: arXiv:1903.01633

Correia, S., P. Guimarães, and T. Zylkin (2020). Fast Poisson estimation with high-dimensional fixed effects. *Stata Journal 20*(1), 95–115

Disdier, A., and K. Head (2008). The puzzling persistence of the distance effect on bilateral trade. *Review of Economics and Statistics 90*(1), 37–48

Egger, P.H., and S. Nigai (2015). Structural gravity with dummies only: Constrained ANOVA-type estimation of gravity models. *Journal of International Economics 97*(1), 86–99

Egger, P.H., and K.E. Staub (2016). GLM estimation of trade gravity models with fixed effects. *Empirical Economics 50*(1), 137–175

Egger, T., M. Larch, S. Nigai, and Y.V. Yotov (2021). Trade costs in the global economy: Measurement, aggregation and decomposition. Staff Working Paper ERSD-2021-2, WTO, Geneva

Gurevich, T., and P. Herman (2018). The dynamic gravity dataset: 1948–2016. USITC Working Paper 2018-02-A

Head, K., and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In: Gopinath, G., E. Helpman, and K. Rogoff (eds.), *Handbook of International Economics 1*(4). Oxford: Elsevier

Honoré, B.E., and M. Kesina (2017). Estimation of some nonlinear panel data models with both time-varying and time-invariant explanatory variables. *Journal of Business & Economic Statistics 35*(4), 543–558

Jacks, D. S., C.M. Meissner, and D. Novy (2008). Trade costs, 1870–2000. *American Economic Review: Papers & Proceedings 98*(2), 529–534.

Spornberger, J. (2022). EU integration and structural gravity: A comprehensive quantification of the border effect on trade. *Review of International Economics 30*(4), 915–938

Weidner, M., and T. Zylkin (2021). Bias and consistency in three-way gravity models. *Journal of International Economics 132*(4), 103513

Yotov, Y.V. (2012). A simple solution to the distance puzzle in international trade. *Economics Letters 117*(3), 794–798

Yotov, Y.V. (2022). Gravity at sixty: The workhorse model of trade. Working Paper 9584, CESifo, Munich

Appendix

Construction of domestic manufacturing trade data

We construct sets of international and domestic manufacturing trade data between 1983 and 2018 using several source data bases. International manufacturing trade from 1995 to 2018 are from the 2021 version of CEPII's BACI data set, reported at HS92 6-digit code and constructed by Gaulier and Zignano (2010), based on Comtrade, United Nations (2016). Earlier data on international manufacturing trade from 1983 to 1994 are from CEPII's TradeProd data set, reported at ISIC Rev. 2 3-digit level, constructed by de Sousa et al. (2012) in a similar manner as the BACI trade data. Domestic manufacturing trade is constructed using gross production data on manufacturing from UNIDO's (2021) Indstat 2 database, reported at ISIC Rev. 3 2-digit level, complemented by CEPII's TradeProd production data. The freely available Indstat 2 database reports data on gross production for the whole period 1983 to 2018 for 174 countries, although there are gaps in the data. CEPII's TradeProd data covers gross production between 1983 and 2006 and is mainly based on older versions of UNIDO's Indstat. Since Indstat 2 covers only manufacturing production, our data set is restricted to trade in manufacturing goods, as is usual for trade data sets which include domestic trade (Anderson and Yotov, 2016). We use Worldbank's World Integrated Trade Solution (WITS) concordance tables to match the different data sets.

We construct domestic trade first on ISIC Rev. 2 or ISIC Rev. 3 2-digit sector level, depending on the respective source data, for each country by subtracting total exports from gross production. The ISIC sector 'Recycling' is excluded since no international trade is reported for this sector. Afterwards, we aggregate domestic trade to the 8 sectors usually used in the gravity literature (Anderson and Yotov, 2016).[5] Thus, we end up with four different combinations of source data: CEPII's BACI trade data is matched with UNIDO's Indstat 2 production data for 1995–2018 and with CEPII's TradeProd production data for 1995–2006. For the period from 1983 to 1994, we separately combine trade data from TradeProd with Indstat 2 and TradeProd production data. We give preference to domestic trade constructed

[5] These are (1) Food, Beverages, and Tobacco Products; (2) Textile, Apparel, and Leather Products; (3) Wood and Wood Products; (4) Paper and Paper Products; (5) Chemicals, Petroleum, Coal, Rubber, and Plastic Products; (6) Other Non-metallic Products; (7) Basic Metal Products; (8) Fabricated Metal Products, Machinery, Equipment. The category 'Other manufacturing' is included in category (8).

with the newer Indstat 2 production data which is corrected *ex-post* and therefore more reliable. Domestic trade constructed with TradeProd production data is only used to fill in missing data between 1983 and 2006.[6]

For some observations, there is no data on internal trade, or it is not positive. This can be due to incomplete or wrong data on gross production, i.e. if small firms are not covered or production is allocated to the wrong sector. Additionally, there could be discrepancies between the year of production and the year of export. To handle this issue, we follow in large parts Baier et al. (2019). First, we replace single missing sectors by linear interpolation between years. If internal trade is non-positive for up to four sectors in a year, we replace them by the average expenditure share on domestic products in the respective year. After these steps, we aggregate domestic trade on country level. In single years without positive data on gross production, we inter- and extrapolate aggregated data linearly from adjacent years. Then, we merge the resulting data on domestic trade with international trade data from CEPII's BACI and TradeProd and thus receive a panel from 1983 to 2018, where the number of countries included varies by year.

From this data, we construct two panels of international and domestic manufacturing trade. Our first, longer panel A covers 55 major trading countries between 1983 and 2018. We apply the dynamic country code identifier of the USITC Dynamic Gravity Dataset (DGD, see Gurevich and Herman, 2018) such that uniquely defined countries do not experience significant changes in geographic or political characteristics while retaining their codes over time. In this sense, our panel A is fully balanced for 49 countries, i.e., for 6 out of 55 countries there are still gaps in constructed domestic trade. Our second, shorter but broader panel B covers 94 countries between 1995 and 2018, coming close to covering worldwide international trade in manufacturing goods. A balanced version of panel B still covers 71 countries, capturing about 75 percent of worldwide international trade in manufacturing goods.

[6] We calculate domestic trade separately for Indstat 2 (reported at ISIC Rev. 3) and TradeProd (reported at ISIC Rev. 2) production data. The resulting data differs sometimes, but a correlation of .99 between sectoral domestic trade constructed with TradeProd data and Indstat 2 data is reassuring. The small differences can be attributed to corrections in newer versions of Indstat 2 and BACI and slightly different sector definitions between ISIC Rev. 2 and ISIC Rev. 3.

Appendix – Tables

Table A1: First-stage PPML decompositions, Eq. (5a)

	Balanced		Unbalanced	
	Panel A	Panel B	Panel A	Panel B
Observations	84,495	111,353	95,980	189,753
Clusters	2,401	4,875	2,992	8,585
Pseudo-R^2	0.506	0.507	0.492	0.491
Imputed R^2	1.000	1.000	1.000	0.999

Notes: Estimations perform unsaturated constrained ANOVA models (5a). Imputed R^2 is the squared correlation between outcome and fitted values (see Egger and Staub, 2016).

Table A2: Figure 1 PPML estimation results, balanced panels

	Panel A		Panel B	
	Eq. (4)	Eq. (5b)	Eq. (4)	Eq. (5b)
ldist_1983–85	–0.138***	–0.142***		
	(0.011)	(0.011)		
ldist_1986–88	–0.107***	–0.111***		
	(0.010)	(0.009)		
ldist_1989–91	–0.101***	–0.105***		
	(0.010)	(0.010)		
ldist_1992–94	–0.0881***	–0.0902***		
	(0.009)	(0.008)		
ldist_1995–97	–0.0450***	–0.0479***	–0.0599***	–0.0566***
	(0.006)	(0.006)	(0.006)	(0.006)
ldist_1998–2000	–0.0257***	–0.0275***	–0.0383***	–0.0359***
	(0.006)	(0.005)	(0.006)	(0.005)
ldist_2001–03	–0.0199***	–0.0219***	–0.0261***	–0.0241***
	(0.005)	(0.005)	(0.006)	(0.005)
ldist_2004–06	–0.0135***	–0.0153***	–0.0196***	–0.0180***
	(0.005)	(0.005)	(0.005)	(0.005)
ldist_2007–09	–0.00989**	–0.0116***	–0.0153***	–0.0142***
	(0.004)	(0.004)	(0.004)	(0.004)
ldist_2010–12	–0.00703**	–0.0082***	–0.0101***	–0.00933***
	(0.003)	(0.003)	(0.003)	(0.003)
ldist_2013–15	–0.00594***	–0.00684***	–0.00757***	–0.00733***
	(0.002)	(0.002)	(0.002)	(0.002)
RTA(t)	0.010		–0.004	
	(0.070)		(0.032)	
RTA(t–1)	0.036***		0.026	
	(0.014)		(0.022)	
RTA(t–2)	0.060***		0.019	
	(0.014)		(0.025)	
RTA(t–3)	0.044**		0.034	
	(0.018)		(0.021)	
RTA(t–4)	0.175***		0.025	
	(0.060)		(0.039)	
Observations	86,436	28,165	115,328	37,334
Clusters	2,401	2,401	4,875	4,842
Pseudo-R^2	0.506	0.741	0.508	0.760
Imputed R^2	0.999	0.999	0.999	0.999

Notes: See Figure 1.

Table A3: Figure 1 PPML estimation results, unbalanced panels

	Panel A		Panel B	
	Eq. (4)	Eq. (5b)	Eq. (4)	Eq. (5b)
ldist_1983–85	–0.142***	–0.145***		
	(0.011)	(0.011)		
ldist_1986–88	–0.113***	–0.117***		
	(0.010)	(0.009)		
ldist_1989–91	–0.102***	–0.105***		
	(0.010)	(0.010)		
ldist_1992–94	–0.091***	–0.093***		
	(0.008)	(0.008)		
ldist_1995–97	–0.051***	–0.053***	–0.0635***	–0.0609***
	(0.007)	(0.006)	(0.006)	(0.006)
ldist_1998–2000	–0.031***	–0.032***	–0.0408***	–0.0388***
	(0.006)	(0.006)	(0.006)	(0.006)
ldist_2001–03	–0.024***	–0.026***	–0.0315***	–0.0296***
	(0.005)	(0.005)	(0.005)	(0.005)
ldist_2004–06	–0.017***	–0.018***	–0.0243***	–0.0225***
	(0.005)	(0.005)	(0.005)	(0.005)
ldist_2007–09	–0.011***	–0.013***	–0.0194***	–0.0181***
	(0.004)	(0.004)	(0.004)	(0.004)
ldist_2010–12	–0.008***	–0.009***	–0.0142***	–0.0131***
	(0.003)	(0.003)	(0.003)	(0.003)
ldist_2013–15	–0.006***	–0.007***	–0.0118***	–0.0113***
	(0.002)	(0.001)	(0.002)	(0.002)
RTA(t)	–0.002		–0.086	
	(0.067)		(0.028)	
RTA(t–1)	0.036***		0.008	
	(0.012)		(0.019)	
RTA(t–2)	0.060***		0.012	
	(0.013)		(0.020)	
RTA(t–3)	0.025		–0.002	
	(0.019)		(0.026)	
RTA(t–4)	0.183***		0.044	
	(0.058)		(0.038)	
Observations	98,562	31,945	200,347	63,478
Clusters	2,993	2.,910	8,585	8,509
Pseudo-R^2	0.493	0.743	0.493	0.755
Imputed R^2	0.999	0.999	0.999	0.999

Notes: See Table A2.

Table A4: Figure 2 PPML estimation results, balanced panels

	Panel A		Panel B	
	PBTC	TBTC	PBTC	TBTC
ldist_1983–85	0.0279***	0.0248***		
	(0.007)	(0.006)		
ldist_1986–88	0.0197***	0.0169***		
	(0.006)	(0.005)		
ldist_1989–91	0.0204***	0.0174***		
	(0.006)	(0.005)		
ldist_1992–94	0.0154**	0.0148***		
	(0.006)	(0.005)		
ldist_1995–97	–0.000447	–0.000261	–0.00165	0.0161***
	(0.004)	(0.003)	(0.004)	(0.003)
ldist_1998–2000	0.00498	0.000633	–0.00518	0.00764***
	(0.004)	(0.003)	(0.004)	(0.003)
ldist_2001–03	0.00882**	0.00223	–0.00182	0.00305
	(0.004)	(0.003)	(0.004)	(0.003)
ldist_2004–06	0.00626	–0.000759	–0.00781**	–0.00449*
	(0.004)	(0.003)	(0.003)	(0.003)
ldist_2007–09	0.00074	–0.00493*	–0.00566*	–0.00517**
	(0.003)	(0.003)	(0.003)	(0.002)
ldist_2010–12	–0.00197	–0.000697	–0.00955***	–0.00273
	(0.003)	(0.003)	(0.003)	(0.002)
ldist_2013–15	–0.00572*	–0.00128	0.000422	0.00597***
	(0.002)	(0.002)	(0.003)	(0.002)
Observations	28,165	28,165	37,334	37,334
Clusters	2,401	2,401	4,842	4,842
Pseudo-R^2	0.129	0.039	0.118	0.042
Imputed R^2	0.752	0.706	0.792	0.773

Notes: See Figure 2.

Table A5: Figure 2 PPML estimation results, unbalanced panels

	Panel A		Panel B	
	PBTC	TBTC	PBTC	TBTC
ldist_1983–85	0.0263***	0.0245***		
	(0.007)	(0.006)		
ldist_1986–88	0.0225***	0.0171***		
	(0.006)	(0.005)		
ldist_1989–91	0.0224***	0.0188***		
	(0.006)	(0.005)		
ldist_1992–94	0.0155***	0.0168***		
	(0.006)	(0.005)		
ldist_1995–97	0.00256	0.000232	0.0133***	0.0176***
	(0.004)	(0.003)	(0.004)	(0.003)
ldist_1998–2000	0.00598*	–0.000626	0.00869**	0.0084***
	(0.004)	(0.003)	(0.004)	(0.003)
ldist_2001–03	0.0099***	0.00141	0.0212***	0.00879***
	(0.004)	(0.003)	(0.003)	(0.003)
ldist_2004–06	0.00476	–0.00309	0.0147***	0.00147
	(0.004)	(0.003)	(0.003)	(0.003)
ldist_2007–09	0.000088	–0.00614**	0.0143***	–0.000612
	(0.003)	(0.003)	(0.003)	(0.002)
ldist_2010–12	–0.00167	–0.000604	0.00942***	0.00396*
	(0.003)	(0.003)	(0.003)	(0.002)
ldist_2013–15	–0.00514**	–0.00123	0.0133***	0.0109***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	31,945	31,945	63,478	63,478
Clusters	2,910	2,910	8,509	8,509
Pseudo- R^2	0.125	0.038	0.119	0.043
Imputed R^2	0.755	0.708	0.761	0.744

Notes: See Table A4.

Table A6: Figure 3 PPML estimation results, balanced panels

	Panel A		Panel B	
	PBTC	TBTC	PBTC	TBTC
$BTC_{od,p-1}(dom)$	0.504***	0.939***	0.474***	0.740***
	(0.100)	(0.082)	(0.104)	(0.087)
$BTC_{od,p-1}(int)$	0.475***	0.406***	0.337***	0.240***
	(0.016)	(0.019)	(0.011)	(0.012)
ldist_1986–88	0.000176	0.000538		
	(0.005)	(0.004)		
ldist_1989–91	0.0106**	0.00824**		
	(0.005)	(0.003)		
ldist_1992–94	0.00563	0.00832***		
	(0.004)	(0.003)		
ldist_1995–97	–0.00887**	–0.00678**		
	(0.004)	(0.003)		
ldist_1998–2000	0.0015	–0.00126	–0.00328	0.00466*
	(0.003)	(0.003)	(0.003)	(0.003)
ldist_2001–03	0.00356	0.00126	–0.00173	0.000944
	(0.003)	(0.003)	(0.003)	(0.003)
ldist_2004–06	0.000774	–0.000982	–0.00684**	–0.00435*
	(0.004)	(0.003)	(0.003)	(0.002)
ldist_2007–09	–0.00329	–0.00329	–0.00306***	–0.00278
	(0.003)	(0.002)	(0.003)	(0.002)
ldist_2010–12	–0.00396	0.00233	–0.00671**	0.000925
	(0.003)	(0.003)	(0.003)	(0.002)
ldist_2013–15	–0.0069**	–0.000443	0.0033	0.00762***
	(0.003)	(0.002)	(0.003)	(0.002)
Observations	25,669	25,669	32,195	32,195
Clusters	2,400	2,400	4,782	4,782
Pseudo-R^2	0.120	0.037	0.114	0.040
Imputed R^2	0.823	0.786	0.835	0.808

Notes: See Figure 3.

Table A7: Figure 3 PPML estimation results, unbalanced panels

	Panel A		Panel B	
	PBTC	TBTC	PBTC	TBTC
$BTC_{od,p-1}(dom)$	0.518***	0.909***	0.516***	0.695***
	(0.093)	(0.080)	(0.076)	(0.073)
$BTC_{od,p-1}(int)$	0.466***	0.399***	0.295***	0.223***
	(0.015)	(0.016)	(0.009)	(0.010)
ldist_1986–88	0.00423	0.000547		
	(0.005)	(0.004)		
ldist_1989–91	0.0117***	0.0100***		
	(0.004)	(0.003)		
ldist_1992–94	0.00532	0.00974***		
	(0.003)	(0.003)		
ldist_1995–97	–0.00624*	–0.00759***		
	(0.003)	(0.003)		
ldist_1998–2000	0.00168	–0.00222***	0.00667**	0.00567**
	(0.003)	(0.002)	(0.003)	(0.002)
ldist_2001–03	0.00434	0.000765	0.0205***	0.00843***
	(0.003)	(0.003)	(0.003)	(0.002)
ldist_2004–06	–0.00115	–0.00321	0.0134***	0.00306
	(0.003)	(0.003)	(0.003)	(0.002)
ldist_2007–09	–0.00301	–0.00368	0.0148***	0.00289
	(0.003)	(0.002)	(0.003)	(0.002)
ldist_2010–12	–0.00311	0.00284	0.00835***	0.00635***
	(0.003)	(0.003)	(0.003)	(0.002)
ldist_2013–15	–0.00632**	–0.000548	0.0156***	0.0131***
	(0.003)	(0.002)	(0.003)	(0.002)
Observations	28,822	28,822	54,053	54,053
Clusters	2,808	2,808	8,170	8,170
Pseudo-R^2	0.115	0.036	0.114	0.041
Imputed R^2	0.822	0.780	0.805	0.786

Notes: See Table A6.

References

Anderson, J.E., and Y.V. Yotov (2016). Terms of trade and global efficiency effects of free trade agreements, 1990–2002. *Journal of International Economics 99*, 279–298

Baier, S. L., Y.V. Yotov, and T. Zylkin (2019). On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration. *Journal of International Economics 116*, 206–226

de Sousa, J., T. Mayer, and S. Zignago (2012). Market access in global and regional trade. *Regional Science and Urban Economics 42*(6), 1037–1052

Egger, P. H., and Staub, K. E. (2016). GLM estimation of trade gravity models with fixed effects. *Empirical Economics 50*(1), 137–175

Gaulier, G., and S. Zignago (2010). BACI: International trade database at the product-level. The 1994–2007 version. CEPII Working Paper, N° 2010-23.

Gurevich, T, and P. Herman (2018). The dynamic gravity dataset: 1948–2016. USITC Working Paper 2018-02-A

UNIDO (2021). INDSTAT 2 Industrial Statistics Database at 2-digit level of ISIC Revision 3. Vienna. Available from http://stat.unido.org

United Nations (2016). Comtrade Database, United Nations Commodity Trade Statistics. Retrieved from http://comtrade.un.org