Received: 15 August 2023 | Revised: 19 February 2024 | Accepted: 27 February 2024

DOI: 10.1111/twec.13569

ORIGINAL ARTICLE

The World Economy WILEY

Estimating the effects of trade agreements: Lessons from 60 years of methods and data

Mario Larch[1,2,3,4] | **Yoto V. Yotov**[3,5]

[1]Department of Law, Business and Economics, University of Bayreuth, Bayreuth, Germany

[2]Centre d'Etudes Prospectives et d'Informations Internationales (CEPII), Paris, France

[3]ifo Institute and CESifo Research Network, Munich, Germany

[4]Nottingham Centre for Research on Globalisation and Economic Policy (GEP), Nottingham, UK

[5]School of Economics, Drexel University, Philadelphia, Pennsylvania, USA

Correspondence
Mario Larch, Department of Law, Business and Economics, University of Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany.
Email: mario.larch@uni-bayreuth.de

Abstract

Starting with Tinbergen (1962, *Shaping the world economy: Suggestions for an international economic policy*, The Twentieth Century Fund), quantifying the effects of regional trade agreements (RTAs) on international trade flows has always been among the most popular topics in the trade literature. Also not surprisingly, to estimate the effects of RTAs, most researchers and policy analysts have relied on the workhorse model of trade—the gravity equation. Over the past 60 years, there have been many important developments in the RTA literature, both in terms of better methods to quantify their effects, and in terms of more and higher quality data. The objective of this paper is to trace the evolution of the methods and data developments in the RTA literature, from Tinbergen's very first exploration until today, and to critically evaluate their significance for our ability to measure the impact of RTAs (and other policies) on international trade.

KEYWORDS

data, estimation, gravity equation, methods, regional trade agreements

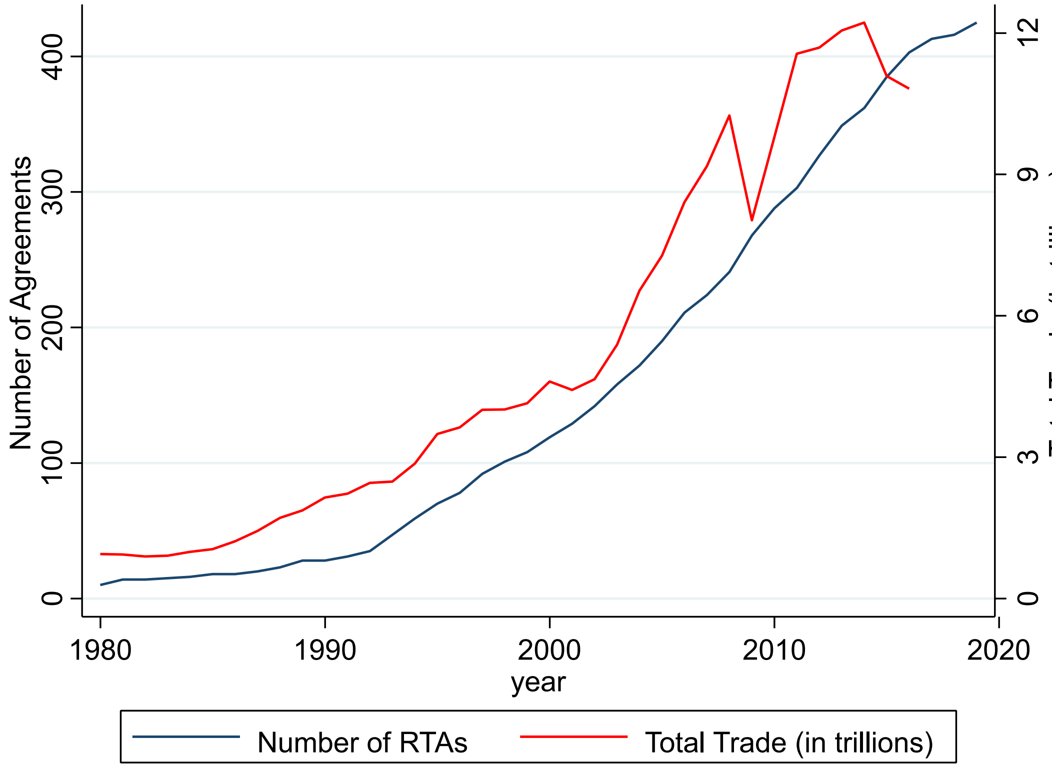


FIGURE 1 The evolution of regional trade agreements (RTAs) and manufacturing trade. This figure plots the evolution of RTAs between 1980 and 2019 on the left *Y*-axis, and the evolution of aggregate manufacturing trade flows between 1980 and 2016 on the right *Y*-axis. The data on RTAs come from Egger and Larch (2008) and the trade data come from Larch et al. (2019). See Section 2 for further details on the data. [Colour figure can be viewed at wileyonlinelibrary.com]

1 | INTRODUCTION

Over the past 60 years, the world economy witnessed a remarkable proliferation of trade agreements, accompanied by a matching and equally remarkable increase in international trade flows. Figure 1 captures these patterns very clearly by plotting the evolution of regional trade agreements (RTAs) over the past 40 years (1980–2019) as well as the corresponding increase in total manufacturing trade between 1980 and 2016 for the 89 countries used in our main analysis. The correlation between the two variables in Figure 1 is strong (0.96) and obvious. Nevertheless, many have questioned the effectiveness of RTAs in promoting international trade; at its extreme, some have even doubted the causal links between RTAs and trade flows altogether. What are the channels through which RTAs promote trade? How to design more efficient and effective trade agreements? How to estimate the partial effects of trade agreements? How to account for the potential endogeneity of trade agreements in our econometric analysis? How to quantify their general equilibrium effects? This is only a small set of questions related to RTAs that have attracted significant attention among academic and policy analysts alike.

Against this backdrop, and perhaps not surprisingly, trade agreements, their design, their implementation, and quantifying their impact on trade flows are among the most popular topics in the international trade literature.[1] It is also a fact that the trade gravity equation, whose 60th anniversary we marked in 2022,[2] is the workhorse model for analysing both the partial and the general equilibrium effects of RTAs. Since Tinbergen's first exploration of the effects of trade agreements in 1962, there have been many significant developments in the gravity and RTA literature, both in terms of better methods to quantify their effects, and in terms of more and higher quality data.

Stimulated by the 60th gravity anniversary and the significant interest in quantifying the effects of RTAs, the objectives of this paper are to trace the evolution of the methods and data developments in the gravity and RTA literature, from Tinbergen's very first exploration until today, and to critically evaluate their significance for our ability to measure the impact of RTAs (and other policies) on international trade. We pursue these goals in two broad steps. First, in Section 2, we use the same dataset to explore the impact of various methods that have been used to estimate the effects of RTAs.[3] Then, in Section 3, we rely on an econometric model, which accounts for all major developments in the literature, to study the heterogeneous effects of RTAs.

To develop the analysis in Section 2, we follow the evolution over time of the methods used to estimate the effects of RTAs. Specifically, we start with a naive gravity specification, which, in the spirit of Tinbergen (1962), 'borrows' gravity directly from physics and only includes a limited number of covariates that proxy for distance and economic mass. Then, we gradually introduce what we believe were the major developments in the related literature, including additional control variables, controlling for multilateral resistance (MR) terms, using pair fixed effects, estimating gravity with the Poisson pseudo maximum likelihood (PPML) estimator, using domestic sales in addition to international trade flows, accounting for globalisation effects, relying on alternative clustering for standard errors, correcting for potential incidental parameter problem (IPP), and using consecutive-year data (as opposed to interval or averaged trade data).

Our main findings are summarised in Figure 2 and, based on this, we draw four main conclusions. First, there have been many important methodological contributions that should be considered when estimating the effects of RTAs with the gravity model. Second, some of the developments we review and implement have had a significant impact on the RTA estimates, while others have not so much. However, third, even if some of our RTA results are not very sensitive to alternative specifications, we show that the same methods could be very important for estimating the effects of other gravity covariates and for quantifying heterogeneous RTA effects across various dimensions. Finally, the most recent methodological contributions focused on standard errors and therefore led to relatively minor changes/improvements in the RTA point estimates. While we find this conclusion to be intuitive and encouraging, we also believe there are significant opportunities for further improvements and contributions, some of which we highlight when presenting the heterogeneous effects of RTAs.

[1]A google scholar search on 'trade agreements' that we performed on 11 December 2022, delivered 3,750,000 results. For comparison, a search on 'economic sanctions' yielded 1,560,000 results, and a search on 'covid pandemic' returned 1,610,000. Moreover, after completing this paper, we estimated that the most relevant papers that we cite amount to about 60,000 citations.

[2]Ravenstein (1885) is the first to apply the gravity equation to economics (migration flows). However, Tinbergen (1962) is the first to apply gravity to international trade flows. Hence, the 60th gravity anniversary in 2022.

[3]To highlight the importance of methods, we rely on a single dataset, which covers total manufacturing for 89 countries (accounting for 95% of world trade) during the period 1980–2016.

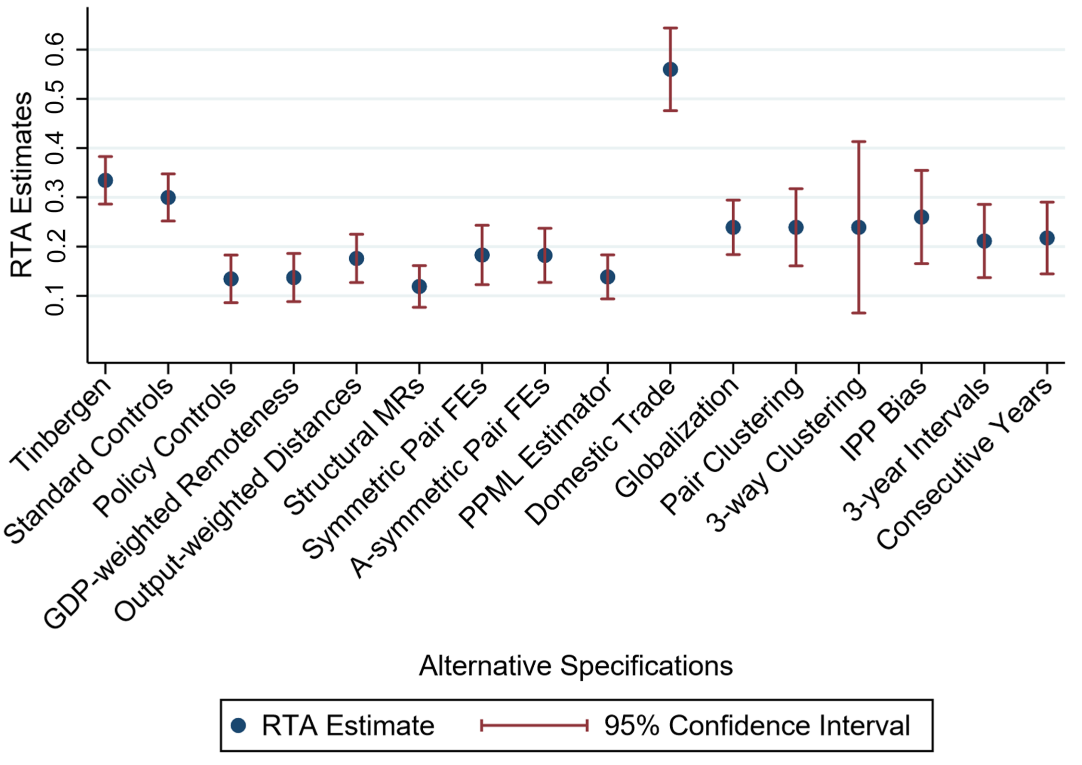


FIGURE 2 Sixty years of regional trade agreement (RTA) estimates with the gravity model of trade. This figure plots the estimates of the effects of RTAs, along with the corresponding confidence intervals, that we obtain in Section 2 and report in Table 1. The *X*-axis of the figure lists the alternative specifications that we use, which follow the evolution of the gravity methods from the 1962 specification of 'Tinbergen' to the structural gravity specification with 'Consecutive Years' of Egger et al. (2022). We refer the reader to Table 1 and to the discussion in the main text for further details. [Colour figure can be viewed at wileyonlinelibrary.com]

We study the heterogeneous effects of RTAs in Section 3. We rely on a benchmark econometric model that incorporates all major developments we reviewed, and we utilise some new datasets and allow for RTA heterogeneity across various dimensions. This analysis reveals significant heterogeneity in the evolution of RTA estimates over time, across sectors, across individual agreements, depending on the type and depth of the RTAs, across country pairs within a multilateral trade agreement, and even within pairs depending on the direction of trade flows.

The wide heterogeneity of RTA effects we obtain has several important implications. First, average RTA estimates may hide significant heterogeneity across various dimensions. Second, researchers should keep this caveat in mind when RTA estimates are used for counterfactual analysis. Third, even though most of the RTA estimates that we obtain are positive (and many of them statistically significant too), we also obtain many negative and statistically significant estimates. From a methodological (and data) perspective, this implies researchers should be careful and ensure that such estimates are sound.

Finally, from a broader perspective, an important implication of our analysis in Section 3 is that the econometric gravity model can be used to obtain not only aggregate but also very disaggregated (e.g., case study) RTA estimates. In combination with the solid theoretical foundations of the gravity model of trade and the fact that the gravity system can be seen as a computable general equilibrium (CGE) model that can be nested in more complex production models, this implies that gravity estimates can be used to quantify both the partial and GE effects of RTAs.

The rest of the paper is organised as follows. Section 2 uses the same data to trace the evolution of the methods to estimate the effects of RTAs. Section 3 uses our most preferred econometric model from Section 2 and relies on alternative datasets to study the heterogeneous effects of RTAs across various dimensions. Finally, Section 4 summarises our main findings and points to directions for future research.

2 | ESTIMATING THE EFFECTS OF RTAS

This section traces the evolution of the methodological developments in the gravity literature and discusses their implications for estimating the effects of RTAs.[4] For consistency, throughout the analysis we rely on the same dataset. Specifically, we use the aggregate manufacturing dataset from Larch et al. (2019), which has several advantages for our purposes. First, it covers a long period of time (1980–2016), which, as demonstrated in the introduction, coincides with the remarkable growth in the formation of RTAs. Second, it covers many countries and, therefore, most trade and RTAs in the world.[5] Third, it includes consistently constructed domestic and international trade. In addition to trade data, we use the RTA dataset of Egger and Larch (2008) for information on trade agreements.[6] Finally, we rely on the *Dynamic Gravity Dataset* of Gurevich and Herman (2018) for the other gravity controls (e.g., distance, colonial ties, etc.) in our analysis.

We start with a simple and naive gravity specification that corresponds to Tinbergen (1962), who 'borrowed' the gravity equation from physics to describe bilateral trade flows as a function of economic masses and trade frictions:

$$\ln\_X_{ij,t} = \beta_1\, RTA_{ij,t} + \beta_2 \ln\_DIST_{ij} + \beta_3\; \ln\_GDP_{i,t} + \beta_4 \ln\_GDP_{j,t} + \epsilon_{ij,t}. \qquad (1)$$

Here, $\ln\_X_{ij,t}$ is the logarithm of exports from source/exporter i to destination/importer j at time t. Consistent with most of the gravity literature, bilateral trade is measured at the yearly level and we use 5-year interval data (Baier & Bergstrand, 2007; Cheng & Wall, 2005; Olivero & Yotov, 2012). $RTA_{ij,t}$ is an indicator variable for the presence of an RTA between countries i and j at time t. This variable will be the focus of our analysis. $\ln\_DIST_{ij}$, which is measured as the logarithm of the population-weighted bilateral distance between i and j, is the most widely used and robust proxy for trade costs (Disdier & Head, 2008). Finally, $\ln\_GDP_{i,t}$ and $\ln\_GDP_{j,t}$ are the gross domestic products of the exporter and the importer, which proxy for country size.

[4]Following most of the existing literature, we focus on the average effect of RTAs in this section. In the next section, we estimate the heterogeneous effects of trade agreements across various dimensions, including by type of RTA (e.g., free trade agreements [FTAs] vs. customs unions [CUs], etc.).

[5]The original dataset of Larch et al. (2019) covers 229 countries. For the analysis in this paper, we limited the number of countries to 89, covering more than 95% of manufacturing trade in the world. Our main results and conclusions are robust to include all 229 countries from the original dataset.

[6]Two other popular trade agreement datasets are the NSF-Kellogg Institute Database on Economic Integration Agreements, developed Baier and Bergstrand (2021) and the World Bank Deep Trade Agreement database, developed by Mattoo et al. (2020). We rely on the latter database in the next section, where we offer estimates of the effects of RTAs depending on their depth.

TABLE 1 Sixty years of free trade agreement estimates with the gravity model of trade.

Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$RTA_{ij,t}$	0.334	0.300	0.134	0.137	0.176	0.119	0.183	0.182
	(0.025)**	(0.024)**	(0.025)**	(0.025)**	(0.025)**	(0.022)**	(0.031)**	(0.028)**
$DIST_{ij}$	−1.257	−1.218	−1.276	−1.268	−1.157	−1.378		
	(0.013)**	(0.013)**	(0.013)**	(0.017)**	(0.016)**	(0.014)**		
$\ln\_GDP_{i,t}$	1.227	1.235	1.211	1.206	1.189			
	(0.007)**	(0.006)**	(0.006)**	(0.006)**	(0.006)**			
$\ln\_GDP_{j,t}$	0.942	0.950	0.925	0.929	0.924			
	(0.007)**	(0.007)**	(0.006)**	(0.007)**	(0.007)**			
$CLNY_{ij}$		0.741	0.744	0.740	0.698	0.514		
		(0.048)**	(0.047)**	(0.048)**	(0.047)**	(0.049)**		
$CNTG_{ij}$		0.283	0.334	0.342	0.460	0.220		
		(0.051)**	(0.050)**	(0.051)**	(0.050)**	(0.048)**		
$LANG_{ij}$		0.623	0.564	0.570	0.661	0.730		
		(0.022)**	(0.022)**	(0.024)**	(0.024)**	(0.023)**		
$WTO_{ij,t}$			0.797	0.798	0.815	0.304	0.010	−0.008
			(0.026)**	(0.026)**	(0.026)**	(0.055)**	(0.062)	(0.057)
$REM\_GDP_{i,t}$				−0.408				
				(0.054)**				
$REM\_GDP_{j,t}$				0.351				
				(0.056)**				
$REM\_OUT_{i,t}$					−0.922			
					(0.049)**			
$REM\_EXP_{j,t}$					−0.022			
					(0.050)			
N	47,035	47,035	47,035	47,035	47,035	48,254	48,247	48,134

TABLE 1 (Continued)

Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R^2	.573	.583	.593	.595	.597	.797	.870	.901
Panel B	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**	**(16)**
RTA	0.138	0.560	0.239	0.239	0.239	0.260	0.211	0.217
	(0.023)**	(0.043)**	(0.028)**	(0.040)**	(0.089)**	(0.052)**	(0.038)**	(0.037)**
WTO	−0.024	0.506	0.224	0.224	0.224	0.209	0.217	0.263
	(0.065)	(0.048)**	(0.036)**	(0.049)**	(0.121)$^{+}$	(0.065)**	(0.055)**	(0.049)**
N	49,085	49,625	49,625	49,625	49,625	49,625	81,614	233,242

Note: This table includes two panels, which report a series of gravity estimates of the effects of regional trade agreements (RTAs) that trace the evolution of the methods from the related literature. The RTA estimates, along with the corresponding confidence intervals, are visualised in Figure 2. The estimates in Panel A (i.e., columns (1) to (8)) are obtained with the OLS estimator. The estimates in columns (1) to (14) are obtained with 5-year interval data. The estimates in columns (1) to (9) are obtained with international data only. Column (1) replicates the results from Tinbergen (1962). Column (2) introduces the so-called 'standard gravity covariates'. Column (3) adds an indicator for GATT/WTO membership. Columns (4) to (6) implement different methods to control for the MR terms. Column (7) and (8) add symmetric and directional pair fixed effects, respectively. Column (9) uses the PPML estimator. Column (10) adds observations for domestic trade flows. Column (11) controls for common globalisation trends. Columns (12) and (13) cluster the standard errors by pair and 3-way (by exporter, importer, and time). Column (14) corrects for possible IPP bias. Finally, columns (15) and (16) employ 3-year interval data and data for consecutive years, respectively. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$. See text for further details.

The estimates of Equation (1) appear in column (1) of Table 1. In addition, Figure 2 traces the evolution of all RTA estimates that we obtain from alternative specifications. Overall, and despite using different data, our results are comparable to Tinbergen's. Most important for our purposes, we obtain a positive and statistically significant estimate of the effects of RTAs, which implies that all else equal, an RTA between two countries in our sample leads to an increase in trade of 39.7% (std.err. 3.44).[7] Consistent with the voluminous gravity literature, we obtain a large, negative, and statistically significant estimate of the impact of distance on international trade (−1.257, std.err. 0.013). The positive estimates on $ln\_GDP_{i,t}$ and $ln\_GDP_{j,t}$ reveal that, as expected, country size is a significant determinant of trade. Finally, we obtain an R^2 of .57, which is a bit lower than Tingergen's ($R^2 = .7$) but, nevertheless, revealing the great predictive power of the gravity equation, even when it just includes four intuitive covariates.

Due to its intuitive appeal and remarkable predictive power, the naive gravity equation of Tinbergen (1962) quickly gained popularity. Given the purposes of this paper, we select a few variables commonly included in gravity regressions and, just like hundreds of early gravity papers, we add them as covariates to the previous specification, which becomes[8]:

$$\begin{aligned} ln\_X_{ij,t} = \; & \beta_1 \, RTA_{ij,t} + \beta_2 \, ln\_DIST_{ij} + \beta_3 \, ln\_GDP_{i,t} + \beta_4 \, ln\_GDP_{j,t} + \beta_5 \, CLNY_{ij} + \\ & \beta_6 \, CNTG_{ij} + \beta_7 \, LANG_{ij} + \beta_8 \, WTO_{ij,t} + \epsilon_{ij,t}, \end{aligned} \tag{2}$$

where we have split the new covariates into two groups. $CLNY_{ij}$, $CNTG_{ij}$, and $LANG_{ij}$ are time-invariant dummy variables for the presence of colonial ties, contiguous borders, and common official language, respectively. Together with the logarithm of the bilateral distance, these variables comprise the group of 'standard gravity covariates'.

In addition, we have added a time-varying indicator variable for membership in the GATT and WTO ($WTO_{ij,t}$). The motivation is twofold. First, it is a widely used policy variable in the gravity model.[9] Second, given the objective of GATT and WTO to promote trade and trade liberalisation between member countries, we expect that omitting this may lead to biased RTA estimates.

We estimate Equation (2) in two steps. First, we only include the time-invariant control variables. Then, we introduce the WTO dummy. Our findings are reported in columns (2) and (3) of Table 1, respectively. The estimates in column (3) reveal that sharing a common language, common borders, and colonial ties are all catalysts for bilateral trade. Second, and more important for our purposes, we see that the estimate on RTA is just a bit smaller than the estimate in column (1). Pushing inference to the limit, this suggests a positive correlation between RTAs and the new covariates. This is intuitive, i.e., countries that already have close economic ties (e.g., due to common language, borders, and colonial ties) are more likely to sign trade agreements.

The results in column (3) of Table 1 are obtained after controlling for GATT/WTO membership. The estimate on $WTO_{ij,t}$ is positive and statistically significant, as expected. Importantly, the introduction of the indicator for GATT/WTO membership has a very significant impact on the RTA estimate, which is more than twice smaller in magnitude. This change is expected, and the intuition is that our RTA variable has captured effects that should be attributed to the GATT or WTO. The broader implication for estimating RTA effects with the gravity model is that the

[7]Obtained as $(\exp(\beta_1) - 1) \times 100 = (\exp(0.334) - 1) \times 100$, where standard errors are obtained with the Delta method.

[8]We recognise that we are abusing notation by not changing the notation for $\epsilon_{ij,t}$. However, for simplicity, we will just use the same notation for the error term throughout the analysis.

[9]See Campos et al. (2022) for a recent survey and meta-analysis of the effects of GATT and WTO.

omission of other trade liberalisation (or protection) policy variables may lead to significant biases in the estimates of the RTA effects.

One of the major theoretical developments in the gravity literature is the introduction of the MR terms by Anderson and van Wincoop (2003). Intuitively, the MRs capture the fact that trade between two countries depends not only on their sizes and the bilateral trade costs between them but also on how remote (geographically and economically) these countries are from the rest of the world.[10] To underscore the importance of the MRs Baldwin and Taglioni (2006) dub the omission to control for them 'the gold medal mistake' in gravity estimations. To account for MRs, we introduce two additional terms to our estimating equation, which becomes:

$$\begin{aligned} \ln\_X_{ij,t} = \; & \beta_1\, RTA_{ij,t} + \beta_2 \ln\_DIST_{ij} + \beta_3 \ln\_GDP_{i,t} + \beta_4 \ln\_GDP_{j,t} + \beta_5\, CLNY_{ij} + \\ & \beta_6\, CNTG_{ij} + \beta_7\, LANG_{ij} + \beta_8\, WTO_{ij,t} + \beta_9 \ln\_OMR_{i,t} + \beta_{10} \ln\_IMR_{j,t} + \epsilon_{ij,t}, \end{aligned} \tag{3}$$

where $OMR_{i,t}$ denotes the 'outward multilateral resistance', which falls on the producers in exporter i, and $IMR_{j,t}$ denotes the 'inward multilateral resistance', which falls on the consumers in importer j.[11]

We use three methods to account for MRs.[12] First, we proxy for them with the so-called 'remoteness indexes' (Wei, 1996) constructed as GDP-weighted distances. The results from column (4) of Table 1 reveal that the RTA estimate has changed very little. Moreover, we see that the estimates on the MR terms have opposing signs. A possible explanation is that the GDP-weighted 'remoteness indexes' have (at least) three deficiencies relative to the structural MR terms: they do not take into account all bilateral trade costs but only distance; they do not account for aggregate trade imbalances; and they do not reflect the inverse structural relationship between the inward and the outward resistances.

The estimates in column (5) of Table 1 are obtained with MR indexes that, once again, use only distance as a proxy for bilateral trade costs, however, instead of GDP weights, we use values for gross output and gross expenditures: these are theory-consistent size variables from the structural gravity system (Arkolakis et al., 2012) and the use of output and expenditure weights reflects the presence of trade imbalances, which, as recently demonstrated by Felbermayr and Yotov (2021), is a necessary condition for the empirical performance of the gravity model. The RTA estimate in column (5) is significantly larger in percentage terms.

As initially advocated by Hummels (2001) and popularised by Feenstra (2004) and Redding and Venables (2004), the most common practice to control for MRs is with exporter and importer fixed effects. Baldwin and Taglioni (2006) emphasise that, as a function of time-varying bilateral trade costs, MRs are also time-varying, which implies that the proper set of fixed effects in panel gravity regressions should be of dimension exporter-time and importer-time.[13] Applied to our

[10]Yotov et al. (2016) summarise several attractive features of the structural multilateral resistance terms.

[11]An alternative structural interpretation of the inward multilateral resistance is as an ideal consumer price index (Anderson & Yotov, 2010). Redding and Venables (2004) interpret the MR terms as 'supply access' and 'market access' indexes.

[12]Baier and Bergstrand (2009) propose a fourth approach, Bonus Vetus OLS (BvOLS), which uses a first-order log-linear Taylor-series expansion of the MR terms. We refer the reader to Egger and Pfaffermayr (2022) for a recent discussion and implementation of BvOLS with fixed effects, as we do in later specifications.

[13]Olivero and Yotov (2012) offer additional motivation for the treatment of the MRs and the use of exporter-time and importer-time fixed effects in panel gravity settings.

model, the exporter-time and importer-time fixed effects ($\phi_{i,t}$ and $\psi_{j,t}$, respectively) would absorb not only the MRs but the size control variables too:

$$\ln\_X_{ij,t} = \beta_1\, RTA_{ij,t} + \beta_2 \ln\_DIST_{ij} + \beta_3\, CLNY_{ij} + \beta_4\, CNTG_{ij} + \beta_5\, WTO_{ij,t} + \phi_{i,t} + \psi_{j,t} + \epsilon_{ij,t}. \quad (4)$$

Comparison between the estimates from specification (4), which appear in column (6) of Table 1, and the corresponding results from column (5) reveals a significant drop (in percentage terms) in the RTA estimate. Moreover, we also see very large changes in the estimates of some of the other covariates. For example, consistent with the main result from Anderson and van Wincoop (2003) regarding the impact of the border between Canada and the United States, we find that once we account fully for the MR terms, the estimate of the effects of contiguous borders decreases in half. Another notable change is that the WTO estimate in column (6) is about three times smaller than the corresponding index from column (5).

Endogeneity has been a major obstacle in gravity models. While the sources of the problem are very clear, e.g., reverse causality and/or omitted factors that simultaneously affect trade and the probability to sign RTAs, the profession has struggled to address this issue using standard techniques (e.g., instrumental variable (IV) treatment) due to the fact that very often what determines the probability to sign a trade agreement also affects the volume of trade flows (Egger et al., 2008; Egger et al., 2011; Helpman et al., 2008; Jochmans & Verardi, 2022). Some authors have proposed matching-type approaches (e.g., Baier & Bergstrand, 2009; Egger et al., 2008; Egger & Tarlea, 2020).

Due to its intuitive appeal and easy implementation, the leading method to handle endogeneity of RTAs is due to Baier and Bergstrand (2007), who, consistent with the approach to control unobserved time-invariant heterogeneity with panel data by Wooldridge (2010), propose the use of pair fixed effects (γ_{ij}), which will control for most of the unobserved correlation between the endogenous RTAs and the error term in gravity models.[14] Applied to our setting, this leads to the following econometric specification:

$$\ln\_X_{ij,t} = \beta_1\, RTA_{ij,t} + \beta_2\, WTO_{ij,t} + \phi_{i,t} + \psi_{j,t} + \gamma_{ij} + \epsilon_{ij,t}. \quad (5)$$

As demonstrated by Egger and Nigai (2015) and Agnosteva et al. (2019), the ‘standard gravity variables’ (e.g., distance, contiguity, etc.) do well in predicting relative trade costs, however, they fail to capture the level of bilateral trade costs (e.g., they under-predict the bilateral trade costs for the poor countries and over-predict them for the more developed countries). Thus, an additional benefit of the use of pair fixed effects is that they will match the bilateral trade costs much better. Consistent with the results from Baier and Bergstrand (2007), the estimates in column (7) of Table 1 reveal that the RTA estimate is larger. Another interesting result, which is also consistent with some prominent findings from the existing literature (e.g. Rose, 2004) is that the estimate of the impact of WTO becomes negligible in magnitude and not statistically significant.

In addition to mitigating endogeneity concerns, another possible source for the differences between the estimates with and without the pair fixed effects, is that the latter will fully absorb

[14]Egger (2002) is an early contribution that advocated the use of panel data for gravity estimation and the inclusion of pair fixed effects. Concerned with time-invariant regressors and the feasibility to explore trade potentials in- and out-of-sample, he compares estimates from fixed effects, between, random effects, and Hausman and Taylor (1981) models, the latter two also allowing for first-order autocorrelation.

and control for the effects of all regional trade agreements that entered into force prior to the period of investigation (except for, possibly, some phasing-in effects). Thus, the RTA estimates we obtain in the presence of pair fixed effects in column (7) of Table 1 are identified by agreements that entered into force during the period of investigation.

Our next specification follows Baier et al. (2019), who advocate the use of directional pair fixed effects ($\overrightarrow{\gamma_{ij}}$) in part because bilateral trade costs may be asymmetric. For example, even if the distance between two countries is perfectly symmetric, the cost of transporting a good from one to the other and vice versa may be asymmetric and the effects of some trade policies could be asymmetric, which, for consistency, implies modelling asymmetric time-invariant trade costs. With this adjustment, our econometric model becomes:

$$\ln\_X_{ij,t} = \beta_1\, RTA_{ij,t} + \beta_2\, WTO_{ij,t} + \phi_{i,t} + \psi_{j,t} + \overrightarrow{\gamma_{ij}} + \epsilon_{ij,t}. \tag{6}$$

Comparison between the estimates from specification (6), which appears in column (8) of Table 1, and the corresponding results from column (7) reveals they are virtually identical. Thus, it seems there is no gain from allowing for asymmetric time-invariant trade costs. However, we still recommend the use of directional pair fixed effects. As we will demonstrate in the next section, this matters a lot for obtaining directional estimates of the effects of trade agreements (e.g., the impact of CETA on Canada's exports to Germany vs. the impact of CETA on Germany's exports to Canada). Moreover, the current state of technology (e.g., Correia, 2016; Correia et al., 2020) is such that the introduction of fixed effects in gravity models has minimal impact on convergence and computational time.[15]

One important empirical contributios to the literature is the introduction of the PPML estimator by Santos Silva and Tenreyro (2006). Due to its ability to handle heteroskedasticity in the trade data, which renders OLS estimates inconsistent, and to estimate gravity in multiplicative form, which does not lead to dropping the zero trade flows from gravity estimations, PPML established itself as the leading gravity estimator (Santos Silva & Tenreyro, 2022).[16] Accordingly, our next specification employs PPML to become:

$$X_{ij,t} = \exp\left[\beta_1\, RTA_{ij,t} + \beta_2\, WTO_{ij,t} + \phi_{i,t} + \psi_{j,t} + \overrightarrow{\gamma_{ij}}\right] + \epsilon_{ij,t}. \tag{7}$$

The main result from column (9) of Table 1 is that the PPML estimate of the effects of RTAs is significantly smaller than its OLS counterpart from column (8). Consistent with the main argument from Santos Silva and Tenreyro (2006), this suggests the OLS estimates may indeed be biased. We further investigate the source of the bias observed. Specifically, given the relatively small number of zeroes in our sample (about 2%), we do not expect the difference between the OLS and the PPML estimates is driven by the zeroes. Nevertheless, we obtained PPML estimates based only on positive trade flows. As expected, the resulting RTA estimate (0.1385, std.err. 0.0228) is virtually identical to the estimate from column (9) (0.1384, std.err. 0.0228). Hence, as demonstrated in Larch et al. (2018) and discussed in Head and Mayer (2014), PPML and OLS weight observations differently. Specifically, PPML gives more weight to larger trade flow observations.

[15]Following Bergstrand et al. (2015), we also experimented by interacting the pair fixed effects in our model with linear time trends. This had no significant impact on our results.

[16]Head and Mayer (2014) advocate the use of the Gamma PML estimator. However, unlike PPML, GPML is subject to the incidental parameter problem (IPP). We offer further discussion on this front below.

Even though all theoretical gravity models (e.g., Anderson, 1979; Anderson & van Wincoop, 2003; Arkolakis et al., 2012; Eaton & Kortum, 2002) and corresponding CGE analysis include domestic trade flows, the benefits of using theory-consistent domestic trade flows in estimating gravity models were recognised only more recently. For example, Yotov (2012) uses domestic trade flows in a gravity estimation to resolve the distance puzzle in trade while Dai et al. (2014) is the first structural gravity model that estimates the effects of RTAs with domestic, in addition to international, trade flows. Some of the benefits of using domestic trade flows for gravity estimations include the ability to estimate the effects of non-discriminatory trade policies (Heid et al., 2021) and country-specific trade policies (Beverelli et al., 2018). Yotov (2022) describes 11 benefits of using domestic trade flows in gravity regressions.

Introducing domestic trade flows does not lead to significant changes in the presentation of our econometric model. The only difference is that the dependent variable now includes domestic trade and we have assigned zeros to the corresponding observations in the RTA and WTO covariates. The set of fixed effects is identical, with the set of pair fixed effects now also including time-invariant fixed effects for domestic trade (γ_{ii}). To capture these differences explicitly, we add the notation '$\forall_{i,j}$' to our estimating model:

$$X_{ij,t} = \exp\left[\beta_1\, RTA_{ij,t} + \beta_2\, WTO_{ij,t} + \phi_{i,t} + \psi_{j,t} + \overrightarrow{\gamma_{ij}}\right] + \epsilon_{ij,t}, \forall_{i,j}. \tag{8}$$

The estimates in column (10) of Table 1 are substantially different from those in column (9). Consistent with the results from Dai et al. (2014), the estimate of the effects of RTAs is four times larger. The intuition for this is that the new specification explicitly allows for diversion from domestic trade flows. The large increase in the RTA estimate implies RTAs divert significant amounts of trade from domestic toward international sales. Another important result from column (10) is that the effect of GATT/WTO membership becomes positive, large, and statistically significant. This is consistent with Larch et al. (2019), who demonstrate that the 'puzzle of missing WTO effects' is resolved once gravity is estimated with (theory-consistent) domestic trade flows.

Bergstrand et al. (2015) argue that the estimates of RTAs may capture common globalisation trends and, therefore, be biased upward. To address this, they propose an adjustment to the gravity equation from the previous specification, which explicitly accounts for globalisation effects. Following Bergstrand et al. (2015), the simple modification we make to Equation (8) is that, in addition to domestic trade flows in the dependent variable, we introduce as covariates a series of time-varying border variables, $\sum_t BRDR_{ij,t}$, which take a value of one for international trade and zero for internal trade for each year in our sample. Our estimating model becomes:

$$X_{ij,t} = \exp\left[\beta_1\, RTA_{ij,t} + \beta_2\, WTO_{ij,t} + \sum_t \beta_t\, BRDR_{ij,t} + \overrightarrow{\gamma_{ij}}\right] + \epsilon_{ij,t}, \forall_{i,j}. \tag{9}$$

Before we discuss our findings, we note that, due to perfect collinearity with the pair fixed effects, we cannot estimate the coefficients of all border variables and need to drop one. We select the border variable for 1980, which is the first year in our sample.

Estimates from the specification with the globalisation effects appear in column (11) of Table 1. The estimates of the globalisation/border dummy variables are positive and increasing over time.[17] Keeping in mind that these estimates are relative to the border effect in 1980, the results reveal that the gravity equation captures the effects of globalisation quite well, which is consistent with Bergstrand et al. (2015) but in contrast to Bhavnani et al. (2002), who argue that gravity models cannot capture the effects of globalisation and dub this 'the missing globalisation puzzle'. Another result is that the estimate of the effects of RTAs is significantly smaller. This is consistent with the main argument from Bergstrand et al. (2015) that the RTA effects from column (10) have indeed captured some common globalisation effects.

Motivated by Egger and Tarlea (2015), Pfaffermayr (2019), and Pfaffermayr (2022), who demonstrate that multi-level clustering makes a big difference, the next two specifications allow for different degrees of correlations among observations. The gravity model that we estimate is identical to specification (9), and the only difference is that in column (12) we cluster the standard errors by country pair (i.e., $Cov[\epsilon_{ij,t}, \epsilon_{ij,d}] \neq 0$, for all t, d, and zero else), while the results in column (13) are obtained with 3-way clustering (by exporter, importer, and year, i.e., $Cov[\epsilon_{ij,t}, \epsilon_{il,d}] \neq 0$, $Cov[\epsilon_{ij,t}, \epsilon_{kj,d}] \neq 0$, $Cov[\epsilon_{ij,t}, \epsilon_{kl,t}] \neq 0$, and zero else). As expected, alternative ways of clustering do not affect our point estimates. However, we see that the standard errors increase when we cluster by pair, and they further increase when we cluster three-way. Thus, our results suggest intra-cluster correlations. Importantly, the RTA effects remain statistically significant. Consistent with most of the existing gravity literature, for the rest of the analysis, we will cluster the standard errors by pair.

The IPP could be an issue for gravity estimations with multi-way fixed effects. Fernández-Val and Weidner (2016) demonstrate that IPP is not an issue for gravity models with two-way fixed effects (i.e., the exporter-time and importer-time fixed effects that are used to control for the MRs). However, Weidner and Zylkin (2021) show that, while still consistent, IPP leads to an asymptotic bias that affects the validity of inferences in PPML specifications with three-way fixed effects, e.g., such as ours, and propose an adjustment based on Taylor expansions to correct for such biases.[18] We follow the procedure of Weidner and Zylkin (2021) in column (14) of Table 1, where we see that the standard error is between the corresponding estimates with the pair vs. 3-way clustering.

Consistent with the critique of Cheng and Wall (2005, p. 52, fn. 8) that '[f]ixed-effects estimation is sometimes criticised when applied to data pooled over consecutive years on the grounds that dependent and independent variables cannot fully adjust in a single year's time', many gravity papers in the existing literature use interval data instead of data for consecutive years. For example, Baier and Bergstrand (2007) use five-year intervals, while Olivero and Yotov (2012) experiment with 3-, 4- and 5-year intervals. Recently, Egger et al. (2022) question this practice on the basis that its implementation requires dropping a lot of data (e.g., using 5-year intervals may lead to dropping up to 80% of the raw data) and that the selection of the intervals is somewhat random. Instead, Egger et al. (2022) argue that gravity estimations with 3-way fixed effects should rely on data for consecutive years. This eliminates the need to randomly drop data and allows for better capturing of the evolution of RTA effects over time. To capture these differences explicitly, we add the notation '$\forall_{i,j,t}$' to our estimating model:

$$X_{ij,t} = \exp\left[\beta_1\, RTA_{ij,t} + \beta_2\, WTO_{ij,t} + \sum_t \beta_t\, BRDR_{ij,t} + \phi_{i,t} + \psi_{j,t} + \overrightarrow{\gamma_{ij}}\right] + \epsilon_{ij,t}, \forall_{i,j,t}. \tag{10}$$

[17]For brevity, we decided not to report the border estimates in a table or a figure. Instead, we list them here. The seven estimates that we obtain (from 1985 to 2015, respectively) are 0.177 (std.err. 0.045), 0.415 (std.err. 0.043), 0.614 (std.err. 0.038), 0.962 (std.err. 0.038), 1.022 (std.err. 0.036), 1.106 (std.err. 0.036), and 1.177 (std.err. 0.041).

[18]Jochmans (2017) uses a quasi-differencing approach and Pfaffermayr (2021) implements jackknife and bootstrap confidence interval methods.

Following the recommendations of Egger et al. (2022), in columns (15) and (16) of Table 1, we experiment with 3-year lags and with consecutive year data, respectively. Other than that, the specification is the same as in column (14). Consistent with one of the main findings from Egger et al. (2022), we see that the RTA estimates that we obtain with different intervals and consecutive-year data are not very different from each other. Based on this, Egger et al. (2022) conclude that there are no reasons to (randomly) drop data. Moreover, and as expected, we see that the standard errors are a bit more precisely estimated when we use more data, again favouring the specification with consecutive years. Finally, as we demonstrate in the next section, the use of consecutive-year data may have significant implications for capturing the evolution of the RTA effects over time.

As visualised in Figure 2 our results send two main messages. First, some of the theory and econometric developments have had a significant impact on the RTA estimates, while others not so much. Second, the most recent contributions, many of which focused on the standard errors, seem to be converging in terms of RTA estimates.

We find these conclusions intuitive, yet potentially misleading. For example, even though Figure 2 implies convergence toward a single RTA effect, we are aware of several recent ongoing trends in the related literature that may influence the RTA estimates. Moreover, even if some of our RTA results are not very sensitive to the methods we employ, we saw that the estimates of some of the other gravity covariates in the same specifications could be quite sensitive. We reinforce this message in Figure 3, where we plot the evolution of the estimates of

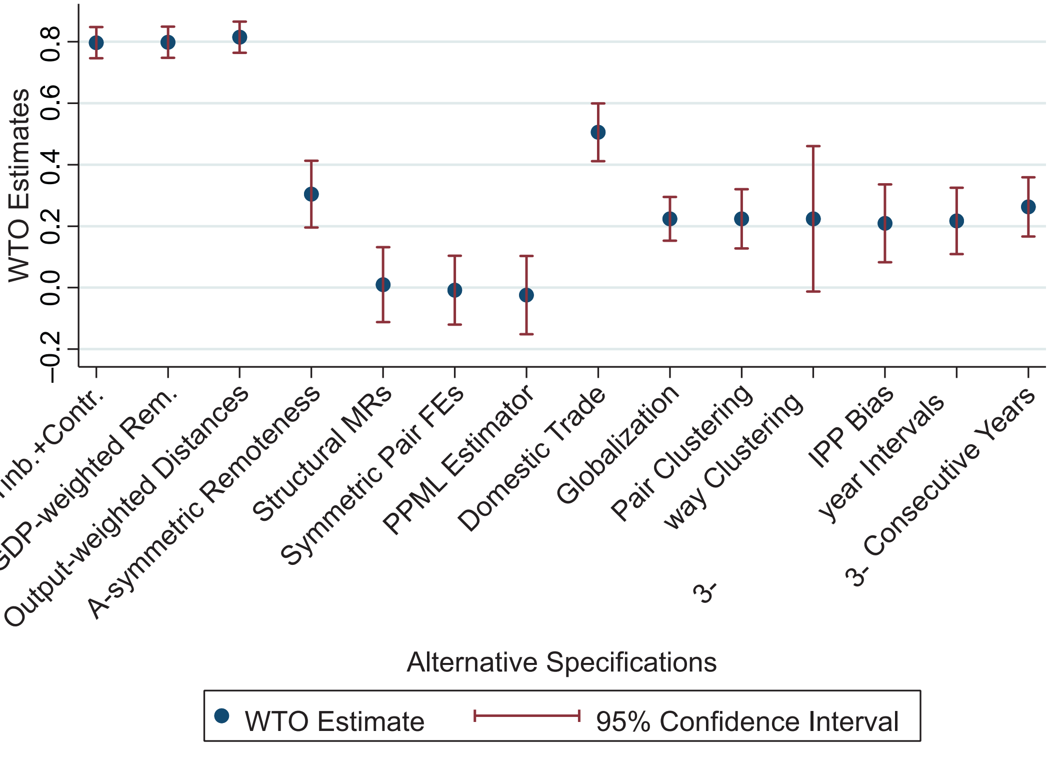


FIGURE 3 Sixty years of WTO estimates with gravity. This figure plots the estimates of the effects of WTO membership, along with the corresponding confidence intervals, which we obtained in Table 1. The *X*-axis of the figure lists the alternative specifications that were used. We refer the reader to Table 1 and to the discussion in the main text for further details. [Colour figure can be viewed at wileyonlinelibrary.com]

the WTO effects from the same specifications, and see they are more heterogeneous. Finally, we note that, even if not important for the average RTA effects, some of the methods that we described may have very significant implications for the heterogeneous RTA effects across various dimensions.

3 | ON THE HETEROGENEOUS EFFECTS OF TRADE AGREEMENTS

In this section, we explore the heterogeneous effects of RTAs across various dimensions and with different data. In each case, we rely on our most preferred specification from the previous section:

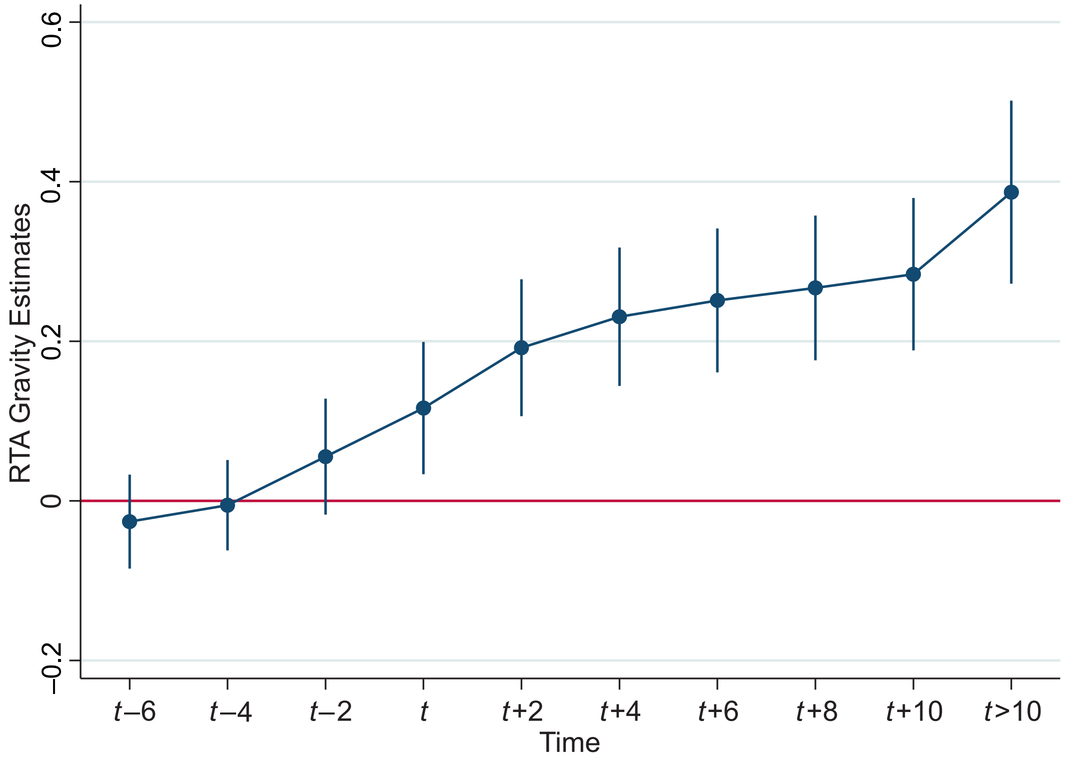
$$X_{ij,t} = \exp\left[\mathbf{RTA_{ij,t}}\,\alpha + \beta_8\, WTO_{ij,t} + \sum_{t=1981}^{2016} \beta_t\, BRDR_{ij,t} + \phi_{i,t} + \psi_{j,t} + \overrightarrow{\gamma_{ij}}\right] + \epsilon_{ij,t}, \forall_{i,j,t}. \quad (11)$$

The single difference between Equations (10) and (11) is that the new specification defines the RTA variables as part of a vector.

First, we investigate the evolution of RTA effects over time. Our results appear in Figure 4. We start with an analysis of the anticipation and phasing-in effects, which appear in the top panel of Figure 4. The motivation for a specification with ‘anticipation’ effects is that some firms may take note when an agreement is announced, and may start adjusting in anticipation of the new trade conditions. Moreover, it is possible that some trade costs between the potential RTA members start falling even before an agreement is signed. A possible motivation for the ‘phasing-in’ effects of RTAs is that RTAs reduce tariffs and NTMs stepwise, and the effects of reduced tariffs and NTMs may take time to materialise (see for example Baier & Bergstrand, 2007; Egger et al., 2022).

Against this backdrop, to capture the evolution of the effects of RTAs over time, we introduce a lead six-year before the agreement, taking a value of one six and five years before the agreement enters into force, and zero otherwise. In a similar way, we introduce leads of 4 years and 2 years before the agreement. The RTA variable stays equal to one in the year of entry into force and the following year. Then we add lags. The lag of 2 years captures the effect after two and 3 years, the lag of four the effect of the agreement four and 5 years after entering into force, and so on, until year 10, which captures years 10 and 11 after the entry into force, and the last one capturing the effect of the agreement thereafter. Note the RTA variables are set to zero in all other periods.

The point estimates that we obtain from this specification are plotted in the top panel of Figure 4. There are no statistically significant anticipation effects. Given that RTAs may potentially be endogenous and anticipation effects may harm the quantification of their effects, this suggests the specification chosen and specifically the included fixed effects take care of large parts of these concerns. In principle, it is possible to obtain statistically significant anticipation RTA estimates (Egger et al., 2022; Moser & Rose, 2014). Given that the decisions to sign RTAs and the process of negotiation usually take at least a couple of years, we would interpret positive 1- or 2-year lead estimates of the effects of RTA as an indicator of economic adjustment in response to the RTA, rather than a sign of unaccounted endogeneity. As noted in Egger et al. (2022), the implication of this is that econometric models should allow for anticipation effects of RTAs with data that are measured at a sufficiently fine granularity in the time dimension.


RTA Gravity Estimates
0.6
0.4
0.2
0
−0.2
t−6
t−4
t−2
t
t+2
t+4
t+6
t+8
t+10
t>10
Time


RTA Gravity Estimates
0.4
0.3
0.2
0.1
0
1980−1989
1990−1999
2000−2009
2010−2016
Time Interval

FIGURE 4 Regional trade agreement (RTA) effects over time. This figure plots estimates that trace the evolution of the effects of RTAs over time. These estimates are obtained from specification (11), after replacing the vector of RTAs ($\mathbf{RTA_{ij,t}}$) with variables that enable us to capture the anticipation and phasing-in effects of RTAs (in the top panel of the figure) and the heterogeneous RTA effects depending on the decade in which they were signed (in the bottom panel of the figure). See text for further details. [Colour figure can be viewed at wileyonlinelibrary.com]

Concerning the lags, we see that it takes some time until the RTAs reach their full potential. In the first years, it is increasing more steeply, then flatting a bit after 6 years. Such evolution is consistent with a gradual adjustment/learning process for trading firms. A large part of the overall increase in trade flows due to an RTA (around 75%) is achieved within 12 years after the agreement has entered into force. This is consistent with findings in the literature that the RTA effects have fully materialised after around 10–15 years (see for example Baier & Bergstrand, 2007; Baier & Standaert, 2022; Egger et al., 2022).

Next, in the bottom panel of Figure 4, we show estimates that allow for differential RTA effects per period, i.e., for the period 1980–1989, 1990–1999, 2000–2009, and 2010–2016. Interestingly we find a stronger (even though not statistically different) effect for the periods 1980–1989 compared to 1990–1999. While in the 1980s, the number of RTAs was still small (see Figure 1), the RTAs that were in place may have on average larger effects. A natural explanation is that earlier RTAs were concluded among natural trading partners (see Baier & Bergstrand, 2004).

The period 1990–1999 is the 'golden age' when the number of RTAs increased dramatically. However, as the bottom panel of Figure 4 suggests, the average effect of the RTA in this period on bilateral trade flows was relatively weak. In the 2000s, we see an increase in the magnitude of the RTA estimates, which is even more pronounced in the last period considered (2010–2016). The evolution of our estimates supports a story of 'learning' to sign more efficient and effective RTAs. The larger estimates we obtain for the post-2000 period are also consistent with the fact that, during recent years, the RTAs have become deeper, including more provisions and a more careful design. We will substantiate this hypothesis next when we investigate the effects of RTAs by type and depth.

Our first step is to distinguish between the effects of different types of RTAs. To this end, we follow the WTO classification and distinguish between Customs Unions (CUs), Free Trade Agreements (FTAs), Partial Scope Agreements (PSAs), Economic Integration Agreements (EIAs), CUs and EIAs (CU&EIAs), FTAs and EIAs (FTA&EIAs), and PSA and EIAs (PSA&EIAs). We obtain this information from the RTA database of Egger and Larch (2008). Results when splitting RTAs into these seven types are presented in the top panel of Figure 5, and we find the variation of our estimates across the different types of agreements to be intuitive for the most part.

Several findings are noteworthy. First CUs, which not only reduce duties and other restrictive regulations of commerce within CU members but also coordinate to a common external policy, have significantly stronger effects on average. Second, FTAs have a positive and statistically significant estimate, but the size of the implied effects is substantially smaller. Third, PSAs, which are agreements that cover only certain products, turn out to be comparably strong. A possible explanation could be that these agreements are designed with a very well-defined purpose. Note also that there are many more FTAs than there are PSAs, hence, the relatively small FTA effect may mask substantial heterogeneity. Last, we estimate an effect for EIAs that is statistically insignificant and economically negligible (even negative). Since the EIAs are agreements liberalising trade in services and our sample includes manufacturing (goods) trade only, this result should not be surprising, and it suggests that services trade liberalising does not directly increase goods trade flows.

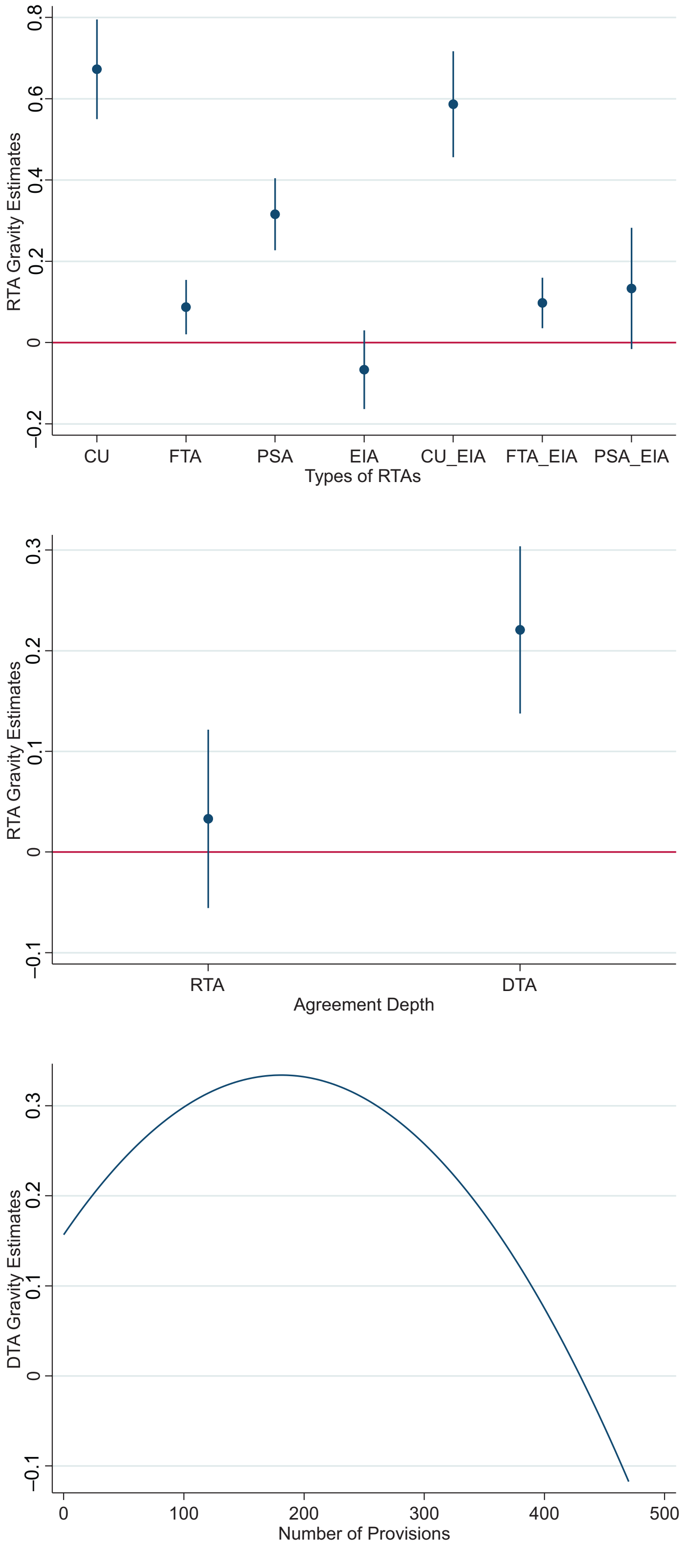

RTA Gravity Estimates
0.8
0.6
0.4
0.2
0
−0.2
CU
FTA
PSA
EIA
CU_EIA
FTA_EIA
PSA_EIA
Types of RTAs
RTA Gravity Estimates
0.3
0.2
0.1
0
−0.1
RTA
DTA
Agreement Depth
DTA Gravity Estimates
0.3
0.2
0.1
0
−0.1
0
100
200
300
400
500
Number of Provisions

FIGURE 5 Regional trade agreement (RTA) estimates by the type of agreement. This figure captures the heterogeneity of the effects of RTAs depending on their type and depth. All estimates are obtained from specification (11), after replacing the vector of RTAs ($\mathbf{RTA}_{ij,t}$) with variables that enable us to capture the type of RTAs (in the top panel of the figure), whether the RTAs were deep or not (in the middle panel of the figure), and depending on the depth of RTAs based on a continuous 'depth' index (in the bottom panel of the figure). See text for further details. [Colour figure can be viewed at wileyonlinelibrary.com]

We next utilise the World Bank's database on Deep Trade Agreements (DTA) (Hofmann et al., 2019; Mattoo et al., 2020),[19] and distinguish between Deep Trade Agreements (DTAs), where the depth index is greater than zero, and other RTAs. The results are presented in the middle panel of Figure 5. As expected, DTAs have a stronger effect than RTAs. Somewhat surprisingly, the estimate of the impact of RTAs in the middle panel, while positive, is not significant. Again, this may mask substantial heterogeneity.

To gain additional insights into whether deeper trade agreements have stronger effects on trade flows, we introduce alongside the two dummy variables the depth variable itself (i.e., a continuous count of the number of provisions within the DTAs in our sample), as well as its square term to allow for a non-linear effect. The bottom panel in Figure 5 plots the number of provisions against the predicted effect on trade for the DTAs. As can be seen, while for a lower number of provisions additional coverage of provisions increases the overall effect on trade, if agreements become too complex, the additional effects turn negative, and may even turn the overall effect negative for the most complex agreements.

We next investigate the effects per agreement, similarly to Baier et al. (2019). We include the dummy for each agreement separately. The results are visualised in the top panel of Figure 6. We sort the RTA estimates by size and plot for each agreement the coefficient (blue dots), as well as the lower 95% confidence interval bound (red line) and the upper 95% confidence bound (green line). As can be seen, effects range from more than minus four to over plus two, showing the huge heterogeneity across agreements. Many estimates are not significant. This may be because there are 390 agreements where we can estimate an agreement-specific effect, and many of them are only between two countries, and there are a lot of country pairs that are members not only of one but more than one agreement. This overlap makes it hard to identify the effects of single agreements.

As demonstrated by Baier et al. (2019), the impact of RTAs varies not only across agreements but also across pairs within agreements and even within pairs depending on the direction of trade flows. We allow for such variations in our next experiments. The estimates in Figure 7 allow for heterogeneous effects across pairs within an RTA. Similarly, as for the single agreement estimates in the top panel in Figure 6, we see huge heterogeneity. Note that we obtain 1316 point estimates in this case. From those about 40% (509) are positive, and 16% (228) are positive and statistically significant at the 5% level. We also obtain 30% (399) negative and statistically significant estimates.[20]

To visualise the heterogeneity across pairs, we follow Baier et al. (2017) and plot in the bottom panel of Figure 7 the single agreement estimates against the pair-wise estimates. As a pair may

[19]Available for download at https://datatopics.worldbank.org/dta/table.html. Specifically, we use all mapped 'PTAs' and the depth variable provided.

[20]Two possible explanations for the negative estimates could be (i) the smaller number of observations used to identify the pair-specific effects, thus making these estimates vulnerable to the influence of outliers, and (ii) the fact that some pairs are involved in more than one RTA, in which case some of the estimates in Figure 7 should be interpreted as deviations.

have more than one agreement, we kept the one with the larger estimate to produce the graph. Hence, not all agreements are visualised. Furthermore, restricted by the number of different colours/markers, all agreements with point estimates between −0.13 and 0.13 appear in blue colour, while the other agreements are highlighted in different colours.

As can be seen from Figure 7, heterogeneity within agreements (indicated by the same colour) is larger, ranging from about −5 to 5, than the estimates across the visualised agreements, which range from −1 to 2. This clearly shows that even when an RTA has positive effects on average, some members may see substantially lower trade effects than others. For example, we obtain the largest negative estimate of −0.895 for the EFTA-Central America (Costa Rica and Panama) agreement (red pluses on the far left), while the large positive effects for many members of 1.689

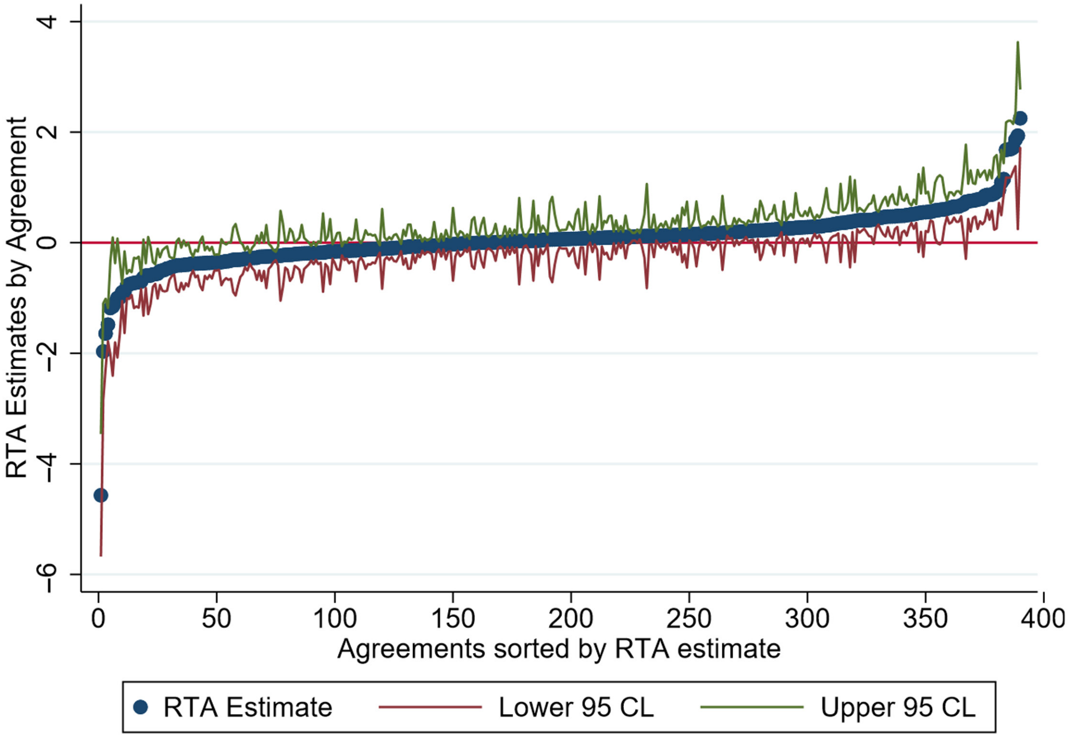
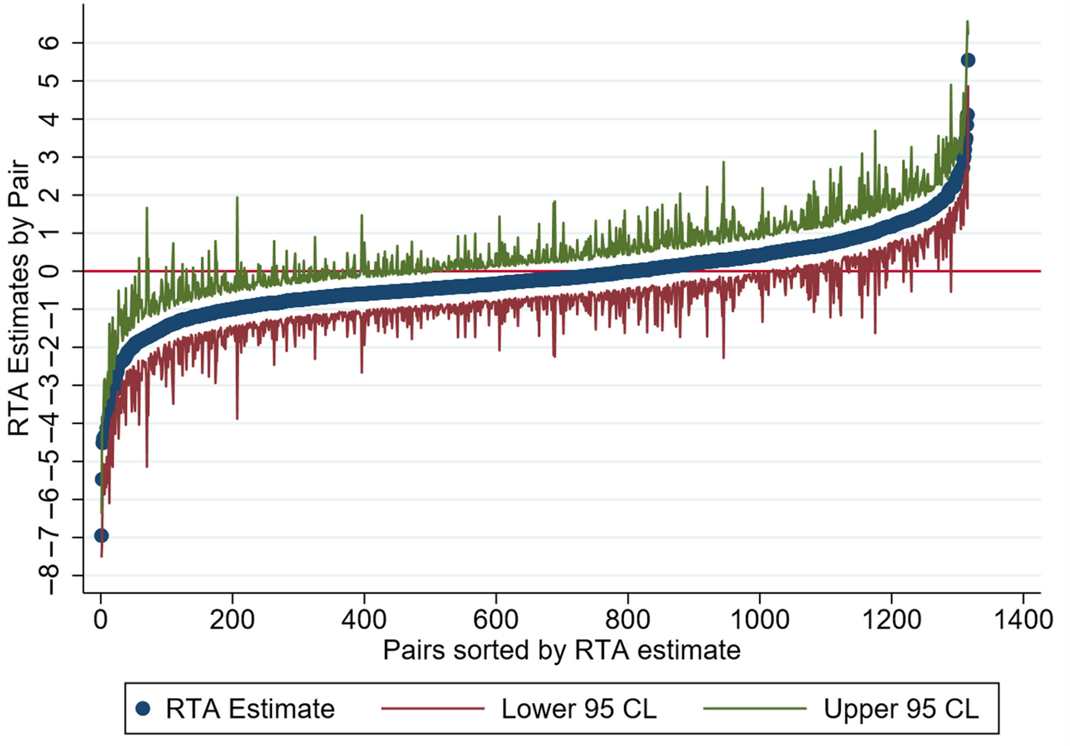
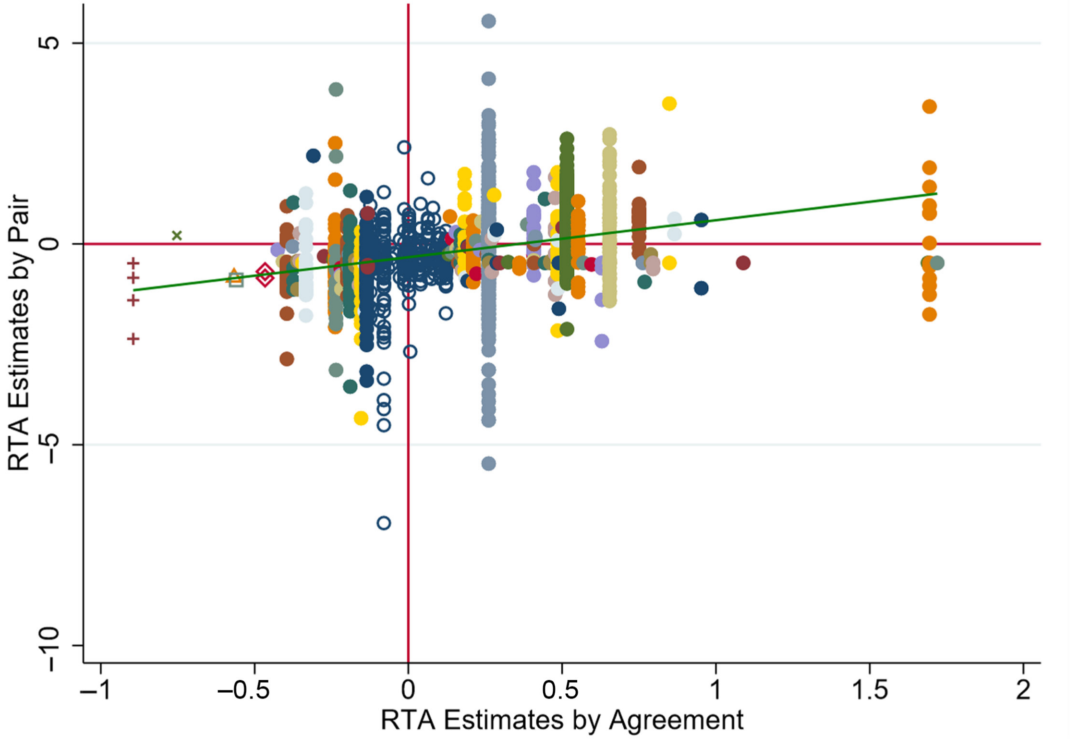
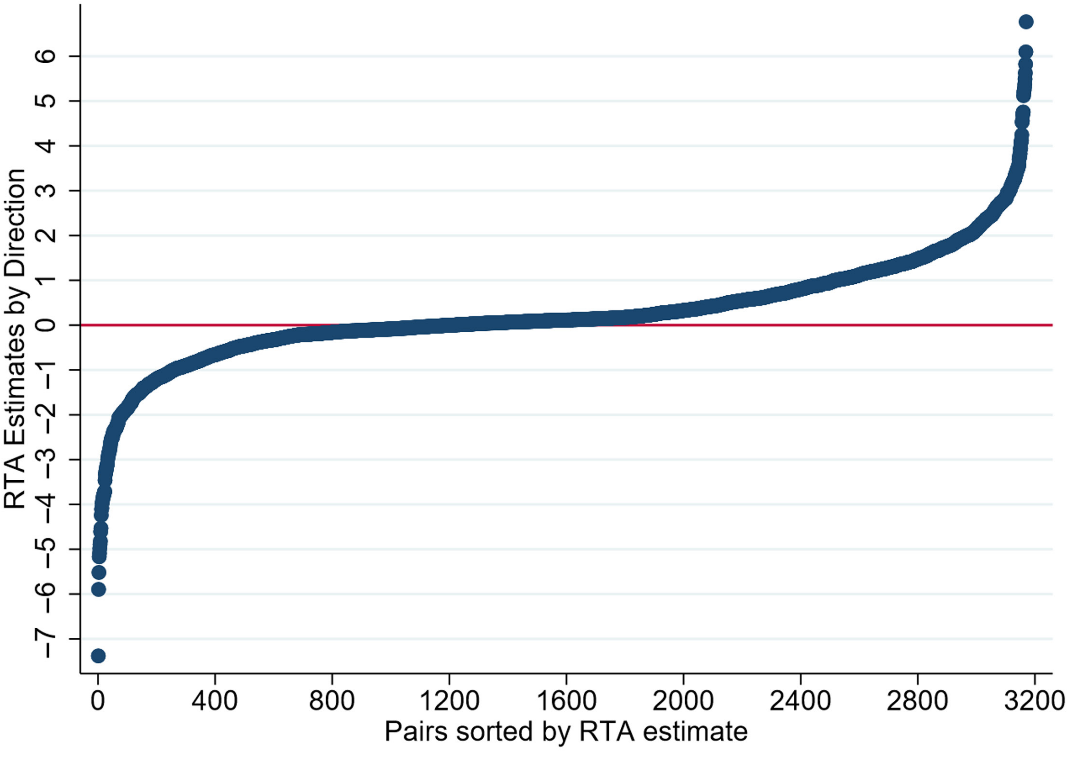


FIGURE 6 Regional trade agreement (RTA) estimates by agreement. This figure plots estimates that capture the heterogeneous effects of RTAs by agreement. These estimates are obtained from specification (11), after replacing the vector of RTAs ($\mathbf{RTA}_{ij,t}$) with indicator variables for each of the agreements in our sample. See text for further details. [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 7 Regional trade agreement (RTA) estimates by pair with agreement. This figure captures the heterogeneity of the effects of RTAs across pairs within an agreement. All estimates are obtained from specification (11), after replacing the vector of RTAs ($\mathbf{RTA}_{ij,t}$) with indicator variables that allow for separate estimates of the effects of RTAs for each pair of countries that signed them. The top panel of the figure orders our pair-specific estimates based on their magnitude and plots the corresponding confidence intervals. The bottom panel captures the heterogeneity within each agreement. Specifically, the different colours in the figure are used for different agreements, and the different dotes within each colour reflect the heterogeneous RTA effects across the pairs within the corresponding agreement. See text for further details. [Colour figure can be viewed at wileyonlinelibrary.com]


RTA Estimates by Pair
−8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6
0 200 400 600 800 1000 1200 1400
Pairs sorted by RTA estimate
RTA Estimate
Lower 95 CL
Upper 95 CL


RTA Estimates by Pair
−10 −5 0 5
−1 −0.5 0 0.5 1 1.5 2
RTA Estimates by Agreement


RTA Estimates by Direction
−7
−6
−5
−4
−3
−2
−1
0
1
2
3
4
5
6
0
400
800
1200
1600
2000
2400
2800
3200
Pairs sorted by RTA estimate

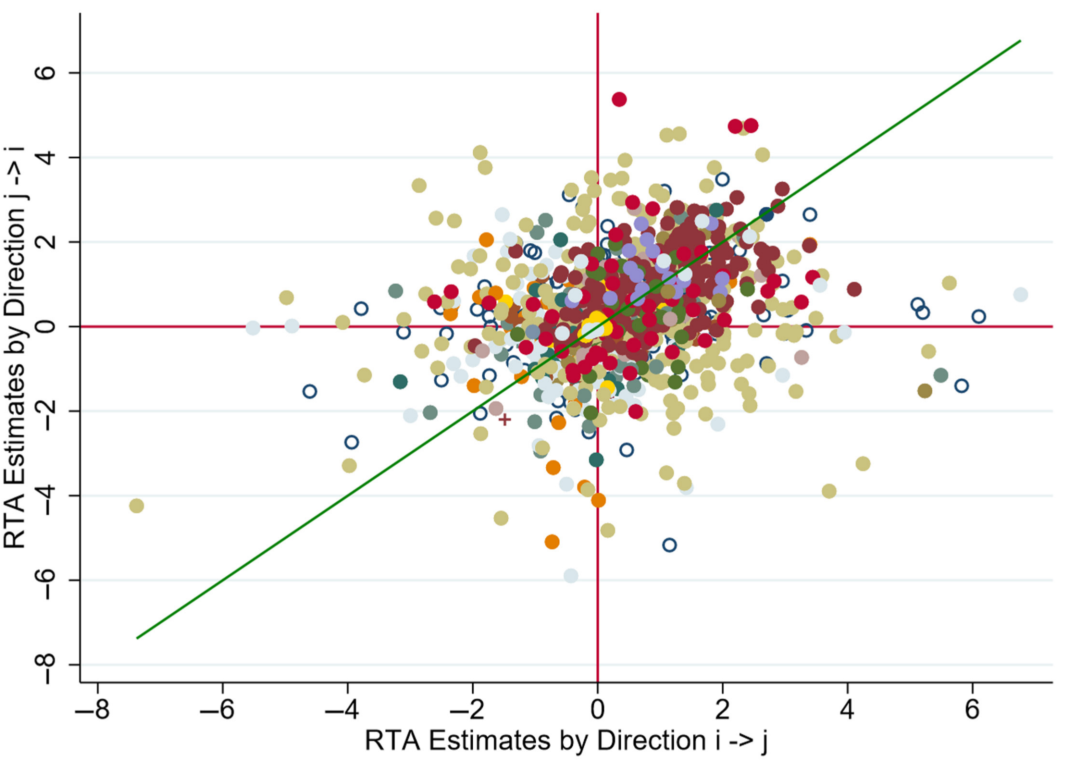

RTA Estimates by Direction j -> i
−8
−6
−4
−2
0
2
4
6
RTA Estimates by Direction i -> j

FIGURE 8 Directional regional trade agreement (RTA) estimates. This figure captures the heterogeneity of the effects of RTAs within pairs and within agreements but depending on the direction of trade flows. All estimates are obtained from specification (11), after replacing the vector of RTAs ($\mathbf{RTA_{ij,t}}$) with indicator variables that allow for separate estimates of the effects of RTAs for each pair of countries that signed them and also depending on the direction of trade within the pair. The top panel of the figure orders our directional-pair-specific estimates based on their magnitude and plots the corresponding confidence intervals. The bottom panel captures the heterogeneity within each pair and agreement. Specifically, the different colours in the figure are used for different agreements, and the different dotes within each colour reflect the heterogenous RTA effects across pairs and depending on the direction of trade within the corresponding agreement. If the effects of RTAs within each pair were symmetric, then all points in the bottom panel of the figure would have been on the 45° line. See the text for further details. [Colour figure can be viewed at wileyonlinelibrary.com]

is for the Protocol on Trade Negotiations (orange dots on the far right). The huge heterogeneity of members of the teal-coloured dotes with a point estimate of 0.261 is for the Global System of Trade Preferences among Developing Countries.

We can push heterogeneity even further by allowing for the effects of RTAs to vary not only by pair but also within pair and depending on the direction of trade (e.g., the impact of NAFTA on Canada's exports to Mexico vs. Mexico's exports to Canada). This is captured in Figure 8. In this case, we obtain nearly 3200 estimates.[21] We find a similar pattern as in Figure 7, but the share of positive estimates is larger (1963 positive estimates, i.e., about 60%). In the bottom panel of Figure 8 we plot the obtained estimates by direction, i.e., trade from country i to country j against trade from country j to country i. If trade would be balanced, we should see the dots aligned at the 45-degree line (the green line). This is not the case, highlighting the huge asymmetric effect of RTAs and reinforcing the need for careful analysis of the heterogeneous effects of RTAs even within member pairs.

The final heterogeneity we explore is across sectors. To obtain sectoral estimates, we rely on data from the USITC's ITPD-E-R02,[22] which contains international and domestic trade for 170 industries covering agriculture, mining and energy, manufacturing, and services for over 200 countries from the 1980s onwards (for services from 2000s).[23] Our results are captured in Figure 9. The top panel presents the results for the four broad sectors, and the results for the 170 industries are reported in the bottom panel of Figure 9.

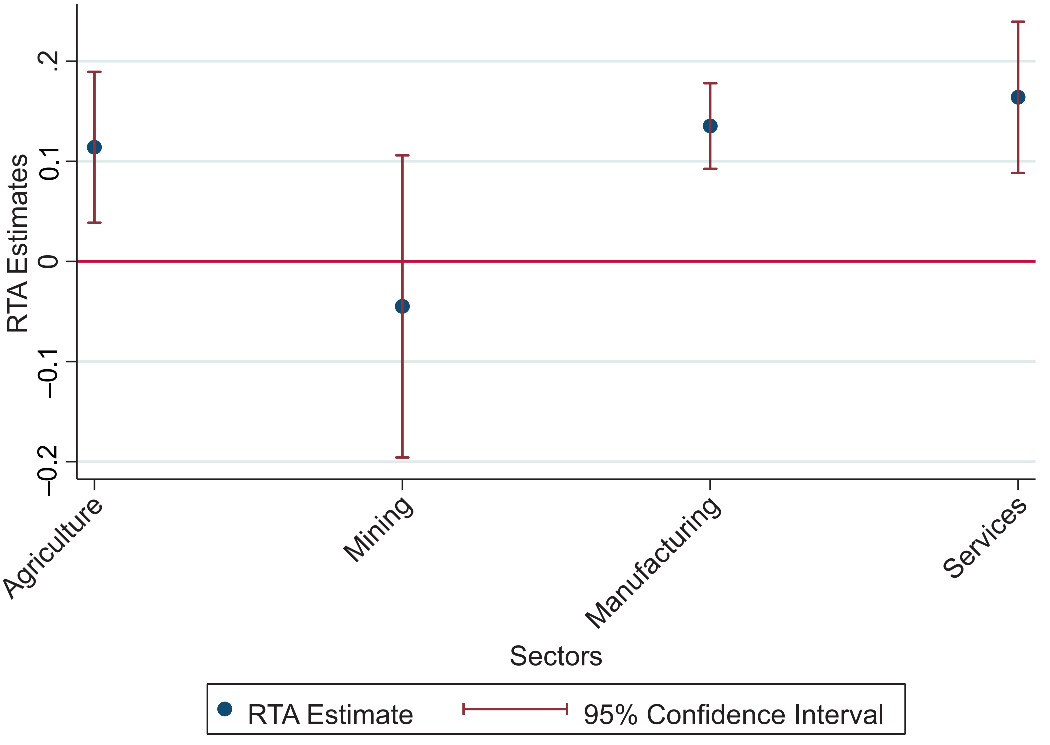
The estimates in the top panel reveal that, besides in mining, the RTAs in our sample have had positive and statistically significant effects in all other broad sectors. The magnitude of the overall RTA effects is similar across agriculture, manufacturing, and services. In mining, we find slight negative, but not statistically significant results. Also, the confidence interval is wide in mining. This suggests that RTAs seem not to be an important determinant of bilateral flows in mining and energy.

Finally, the bottom panel of Figure 9 visualises our estimates for each of the 170 industries in the ITPD-E-R02. For clarity, we do not plot the top and bottom 5% of the estimates. The main message from this figure is that the effects of RTAs vary significantly across sectors. In 60% of the industries (102 out of the 170) we obtain positive estimates, and in 39 industries the positive estimates are also statistically significant at the 5%-level. From the 68 negative estimates, 12 are

[21]Due to the many explanatory variables in this specification, we had to run the regression in a way that did not produce standard errors for the directional RTA estimates.

[22]Available for download at https://www.usitc.gov/data/gravity/index.htm.

[23]Borchert et al. (2021, 2022a, 2022b) provided descriptive statistics for ITPD-E and gravity estimates showing the suitability of this database for estimation.


RTA Estimates
.2
0.1
0
−0.1
−0.2
Agriculture
Mining
Manufacturing
Services
Sectors
RTA Estimate
95% Confidence Interval

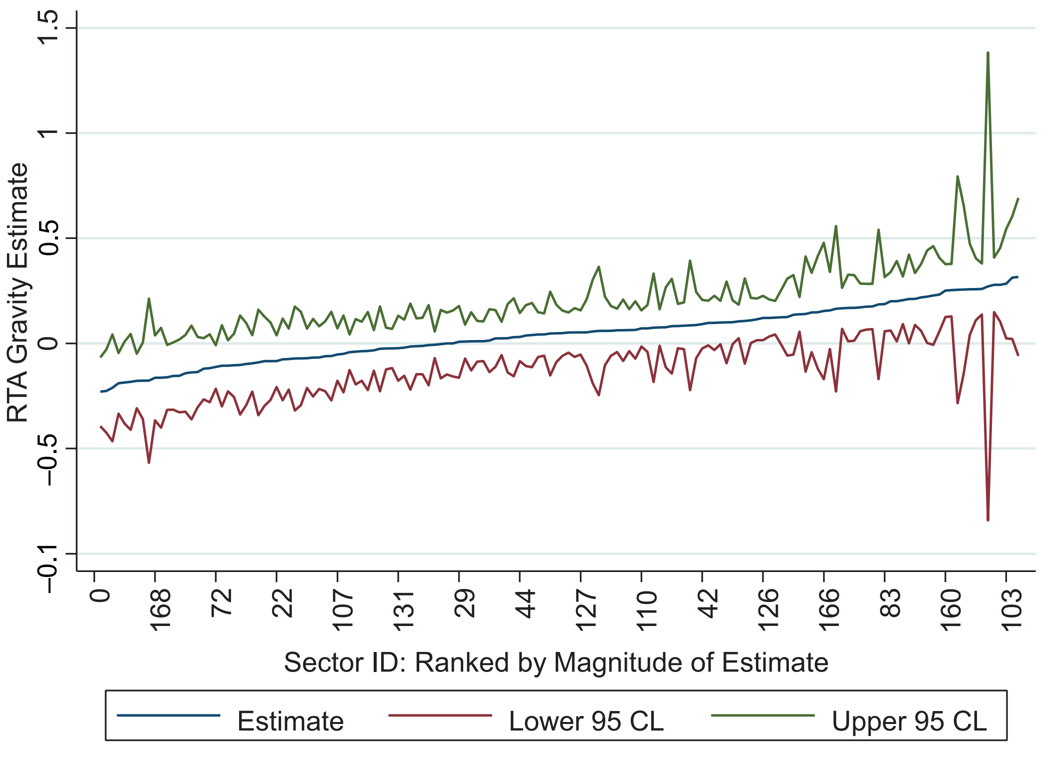

RTA Gravity Estimate
1.5
1
0.5
0
−0.5
−0.1
0
168
72
22
107
131
29
44
127
110
42
126
166
83
160
103
Sector ID: Ranked by Magnitude of Estimate
Estimate
Lower 95 CL
Upper 95 CL

FIGURE 9 Regional trade agreement (RTA) estimates across sectors and industries. This figure captures the heterogeneity of the effects of RTAs across sectors, based on data from the ITPD-E-R02. All estimates are obtained from specification (11), which is estimated at alternative levels of aggregation. Specifically, the estimates in the top panel of the figure are obtained for each of the four broad sectors (Agriculture, Mining and energy, Manufacturing, and Services) by pooling data on all individual industries within each of the broad sectors and using appropriate (exporter-industry-time, importer-industry-time, etc.) fixed effects. The bottom panel of the figure reports estimates (and their corresponding confidence intervals) for the individual industries from the ITPD-E-R02. For clarity, we do not plot the top and bottom 5% of the estimates. See text for further details. [Colour figure can be viewed at wileyonlinelibrary.com]

statistically different from zero at the 5%-level. The distribution of positive and negative estimates within broad sectors is relatively similar. Hence, the main variation is within broad sectors, not across broad sectors.

We draw several conclusions from this. We provide ample evidence that the effects of RTAs are very heterogeneous over time, across sectors, across individual agreements, depending on the type and depth of the RTAs, across country pairs within a multilateral trade agreement, and even within pairs depending on the direction of trade flows. Two important, and related, implications of this result are that (i) the average RTA estimates may hide significant heterogeneity across the pairs of countries in them, and (ii) that one has to keep this caveat in mind when RTA estimates are used for counterfactual analysis. Another important result from our analysis is that, even though most of the RTA estimates that we obtained were positive (and many of them statistically significant), we also obtained many negative and statistically significant estimates. This result also has two implications. From a methodological (and data) perspective, it is important for researchers to be careful and ensure that such estimates are sound. From a policy perspective, such negative estimates should play an important role in the negotiation and evaluation of RTAs.

4 | CONCLUDING REMARKS

Since Jan Tinbergen used a gravity equation to explain bilateral trade flows 60 years ago, researchers developed a thorough theoretical underpinning, made numerous refinements to the specification, tackled omitted variables biases and endogeneity, investigated various correlation structures of the error term, and developed databases that include domestic sales and more disaggregated data covering many sectors. We demonstrate these developments by respective gravity equation specifications to estimate the effects of RTAs. Our estimates show that the most recent developments lead to relatively robust estimates, suggesting a positive, significant average effect of RTAs on bilateral trade flows of about 22%.

This average effect of 22% masks huge heterogeneity. Hence, the second part of our analysis demonstrates the heterogeneity over time, depending on the type and depth of the RTAs, across individual agreements, across country pairs within agreements, and across sectors. While some recent contributions explored some of these dimensions, better data and new estimating developments will allow exploring these and possibly other dimensions of heterogeneity further.

There are several alternative relaxations that we did not explore, such as taking into account spatial/network effects (Morales et al., 2019), allowing for a non-CES structure (Fieler, 2011; Novy, 2013), taking the growth of trade into account (Anderson et al., 2020; Baier & Standaert, 2022; Eaton et al., 2016), allowing for heterogeneous trade cost elasticities (Chen

& Novy, 2022; Czarnowske & Stammann, 2022), or distinguishing the extensive and intensive margin of trade (Chaney, 2008; Eaton & Sotelo, 2013; Egger et al., 2011; Helpman et al., 2008). Furthermore, we did not utilise the theoretical underpinning to quantify the full, general equilibrium effects of RTAs. Recent developments that combine estimation and general equilibrium quantification are Adão et al. (2017), Allen et al. (2020), Adão et al. (2022), and Anderson (2022). We also look forward to new impactful contributions to the gravity and RTA literatures in the coming decades.

ACKNOWLEDGEMENTS

We are very grateful to the participants in the 'Gravity at Sixty' workshop at the Austrian Institute for Economic Research (WIFO, Vienna) for their positive and constructive feedback. We also thank the many colleagues who have made contributions to the related literature, and we apologise for possible omissions of relevant papers. All errors are our own. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

DATA AVAILABILITY STATEMENT

We confirm that we will provide all program (script) files and output files should the paper be accepted.

ORCID

Mario Larch https://orcid.org/0000-0001-9355-2004

REFERENCES

Adão, R., Costinot, A., & Donaldson, D. (2017). Nonparametric counterfactual predictions in neoclassical models of international trade. *American Economic Review*, *107*(3), 633–689.

Adão, R., Costinot, A., & Donaldson, D. (2022). *Putting quantitative models to the test: Evidence from the Trump tariffs*. Unpublished manuscript.

Agnosteva, D. E., Anderson, J. E., & Yotov, Y. V. (2019). Intra-national trade costs: Assaying regional frictions. *European Economic Review*, *112*, 32–50.

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

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *American Economic Review*, *69*(1), 106–116.

Anderson, J. E. (2022). *Non-parametric gravity*. Unpublished manuscript.

Anderson, J. E., Larch, M., & Yotov, Y. V. (2020). Transitional growth and trade with frictions: A structural estimation framework. *Economic Journal*, *130*(630), 1583–1607.

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

Anderson, J. E., & Yotov, Y. V. (2010). The changing incidence of geography. *American Economic Review*, *100*(5), 2157–2186.

Arkolakis, C., Costinot, A., & Rodríguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, *102*(1), 94–130.

Baier, S. L., & Bergstrand, J. H. (2004). The economic determinants of free trade agreements. *Journal of International Economics*, *64*(1), 29–63.

Baier, S. L., & Bergstrand, J. H. (2007). Do free trade agreements actually increase members' international trade? *Journal of International Economics*, *71*(1), 72–95.

Baier, S. L., & Bergstrand, J. H. (2009). Estimating the effects of free trade agreements on international trade flows using matching econometrics. *Journal of International Economics*, *77*(1), 63–76.

Baier, S. L., & Bergstrand, J. H. (2021). *NSF-Kellogg Institute data base on economic integration agreements*. Kellogg Institute. https://kellogg.nd.edu/nsf-kellogg-institute-data-base-economic-integration-agreements

Baier, S. L., & Standaert, S. (2022). *Gravity with trade shocks*. Unpublished manuscript.

Baier, S. L., Yotov, Y. V., & Zylkin, T. (2017, April 28). One size does not fit all: On the heterogeneous impact of free trade agreements. *VoxEU*, 116.

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

Baldwin, R. E., & Taglioni, D. (2006). *Gravity for dummies and dummies for gravity equations*. NBER Working Paper No. 12516.

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

Beverelli, C., Keck, A., Larch, M., & Yotov, Y. V. (2018). *Institutions, trade and development: A quantitative analysis*. School of Economics Working Paper Series 2018-3.

Bhavnani, R., Tamirisa, N. T., Subramanian, A., & Coe, D. T. (2002). *The missing globalization puzzle*. IMF Working Papers 2002/171.

Borchert, I., Larch, M., Shikher, S., & Yotov, Y. V. (2021). The international trade and production database for estimation (ITPD-E). *International Economics*, *166*, 140–166.

Borchert, I., Larch, M., Shikher, S., & Yotov, Y. V. (2022a). Disaggregated gravity: Benchmark estimates and stylized facts from a new database. *Review of International Economics*, *30*, 113–136.

Borchert, I., Larch, M., Shikher, S., & Yotov, Y. V. (2022b). *The international trade and production database for estimation (ITPD-E): An update*. Manuscript.

Campos, R. G., Vidal, E., Larch, M., Timini, J., & Yotov, Y. V. (2022). *Star Wars – Return of the GATT/WTO trade effect significance*. Unpublished manuscript.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review*, *98*(4), 1707–1721.

Chen, N., & Novy, D. (2022). Gravity and heterogeneous trade cost elasticities. *The Economic Journal*, *132*(644), 1349–1377.

Cheng, I.-H., & Wall, H. J. (2005). Controlling for heterogeneity in gravity models of trade and integration. *Federal Reserve Bank of St. Louis Review*, *87*(1), 49–63.

Correia, S. (2016). *Linear models with high-dimensional fixed effects: An efficient and feasible estimator*. Working paper.

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

Czarnowske, D., & Stammann, A. (2022). *Latent structures in directed network models: with an application to cross-section gravity models*. Unpublished manuscript.

Dai, M., Yotov, Y. V., & Zylkin, T. (2014). On the trade-diversion effects of free trade agreements. *Economics Letters*, *122*(2), 321–325.

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

Eaton, J., & Kortum, S. (2002). Technology, geography and trade. *Econometrica*, *70*(5), 1741–1779.

Eaton, J., Neiman, B., & Romalis, J. (2016). Trade and the global recession. *American Economic Review*, *106*(11), 3401–3438.

Eaton, J., & Sotelo, S. (2013). International trade: Linking micro and macro. In D. Acemoglu, M. Arellano, & E. Dekel (Eds.), *Advances in economics and econometrics - tenth world congress. Volume II, Applied economics* (pp. 329–370). Cambridge University Press.

Egger, H., Egger, P. H., & Greenaway, D. (2008). The trade structure effects of endogenous regional trade agreements. *Journal of International Economics*, *74*(2), 278–298.

Egger, P. H. (2002). An econometric view on the estimation of gravity models and the calculation of trade potentials. *The World Economy*, *25*(2), 297–312.

Egger, P. H., & Larch, M. (2008). Interdependent preferential trade agreement memberships: An empirical analysis. *Journal of International Economics*, *76*(2), 384–399.

Egger, P. H., Larch, M., & Yotov, Y. V. (2022). Gravity estimations with interval data: Revisiting the impact of free trade agreements. *Economica*, *89*(353), 44–61.

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

Egger, P. H., & Pfaffermayr, M. (2022). *Bonus-Vetus-OLS for gravity models and its approximation error*. Unpublished manuscript.

Egger, P. H., Staub, K. E., & Winkelmann, R. (2011). The trade effects of endogenous preferential trade agreements. *American Economic Journal: Economic Policy*, *3*(3), 113–143.

Egger, P. H., & Tarlea, F. (2015). Multi-way clustering estimation of standard errors in gravity models. *Economics Letters*, *134*, 144–147.

Egger, P. H., & Tarlea, F. (2020). Comparing apples to apples: Estimating consistent partial effects of preferential economic integration agreements. *Economica*, *88*(350), 456–473.

Feenstra, R. C. (2004). *Advanced international trade: Theory and evidence*. Princeton University Press.

Felbermayr, G., & Yotov, Y. V. (2021). From theory to policy with gravitas: A solution to the mystery of the excess trade balances. *European Economic Review*, *139*, 103875.

Fernández-Val, I., & Weidner, M. (2016). Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics*, *192*(1), 291–312.

Fieler, A. C. (2011). Nonhomotheticity and bilateral trade: Evidence and a quantitative explanation. *Econometrica*, *79*(4), 1069–1101.

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

Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica*, *49*(6), 1377–1398.

Head, K., & Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, & K. S. Rogoff (Eds.), *Chapter 3 in the Handbook of international economics* (Vol. *4*). Elsevier Ltd.

Heid, B., Larch, M., & Yotov, Y. V. (2021). Estimating the effects of nondiscriminatory trade policies within structural gravity models. *Canadian Journal of Economics*, *54*(1), 376–409.

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Trading partners and trading volumes. *Quarterly Journal of Economics*, *123*(2), 441–487.

Hofmann, C., Osnago, A., & Ruta, M. (2019). The content of preferential trade agreements. *World Trade Review*, *18*(3), 365–398.

Hummels, D. (2001). *Toward a geography of trade costs*. Unpublished manuscript.

Jochmans, K. (2017). Two-way models for gravity. *The Review of Economics and Statistics*, *99*(3), 478–485.

Jochmans, K., & Verardi, V. (2022). Instrumental-variable estimation of exponential-regression models with two-way fixed effects with an application to gravity equations. *Journal of Applied Econometrics*, *37*(6), 1121–1137.

Larch, M., Monteiro, J.-A., Piermartini, R., & Yotov, Y. (2019). *On the effects of GATT/WTO membership on trade: They are positive and large after all*. School of Economics Working Paper Series 2019-4, LeBow College of Business, Drexel University.

Larch, M., Wanner, J., Yotov, Y. V., & Zylkin, T. (2018). Currency unions and trade: A PPML Re-assessment with high-dimensional fixed effects. *Oxford Bulletin of Economics and Statistics*, *81*(3), 487–510.

Mattoo, A., Rocha, N., & Ruta, M. (Eds.). (2020). *Handbook of deep trade agreements*. World Bank. https://openknowledge.worldbank.org/handle/10986/34055

Morales, E., Sheu, G., & Zahler, A. (2019). Extended gravity. *The Review of Economic Studies*, *86*(6), 2668–2712.

Moser, C., & Rose, A. K. (2014). Who benefits from regional trade agreements? The view from the stock market. *European Economic Review*, *68*, 31–47.

Novy, D. (2013). International trade without CES: Estimating Translog gravity. *Journal of International Economics*, *89*(2), 271–282.

Olivero, M. P., & Yotov, Y. V. (2012). Dynamic gravity: Endogenous country size and asset accumulation. *Canadian Journal of Economics*, *45*(1), 64–92.

Pfaffermayr, M. (2019). Gravity models, PPML estimation and the bias of the robust standard errors. *Applied Economics Letters*, *26*(18), 1467–1471.

Pfaffermayr, M. (2021). Confidence intervals for the trade cost parameters of cross-section gravity models. *Economics Letters*, *201*, 109787.

Pfaffermayr, M. (2022). *Cross-section gravity models, PPML estimation and the bias correction of the two-way cluster-robust standard error*. Unpublished manuscript.

Ravenstein, E. G. (1885). The laws of migration: Part 1. *Journal of the Statistical Society of London*, *48*(2), 167–235.

Redding, S. J., & Venables, A. J. (2004). Economic geography and international inequality. *Journal of International Economics*, *62*(1), 53–82.

Rose, A. K. (2004). Do we really know that the WTO increases trade? *American Economic Review*, *94*(1), 98–114.

Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, *88*(4), 641–658.

Santos Silva, J. M. C., & Tenreyro, S. (2022). The log of gravity at 15. *Portuguese Economic Journal*, *21*(3), 423–437.

Tinbergen, J. (1962). *Shaping the world economy: Suggestions for an international economic policy*. The Twentieth Century Fund.

Wei, S.-J. (1996). *Intra-national versus international trade: How stubborn are nations in global integration?* NBER Working Papers No. 5531.

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

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). The MIT Press.

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

Yotov, Y. V. (2022). On the role of domestic trade flows for estimating the gravity model of trade. *Contemporary Economic Policy*, *40*, 526–540.

Yotov, Y. V., Piermartini, R., Monteiro, J.-A., & Larch, M. (2016). *An advanced guide to trade policy analysis: The structural gravity model*. UNCTAD and WTO.

How to cite this article: Larch, M., & Yotov, Y. V. (2024). Estimating the effects of trade agreements: Lessons from 60 years of methods and data. *The World Economy*, *47*, 1771–1799. https://doi.org/10.1111/twec.13569