CESifo WORKING PAPERS

10176
2022

December 2022

Estimating Gravity from the Short to the Long Run: A Simple Solution to the 'International Elasticity Puzzle'

James E. Anderson, Yoto V. Yotov

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10176*

Estimating Gravity from the Short to the Long Run: A Simple Solution to the 'International Elasticity Puzzle'

Abstract

We propose a simple and flexible reduced-form econometric approach to estimate gravity models in the short and the long run. The theoretical lens for interpreting our methods amends the canonical Lucas-Prescott adjustment formulation to allow for time-interval-varying depreciation-*cum*-adjustment. A time-varying trade elasticity in the structural gravity model is implied. Our methods explain the 'international elasticity puzzle,' the discrepancy between trade elasticity estimates from the trade literature and the international real business cycle literature. The same theory-motivated estimating equation applied to the same data generates a distribution of trade elasticity estimates that vary from 0.4 in the short run to 4.8 in the long run. The results offer support for some existing theories of dynamic adjustment in trade costs and imply that the long-run equilibrium in our sample is reached in about 16 to 17 years.

JEL-Codes: F130, F140, F160.

Keywords: short vs. long run, gravity estimation, trade elasticity.

James E. Anderson
Department of Economics
Boston College
USA - Chestnut Hill, MA 02467
james.anderson.1@bc.edu

Yoto V. Yotov
School of Economics
Drexel University
USA – Philadelphia, PA 19104
yotov@drexel.edu

December 18, 2022

1 Introduction

"International real business cycle models need low elasticities, in the range of 1 to 2, to match the quarterly fluctuations in trade balances and the terms of trade, but static applied general equilibrium models need high elasticities, between 10 and 15, to account for the growth in trade following trade liberalization." (Ruhl, 2008). Ruhl dubs this discrepancy between the trade and the IRBC literatures *'The International Elasticity Puzzle'*. More recently, Fontagné et al. (2018) use French firm-level data to revisit the puzzle, and they conclude that it was even more severe than originally thought.

We propose a simple and flexible reduced-form econometric approach to explain the international elasticity puzzle. The lens of the structural gravity model is extended to include bilateral marketing capacity adjustment with varying adjustment lag intervals. The extension offers a theoretical foundation that is consistent with a number of theories of the dynamic evolution of bilateral trade links, e.g., Arkolakis (2010), Crucini and Davis (2016), Chaney (2014), and Anderson and Yotov (2020).

Our dynamic adjustment model builds on the Lucas and Prescott (1971) formulation that reduces the roles of adjustment cost and depreciation to a single adjustment parameter of a log-linear difference equation. The Lucas-Prescott specification is extended to incorporate a combination of two features that result in a reduced-form adjustment/depreciation parameter that varies with respect to the time lag interval chosen to discretize the (almost) continuous real world process. The variation in lag length is explained by two elements. First, lags are due to the familiar macro literature combination of a recognition lag and an action lag (itself the sum of decision and implementation lags). The second element is firm heterogeneity in the lag lengths. The combination of the two features is that at any point in the transition from an initial shock to the long run, a time varying fraction of active firms have proceeded from recognition to action, based on their varying productivity in reading evidence, decision-making and implementation.[1] For any chosen (by the analyst) lag time interval, there is an

[1]This feature resembles the Calvo pricing mechanism in the macro literature.

associated Lucas-Prescott adjustment parameter.

The lag-interval dependence of the Lucas-Prescott adjustment parameter implies a lag-interval varying trade elasticity. In the structural gravity model behind our reduced-form estimation, the long run trade elasticity is a fixed parameter. The short run elasticity is a simple function of the long run trade elasticity and the lag-interval-varying adjustment parameter. The variation of the adjustment parameter across each discrete lag interval proxies the effect of the changing fraction of firms that have adjusted.

The resulting econometric gravity model delivers time-interval-varying estimates of the trade elasticity, one for each lag-interval specification. The profile of lag interval-pair fixed effects in the reduced-form estimations is interpreted in the lens of the structural model to reflect the variation in the proportion of firms that have adjusted. We apply the reduced form model to the era of globalization, 1988-2006. The structural model lens becomes plausible when treating globalization as a common shock that dominates other systematic shocks in its effect on bilateral trade capacity adjustment. To eliminate seasonal and other high frequency variation, gravity models are typically estimated on annual data, so the minimum lag is one year. The remaining lag interval choices are necessarily somewhat arbitrary to generate stylized reduced-form "facts".

Our method relies on two standard elements of the gravity modeling literature. First, the estimate of the coefficient of tariffs in the structural gravity model is only a function of the trade elasticity (Anderson and van Wincoop (2001), Fontagné et al. (2022)). Thus, the trade elasticity is obtained directly from the reduced-form estimate of the coefficient of tariffs. Second, we rely on and extend the standard fixed effects treatment of bilateral trade costs in empirical gravity equations (Baier and Bergstrand (2007), Egger and Nigai (2015), Agnosteva et al. (2019)) to allow for short vs. long run modeling of bilateral trade costs with interval-pair fixed effects.

The proposed methods have several advantages. They are flexible because the interval-pair fixed effects can be defined over any desired time span. They are also comprehensive because the fixed effects will control for and absorb all possible determinants of bilateral trade flows that are of the same pair-time dimension. Thus, in the spirit of Baier and Bergstrand (2007), our methods may further mitigate potential endogeneity concerns with bilateral trade policies. The lag-interval-pair fixed effects lead to better estimates of the effects of bilateral trade costs and their changes, a crucial advantage for counterfactual general equilibrium analysis. The methods are easy to implement because all they require is the creation of fixed effects. Finally, our reduced-form estimates are interpretable in the lens of the set of structural gravity models with closed forms.

The proposed methods also come with some caveats. First, the interval-pair fixed effects may not allow identification of the effects of some covariates that are of interest to the researcher. Second, setting the length of the intervals is rather arbitrary. Thus, the use of a variety of lag intervals and their interpretation needs discussion and defense that is appropriate to the product-country-time characteristics of the dataset. Our recommendation is to consider using interval-pair fixed effects, especially when the time span of the estimating sample is long, but to experiment with alternative intervals and to think carefully about the optimal interval length based on the dimensions of the data and given the specific policy evaluation in question.

The reduced-form gravity model is estimated with data on aggregate manufacturing trade and tariffs over the period 1989-2006. We present the results and quantitative implications from four time-interval-pair fixed effect specifications. The results explain the 'international elasticity puzzle' when viewed through the lens of the model. The alternative fixed-effect specification generate a distribution of time-varying trade elasticity estimates. These vary from 0.4 in the short run, through 1.4 and 1.9 in the medium run, to 4.8 in the long run. The structural model combines with some external parameters to recover estimates of the lag-interval-varying adjustment parameters. These are within the theoretical bounds and, consistent with expectations based on theory, increase gradually when we move from the short to the long run. A broader (and somewhat obvious) implication of our analysis is that comparisons between trade elasticities from different studies should be conditional on the length of the time span of the corresponding estimating samples.

Taking the structural model behind the estimated reduced form most literally eliminates reliance on external parameters in forming the adjustment parameter for each lag interval. The resulting profile of the adjustment parameter implies that as time increases, the proportion of firms that have adjusted rises, but at a decreasing rate. This finding is consistent with an S-shaped distribution of firm adjustment such as the logistic. It is also consistent with the theory of decreasing marginal trade costs of Arkolakis (2010). We demonstrate that our theory and econometric specification explain the evolution of the phasing-in effects of free trade agreements. Finally, we confirm our main conclusions with two alternative estimating samples. The estimates of the evolution of the adjustment parameter from these experiment imply that the long-run equilibrium in our sample is reached in about 16 to 17 years.

Our simple methods and corresponding analysis are related to several strands of the literature. From a theoretical perspective, our analysis is motivated by a series of papers, e.g., Arkolakis (2010), Head et al. (2010), Chaney (2014), Crucini and Davis (2016), and Anderson and Yotov (2020), which are concerned with the dynamic evolution of bilateral trade costs. Our main contribution to this literature is that we propose a simple and flexible econometric approach that captures the main idea from these papers and enables us to test some of their implications about the evolution of bilateral trade costs within the structural gravity model.

Most closely related to our work from this literature is Anderson and Yotov (2020), whose short run gravity model is the starting point for our analysis. Anderson and Yotov (2020) assume a common, time invariant adjustment parameter, which they borrow from outside, and, as a result, they can only estimate one value for short run elasticity and one value for the long run trade elasticities. In contrast, our theory allows for a time-varying adjustment parameter. This leads to a more flexible econometric treatment of the effect of bilateral trade cost changes that delivers the dynamic evolution of the adjustment parameter when moving from the short to the long run. The empirical analysis that we perform demonstrates that the evolution of the adjustment parameter over time is not linear, which has potentially important empirical and policy implications.

From a modeling perspective, we depart from the canonical formulation of Lucas and Prescott (1971) to capture the process of dynamic evolution of bilateral trade costs in our setting. More recently, the Lucas-Prescott adjustment process has been utilized in structural gravity settings by, e.g., Eaton et al. (2016) and Anderson et al. (2020). Our contribution generalizes the Lucas-Prescott adjustment parameter to allow it to vary over time-lag-intervals.

We also contribute to the broad literature on gravity estimation methods and applications. The most closely related paper from this literature is Baier and Bergstrand (2007), who advocate the use of pair fixed effects to mitigate endogeneity of bilateral policy variables in gravity estimations. In relation to Baier and Bergstrand (2007), we propose the use of interval-pair fixed fixed effects. This enables us to specify an econometric model that can capture the short vs. long run effects of trade policies. In addition, as mentioned earlier, the use of interval-pair fixed effects may further mitigate potential endogeneity concerns and it leads to better modeling of the vector of bilateral trade costs with implications for counterfactual analysis with new quantitative trade models.

The fourth strand of the literature that we contribute to is the one that aims at estimating trade elasticities, e.g., Egger et al. (2012), Simonovska and Waugh (2014), Soderbery (2015), Giri et al. (2021), and Fontagné et al. (2022). Most recent, and most closely related to our work is Fontagné et al. (2022), who use very disaggregated data on tariffs and the structural gravity model to estimate a large number of (long run) trade elasticities at the product level. Similar to Fontagné et al. (2022), we also rely on a version of the structural gravity model to obtain estimates of the trade elasticities. The key differences are that we offer a dynamic foundation for the evolution of the trade elasticity over time, and we obtain estimates of the trade elasticities when we move from short to the long run.

Finally, we contribute to the literature offering explanations of the international elasticity puzzle. Ruhl (2008) builds a model where the gap between the trade and the IRBC literatures is explained by entry of new exporters who do not act in response to temporary shocks but only in response to tariff decreases, which are permanent. Yilmazkuday (2019b) solves the puzzle by building a model of nested CES frameworks, which is consistent with the two literatures, and he shows theoretically and empirically that the trade elasticity is a weighted average of the macro elasticity. Most closely related to our work from this literature is Yilmazkuday (2019a), who uses a panel structural vector autoregressive model for the imports of a single country (the United States) to obtain estimates of the trade elasticities in the short and the long run. Similar to Yilmazkuday (2019a), we also obtain estimates of the trade elasticities from the short to the long run. The main difference is that we do so within a multi-country structural gravity setting and based on data on bilateral trade flows and tariffs.

The rest of the paper is organized as follows. Section 2 provides the theoretical foundation for our analysis. Motivated by theory and capitalizing on the developments in the empirical gravity literature, Section 3 sets up our econometric model. We offer a brief description of the data in Section 4. Section 5 presents our main findings and Section 6 offers concluding remarks. Finally, the Supplementary Appendix includes some derivations to aid the reader.

2 Theoretical Motivation

The objective of this section is to offer theoretical foundations for our empirical methods. To this end, we capitalize on the common idea from some recent papers (e.g., Arkolakis (2010), Chaney (2014), Crucini and Davis (2016), Anderson and Yotov (2020)) that the bilateral links between trading partners evolve dynamically over time. Dubbing these links 'bilateral capacity' and using $\lambda_{ij,t}$ to denote them, our departing point is the short-run gravity equation of Anderson and Yotov (2020):[2]

[2]For the convenience of the reader, we include the derivations of equation (1) in the Supplementary Appendix. For further details see Anderson and Yotov (2020).

$$X_{ij,t} = \frac{Y_{i,t} E_{j,t}}{Y_t} \left[\frac{T_{ij,t}}{\Pi_{i,t} \widetilde{P}_{j,t}} \right]^{(1-\sigma)\rho} \lambda_{ij,t}^{1-\rho}. \tag{1}$$

Here, following conventional notation, X_{ij} denotes nominal bilateral exports (at end user prices) from source i to destination j in year t. $Y_{i,t}$ is the value of gross output in source i, as share of world output $Y_t = \sum_i Y_{i,t}$, and $E_{j,t}$ is the total expenditure in destination j on goods from all origins. $T_{ij,t}$ is a vector of bilateral trade costs on shipments from i to j at time t, including *ad-valorem* tariffs. $\Pi_{i,t}$ and $\widetilde{P}_{j,t}$ are the equivalents of the multilateral resistance terms of Anderson and van Wincoop (2003), and σ is the elasticity of substitution between domestic and foreign varieties.

Two terms distinguish the short-run gravity model (1) from the traditional, long term gravity model (e.g., Eaton and Kortum (2002), Anderson and van Wincoop (2003), and Arkolakis et al. (2012)): (i) $\lambda_{ij,\tau}$ is a term that captures origin-destination-specific investment in bilateral links – 'bilateral capacity', which is consistent with the network link dynamics of Chaney (2014) or the 'marketing capital' notion of Head et al. (2010); (ii) $\rho \in (0, 1]$ is a micro-founded parameter, dubbed 'incidence elasticity', which is defined as a combination of the elasticity of substitution in demand and the elasticity of supply. It captures the proportion by which the short run trade elasticity is reduced from the long run trade elasticity, and when $\rho = 1$, equation (1) collapses to the traditional, long-run gravity equation.

The main objectives of this paper are to describe the transition between the short and the long run and to offer a simple method to implement this empirically in the lens of the structural gravity model. Inspired by Lucas and Prescott (1971), we assume that the 'bilateral capacity' adjustment process can be represented by the following Cobb-Douglas function:

$$\lambda_{ij,t} = (\lambda_{ij,t}^{*})^{\delta_{\Delta t}} \bar{\lambda}_{ij,t-\Delta t}^{1-\delta_{\Delta t}}, \qquad \delta_{\Delta t} \in (0, 1]. \tag{2}$$

Here, $\lambda_{ij,t}^{*}$ is the desired (long run) capacity and $\lambda_{ij,t}$ is the actual capacity at time t. The logic is a bit easier to see after dividing both sides of (2) by $\lambda_{ij,t-\Delta t}$:

$$\frac{\lambda_{ij,t}}{\bar{\lambda}_{ij,t-\Delta t}} = \left(\frac{\lambda^*_{ij,t}}{\bar{\lambda}_{ij,t-\Delta t}}\right)^{\delta_{\Delta_t}}.$$

Log-differentiating and rearranging terms,

$$\delta_{\Delta i} = \frac{d\ln\lambda_{ij,t} - d\ln\bar{\lambda}_{ij,t-\Delta t}}{d\ln\lambda^*_{ij,t} - d\ln\bar{\lambda}_{ij,t-\Delta t}},$$

the actual capacity rate of adjustment is a fraction of the percentage difference between current capacity and desired capacity.

The adjustment process specification is modified in the present paper in two related ways. First, the time lag interval Δt varies across specifications vs. the fixed $\Delta t = 1$ case of Lucas and Prescott. Second, plausibly, the adjustment parameter $\delta_{\Delta t}$ varies with the time lag interval chosen. The parameter combines depreciation with costs of adjustment (e.g., as in Eaton et al. (2016) and Anderson et al. (2020)). Its variability by length of the lag interval Δt reflects heterogeneity in the response rate of firms as well as possible non-linear effects of speed, as with the classic logistic lazy S applied to the accumulation of inventory.[3] Thus, for any lag interval, Δt imposed on the data by the econometrician, there is a potentially different adjustment parameter $\delta_{\Delta t}$. We anticipate that when we move from shorter to longer lag intervals, $\delta_{\Delta t}$ should increase and move closer to 1, which is its long-run theoretical bound.

We see four potentially important implications of the lag-interval-varying adjustment parameter. First is a specification with lag-interval-varying trade elasticities. This enables an elasticity transition from the short to the long run. Second, the structure may explain the non-linear phasing-in effects of free trade agreements from the recent gravity literature, e.g., Egger et al. (2022). Third, lag-interval-rising trade costs are consistent with the rising cost model of Arkolakis (2010), due to increasing difficulty of reaching new customers. Finally,

[3] Another motivation for this modeling choice is that longer lags are not as informative as shorter lags.

the time-lag variation of the trade elasticity has important implications for the quantification and evolution of general equilibrium effects in response to trade cost changes.

The other departure from the Lucas and Prescott (1971) adjustment process that we propose in specification (2) is implied by the first departure. The varying lag interval structure implies that the lagged bilateral capacity varies over still smaller lag intervals within the druation reduced to the point $t - \Delta t$. We denote the average capacity associated with the point $t - \Delta t$ as $\bar{\lambda}_{ij,t-\Delta t}$. This nested structure could potentially be exploited to more tightly inform the implications of the lag interval-pair fixed effects of the reduced-form econometric model.

The last step to the econometric model replaces the unobservable λs with observables. To do so, we adopt an ad-hoc structural approach. The agents know that long run efficient allocation implies that $\lambda_{ij}^* = s_{ij}^*$, where $s_{ij}^* = \frac{X_{ij}^*}{Y_i^*}$ is the efficient trade share for exporter i, and we assume that they use a boundedly rational specification that replaces s_{ij}^* with $s_{ij,t}$. Thus, replace the unobservable $\lambda_{ij}^* = s_{ij}^*$ with $s_{ij,t}$ and $\bar{\lambda}_{ij,t-\Delta t}$ with $\bar{s}_{ij,t-\Delta t}$ in (2), which becomes:

$$\lambda_{ij,\tau} = s_{ij,t}^{\delta_{\Delta t}} \bar{s}_{ij,\Delta t}^{1-\delta_{\Delta t}}.$$

Substitute the right-hand side of the preceding equation in gravity equation (1). Then use the definition of $s_{ij,t} = \frac{X_{ij,t}}{Y_{i,t}}$ and solve for $X_{ij,t}$. The result is:

$$X_{ij,t} = Y_{i,t} \left(\frac{E_{j,t}}{Y_t} \right)^{\frac{1}{1-\delta_{\Delta t}(1-\rho)}} \left[\frac{T_{ij,t}}{\Pi_{i,t} \widetilde{P}_{j,t}} \right]^{\frac{(1-\sigma)\rho}{1-\delta_{\Delta t}(1-\rho)}} \left(\bar{s}_{ij,t-\Delta t} \right)^{\frac{(1-\delta_{\Delta t})(1-\rho)}{1-\delta_{\Delta t}(1-\rho)}}. \tag{3}$$

Comparison between equations (1) and (3) reveals two important differences. First, the exponent of the bilateral trade cost term is $\frac{(1-\sigma)\rho}{1-\delta_{\Delta t}(1-\rho)}$, now lag interval-varying, i.e., the trade costs elasticity is now varying over time. Estimating the lag interval-varying parameter $\delta_{\Delta t}$ is a key objective in the empirical analysis. $\delta_{\Delta t}$ is the only lag interval-varying component of this term, so the sequence of lag-interval varying trade elasticities is driven by interval-variation of the adjustment parameter. The sequence traces the evolution of the trade elasticity from the short to the long run.

The second difference is that the last term in equation (3) is new, i.e., it does not appear in (1). Note that this term is directly liked to the power on the trade costs through $\delta_{\Delta t}$. Specifically, our theory implies that the time-varying trade costs elasticities that we will estimate would vary depending on the definition of the interval that is used to construct $\bar{s}_{ij,\Delta t}$. The practical importance of this term for estimation purposes is that it will guide our econometric model. In particular, it will motivate the use of flexibly defined interval-pair fixed effects. If our theory is correct, using pair fixed effects with different intervals will lead to different estimates of the trade elasticity. Theory also predicts the direction of the changes – longer intervals should correspond to larger values for the adjustment parameter.

3 Estimating gravity from the short to the long run

This section describes a reduced-form gravity specification with lag-interval-pair fixed effects based on the theory from the previous section. The reduced form extends several fixed effects treatments from the empirical gravity literature.

Equation (3) motivates the use of three sets of fixed effects. The first two are the standard exporter-time fixed effects ($\pi_{i,t}$) and importer-time fixed effects ($\chi_{j,t}$) to control for any exporter-specific and importer-specific (time-varying) characteristics that may affect bilateral trade flows. These include the multilateral resistances, output, and expenditure. The structural terms from equation (3), which are controlled for by the country-time fixed effects, differ from those in the standard, long run gravity literature, but the fixed effects fully absorb all exporter-time and importer-time specific components of structural equation (3).

The third set of fixed effects comprises the lag-interval-time-pair fixed effects that control for variation the average trade share term $(\bar{s}_{ij,\Delta t})^{\frac{(1-\delta_{\Delta t})(1-\rho)}{1-\delta_{\Delta t}(1-\rho)}}$. In principle, it is possible to rely on actual trade data to capture this variable and to estimate its coefficient $\left(\frac{(1-\delta_{\Delta t})(1-\rho)}{1-\delta_{\Delta t}(1-\rho)}\right)$. In turn, in combination with the estimate of the coefficient on tariffs $\left(\frac{(1-\sigma)\rho}{1-\delta_{\Delta t}(1-\rho)}\right)$ from our structural model and with an external value of the elasticity of substitution ($\hat{\sigma}$), we could recover the structural incidence parameter ($\hat{\rho}$) and, more importantly, the sequence of adjustment parameters ($\hat{\delta}_{\Delta t}$), which are of central interest to us.

The downside of using lagged/interval trade as a covariate is the well-known issue of a potential dynamic panel bias (Nickell (1981) and Roodman (2009)), which may vary with the length of the interval that we use. Thus, even if we use trade data that are consistent with our model, we would not be confident whether the evolution of the transition parameter that we will obtain is due to the forces that we model or to the dynamic panel bias of our estimations.

To avoid such concerns, we will employ a reduced-form approach. Specifically, we will rely on interval-pair fixed effects ($\gamma_{ij,\Delta t}$). This approach has several advantages. First, generating and adding fixed effects to the gravity model is easy to implement. Second, the method is flexible, because the intervals can be defined over any desired time span, which is consistent with specification (2). Third, the approach is comprehensive because, in addition to controlling for averaged trade flows, the interval-pair fixed effects will capture and absorb the effects of any other bilateral factors that may impact the adjustment of trade costs. Fourth, on a related note and in the spirit of Baier and Bergstrand (2007), the use of interval-pair fixed effects may further mitigate potential endogeneity concerns with bilateral trade policy variables in the gravity model. Finally, the use of interval-pair fixed effects would lead to better estimates of the vector of bilateral trade costs and their changes, which may be very beneficial for counterfactual general equilibrium analysis.[4]

The lag-interval-time-pair fixed effects approach comes with caveats. For example, the interval-pair fixed effects may not allow identification of the effects of some covariates that are of interest to the researcher. Moreover, setting the length of the intervals is somewhat arbitrary, and this selection should be done carefully based on the features/dimensions of

[4]Note also that, even though this is not the objective of the current paper, in principle, it should be possible to recover the distribution of estimated time-varying bilateral trade costs, which could be useful for many reasons.

the data and the specific policy that is being evaluated. Finally, from a structural estimation perspective, we need external values for both $\hat{\rho}$ and $\hat{\sigma}$ to recover the key structural parameter of interest to us ($\delta_{\Delta t}$) from the estimate of the coefficient on tariffs. Fortunately, such estimates are available in the existing literature and we will use them to recover estimates of $\hat{\delta}_{\Delta t}$ in the empirical analysis.

Note, however, that no external parameters are needed to capture the change profile of the adjustment parameter. $\delta_{\Delta t}$ is the only time-varying component in the coefficient on tariffs $\left(\frac{(1-\sigma)\rho}{1-\delta_{\Delta t}(1-\rho)}\right)$. Inference about the rate of change of $\delta_{\Delta t}$ is based on the differences across lag interval specifications in the estimates of the coefficients on tariffs. The change profile allows us to 'test' some existing theories of the dynamic evolution of bilateral trade costs, e.g., Arkolakis (2010), in the empirical analysis.

Two steps complete the specification of our econometric model. First, we introduce two policy covariates: (i) a dummy variable for the presence of free trade agreements between countries i and j at time t ($FTA_{ij,t}$); and (ii) the log of applied tariffs ($LN\_TARIFF_{ij,\tau} = ln(1 + tar_{ij,\tau})$), where $tar_{ij,\tau}$ is the applied *ad-valorem* tariff on imports in destination j from source i. Our primary focus in the empirical analysis would be on the estimates of the coefficients on tariffs. We also analyze the implications of lag interval variation for explaining the phasing-in profile of the FTA estimates (e.g., Baier and Bergstrand (2007), Anderson and Yotov (2016)).

Following the recommendations of Santos Silva and Tenreyro (2006), we estimate the model with the Poisson Pseudo-Maximum-Likelihood (PPML) estimator, which simultaneously accounts for heteroskedasticity in the trade data and takes into account the information that is contained in the zero trade flows. Based on the above, our estimating model becomes:

$$X_{ij,t} = \exp[\beta_1 FTA_{ij,t} + \beta_2 LN\_TARIFF_{ij,t} + \pi_{i,t} + \chi_{j,t} + \gamma_{ij,\Delta t}] + \epsilon_{ij,t}. \quad (4)$$

The dependent variable in (4) is nominal bilateral trade flows in levels, including domestic trade flows (Yotov (2022)), and, following the recent recommendations of Egger et al. (2022), we use data for all years in the sample. We employ equation (4) to obtain the results in Section 5. Before that, we briefly describe our data.

4 Data: Description and Sources

To perform the empirical analysis, we use the dataset of Baier et al. (2019), which covers international and domestic aggregate manufacturing trade, free trade agreements, and tariffs for 52 countries over the period 1988-2006.[5] The two advantages of this dataset for our purposes are that (i) its time coverage is relatively long, and (ii) it includes data on tariffs, which we will use to identify the evolution of the trade elasticity in our setting. The only adjustment that we make to the original sample is to drop the year 1988. The reason is that the remaining 18 years of data would enable us to perform the estimations with balanced intervals of 3 years, 6 years, 9 years, and 18 years. Our conclusions do not change if we add 1988 to the sample and extend the first interval in each specification by one year.

The sources for the original data are standard and can be summarized as follows. Trade data come from the UN COMTRADE database, accessed via WITS. Domestic trade flows are constructed as apparent consumption, i.e., as the difference between the gross value of total production and total exports, and the data on total gross production come from the CEPII TradeProd database and UNIDO IndStat database. The original source of the tariff data is the United Nation's TRAINS database, and tariffs are aggregated to the level of the current analysis with import shares used as weights. Finally, the data on free trade agreements (FTAs) come from the NSF-Kellogg Database on Economic Integration Agreements of Jeff Bergstrand. For further description of the main dataset we refer the reader to Baier et al.

[5] The 52 countries/regions in the sample include: Argentina, Australia, Austria, Bulgaria, Belgium-Luxembourg, Bolivia, Brazil, Canada, Switzerland, Chile, China, Colombia, Costa Rica, Cyprus, Germany, Denmark, Ecuador, Egypt, Spain, Finland, France, United Kingdom, Greece, Hungary, Indonesia, Ireland, Iceland, Israel, Italy, Jordan, Japan, South Korea, Kuwait, Morocco, Mexico, Malta, Myanmar, Malaysia, Netherlands, Norway, Philippines, Poland, Portugal, Qatar, Romania, Singapore, Sweden, Thailand, Tunisia, Turkey, Uruguay, and the United States.

(2019).

5 Estimation results and quantitative implications

Our main findings are presented in Table 1 and visualized in Figure 1. The estimates in each of the four columns of this table are based on specification (4) with the only difference that, when we move from column (1) to column (4), we change the definition of $\gamma_{ij,\Delta t}$. Specifically, the estimates in column (1) are obtained with 3-year interval-pair fixed effects, which are constructed as the interaction between the pair fixed effects and 3-year intervals, i.e., 1989-1991, 1992-1994, ..., 2004-2006. The estimates in column (2) are obtained with 6-year interval-pair fixed effects, the estimates in column (3) are obtained with with 9-year interval-pair fixed effects, and, finally, the estimates in the last column are obtained with pair fixed effects that do not vary over time, i.e., for the whole 18-year period.

Table 1: Trade elasticities: From the short to the long run

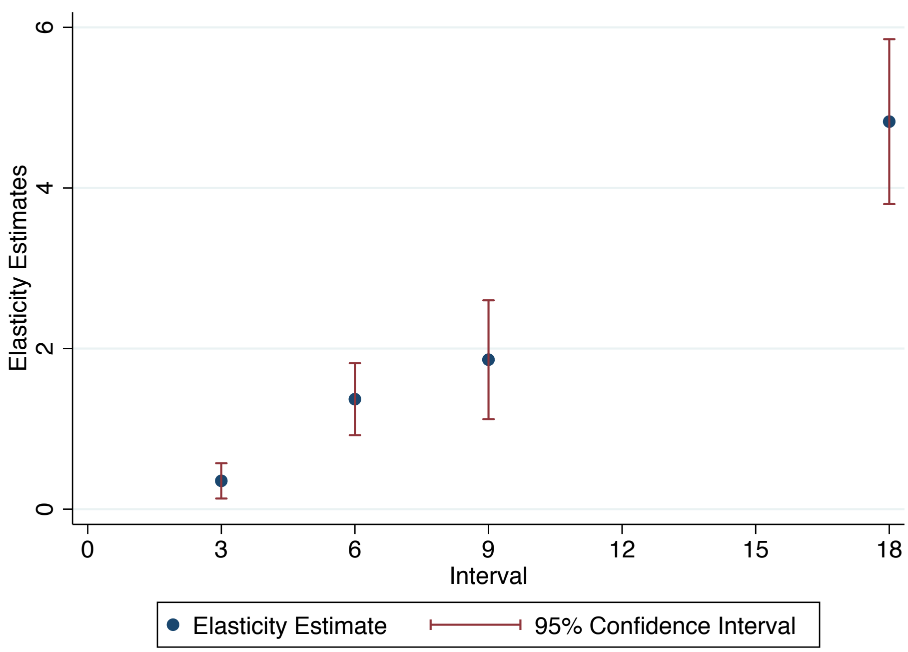
	(1) 3YRS	(2) 6YRS	(3) 9YRS	(4) 18YRS
LN_TARIFF	-0.352	-1.369	-1.861	-4.826
	(0.112)**	(0.229)**	(0.378)**	(0.524)**
FTA	0.028	0.096	0.180	0.224
	(0.024)	(0.037)**	(0.054)**	(0.059)**
N	48078	48390	48510	48636

Notes: This table reports gravity estimates of the effects of tariffs and FTAs. The results in each column are based on specification (4) and all estimates are obtained with the PPML estimator and exporter-time, importer-time, and directional country-pair fixed effects. The only difference when we move from column (1) to column (4) is that we change the definition of the pair fixed effects. Specifically, the estimates in column (1) are obtained with pair fixed effects that vary by 3-year intervals, the estimates in column (2) are obtained with pair fixed effects that vary by 6-year intervals, the estimates in column (3) are obtained with pair fixed effects that vary by 9-year intervals, and, finally, the estimates in the last column are obtained with pair fixed effects that do not vary over time. Standard errors are clustered by country pair. $^{+}$ $p < 0.10$, * $p < .05$, ** $p < .01$. See text for further details.

Three main findings stand out from Table 1. First, the estimates on tariffs in each column are negative and statistically significant. This confirms that, as expected, tariffs are an important impediment to international trade. Second, and most important for our purposes, we see from Table 1 and from Figure 1 that the estimates of the effects of tariffs increase in absolute value when we move from column (1) to column (4), varying from 0.4 to 1.4, to 1.9, and 4.8. This reveals that, as predicted by our theory, the trade elasticity gradually increases when we move from the short-run (macro) to the long-run (trade) trade elasticity. This is our solution to the 'international elasticity puzzle'.

Third, we see that the estimate with the 3-year interval-pair fixed effects is significantly smaller than the theory constraint lower bound of 1. We have two possible explanations for this result. One is mechanical, i.e., this is the specification with the largest number of fixed effects, which leaves us with relatively few degrees of freedom. Thus, the estimate may be driven by outliers. Second, on a related note but from a data perspective, when we estimated the model using only the first three years of the sample (1989-1991), we actually obtained a positive (and insignificant) estimate of the effects of tariffs.

Figure 1: Trade elasticities: From the short to the long run



Note: This figure plots estimates of the short and the long run trade elasticities that we report in Table 4, along with their corresponding confidence intervals. All are based on specification (4) and are obtained with the PPML estimator and exporter-time, importer-time, and directional country-pair fixed effects. The only difference when we move along the X-axis is that we change the definition of the pair fixed effects, which vary from 3-year, to 6-year, to 9-year, to 18-year interval-pair fixed effects. Standard errors are clustered by country pair.

Next, we use the estimates in Table 1, in combination with our theory, to draw some quantitative implications. Based on the previous discussion, we will not rely on the 3-year interval-pair fixed effects estimate for this analysis. In our first experiment, we recover the estimates of the adjustment parameter ($\delta_{\Delta t}$) that correspond to each of our specifications. To this end, we rely on the structural interpretation of the coefficient on tariffs and we borrow estimates of the elasticity of substitution and the incidence parameter from the literature. Specifically, we use $\hat{\sigma} = 5$, which is about the average from the existing literature, and $\hat{\rho} = 0.2$, which comes from Anderson and Yotov (2020), to solve:

$$\hat{\delta}_{\Delta t} = \left(\frac{\hat{\beta}_2 - (1-\hat{\sigma})\hat{\rho}}{\hat{\beta}_2(1-\hat{\rho})} \right) \tag{5}$$

Our conclusions are robust to using alternative (within theory bounds) values for both, the elasticity of substitution and the incidence elasticity parameters.

The adjustment parameters that we recover are within the theoretical bounds [0;1]. Specifically, the estimate of $\delta_{\Delta t}$ from the specification with 6-year interval-pair fixed effects is 0.537 (std.err. 0.118). $\delta_{\Delta t}$ increases to 0.723 (std.err. 0.105) when we estimate the model with with 9-year interval-pair fixed effects, and it becomes 1.041 (std.err. 0..022) when we use time-invariant pair fixed effects, i.e., with 18-year interval-pair fixed effects. Based on this, we draw two conclusions. First, as predicted based on our theory, the estimates of the adjustment parameter increase over time with the increase of the intervals we use. Second, the estimate that we obtain with the time-invariant fixed effects is very close to 1, which is the upper, long term theoretical bound for the adjustment parameter. One implication of these results is that, at least in our sample, 18 years are sufficient to reach the long-run trade equilibrium, while 9 years are still far from the long run.

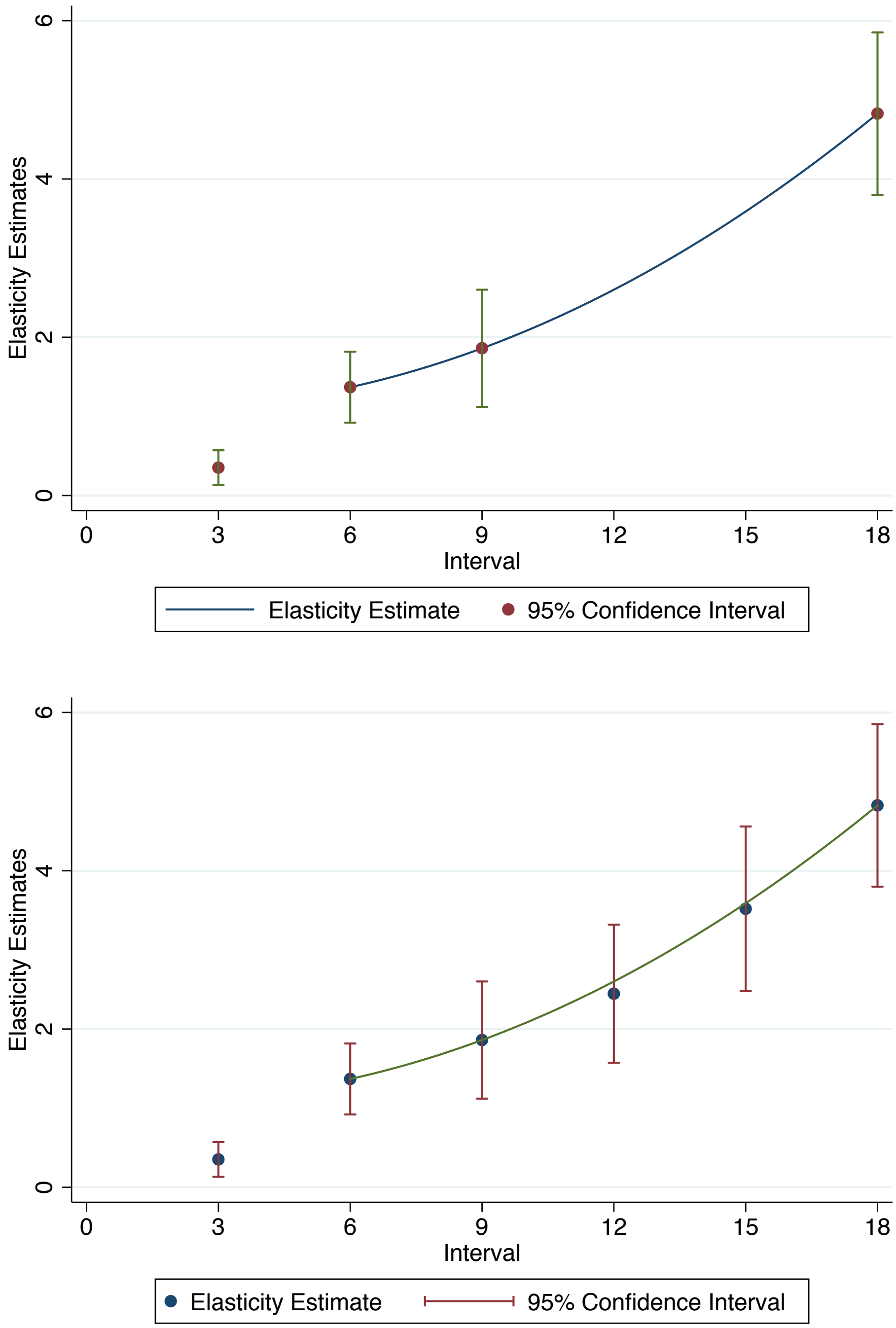
The next implication of our estimates is about the rate of change of $\delta_{\Delta t}$. As discussed earlier, we actually do not need external parameter estimates for this analysis. Instead, we use the estimates of the coefficients of tariffs from each of our specifications (except for the one with the 3-year interval-pair fixed effects) to construct a fitted line/curve, which appears in the top panel of Figure 2. Since $\delta_{\Delta t}$ is the only time-varying element in our estimates, it is exactly the evolution of this parameter that would determine that shape of the fitted curve. In other words, without $\delta_{\Delta t}$, the fitted curve would be just a line that would connect the single estimates of the short and long run elasticities (as in Anderson and Yotov (2020)).

The top panel of Figure 2 revels that the shape of the elasticities curve is convex. The implication for the evolution of the adjustment parameter is that when we move from the short to the long run $\delta_{\Delta t}$ increases at a decreasing rate, i.e., initial adjustment is faster, then it gradually becomes slower on the way to the long run equilibrium. The economic intuition for this result is consistent the theory of increasing marginal costs of trading from Arkolakis (2010).

The broad objective of our next experiment is to test and reinforce our conclusions thus far with alternative estimating samples. To this end, we perform two additional long-run estimations, where we limit the estimating sample to 12 years and 15 years, respectively, and we use time invariant pair fixed effects. The estimate of the coefficient on tariffs that we obtain with the 12-year sample is -2.447 (std.err. 0.445), and it implies a value of $\delta_{\Delta t} = 0.847$ (std.err. 0.072). The corresponding estimate that we obtain with the 15-year sample is -3.519 (std.err. 0.530), implying a value of $\delta_{\Delta t} = 0.967$ (std.err. 0.041).

This analysis validates and complements our previous findings. Specifically, if we add the new 12-year and 15-year estimates to the top panel of Figure 2, we obtain the bottom panel of the same figure. It reveals that the new estimates fall strikingly close next to the fitted curve. This has several implications in relation to our previous results and conclusions. First, the new estimates confirm the increasing trend in the adjustment parameter when we move from the short to the long run. Second, the new results are also consistent with our conclusion about the shape of adjustment. Third, in combination with the estimate that we obtain with the 18-year sample (1.041, std.err. 0..022), our 15-year-sample estimate implies that the long-run is reached at 16 or 17 years. Finally, a broad (and somewhat obvious)

Figure 2: Evolution of the trade elasticities



Note: This figure replicates the results from Figure 1 with two additions. First, the top panel of the figure introduced the fitted curve based on the estimates that we obtain with the 6-year, 9-year, and 18-year interval-pair fixed effects. In addition, the bottom panel of the figure introduces two more estimates, which are obtained with time-invariant fixed effects and are based on 12-year and 15-year samples, respectively. See text for further details.

implication of our analysis is that the estimates of the trade elasticity depend on the the length of the time period used to obtain them. Thus, one should be careful when comparing elasticity estimates from different studies, and do so only after conditioning on the time-spans of the corresponding samples used.

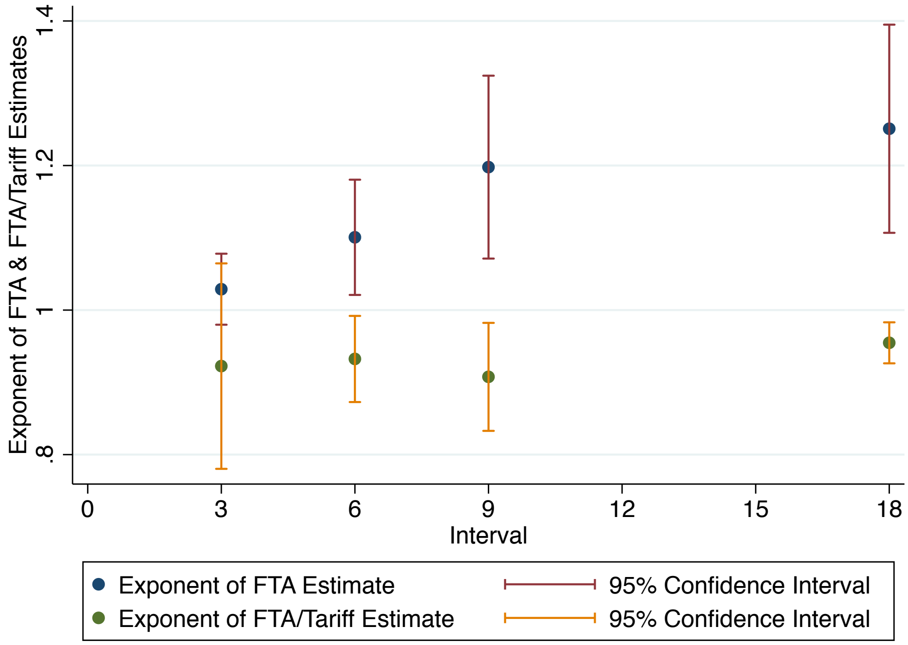
Unlike tariffs, the estimates of all other standard gravity variables in gravity regressions (e.g., distance, contiguity, FTAs, etc.) are a product of the structural power term $\frac{(1-\delta_{\Delta t})(1-\rho)}{1-\delta_{\Delta t}(1-\rho)}$ from equation (3) and another elasticity term (which we would call 'direct elasticity') that is specific to the gravity variable in question. For example, the coefficient of the effects of FTAs in our setting can be expressed as follows:

$$\beta_1 = \frac{(1-\delta_{\Delta t})(1-\rho)}{1-\delta_{\Delta t}(1-\rho)} \times \beta_{FTA} = \beta_2 \times \beta_{FTA}, \tag{6}$$

where, β_1 captures the total effects of FTAs on bilateral trade flows, and β_{FTA} is the 'direct elasticity' of trade with respect to FTAs. FTAs have been one of the most widely studied variables in gravity models and there is significant evidence that the phasing-in effects of FTAs follow a convex path, i.e., over time, they increase at a decreasing rate (Baier and Bergstrand (2007); Anderson and Yotov (2016); and Egger et al. (2022)). Our goal in the next experiment is to check how much of the evolution of the effects of FTAs over time can be explained by our theory and what is left for their direct elasticity to explain.

Figure 3 presents our findings. The figure plots two sets of estimates that correspond to each of the specifications from Table 1. Specifically, for each interval, the 'blue' estimates, along with the corresponding 95% confidence intervals (in red), are obtained as the exponent of the FTA estimates ($\hat{\beta}_1$) from Table 1. We label these estimates 'total' FTA effects. The 'green' estimates, along with the corresponding 95% confidence intervals (in orange), are obtained as the exponent of the ratio between the FTA and the LN_TARIFF estimates from Table 1. Thus, the green estimates are designed to capture the impact of the changes in the direct FTA elasticity (β_{FTA}) in the evolution of the total FTA effects.

Figure 3: Evolution of the FTA estimates

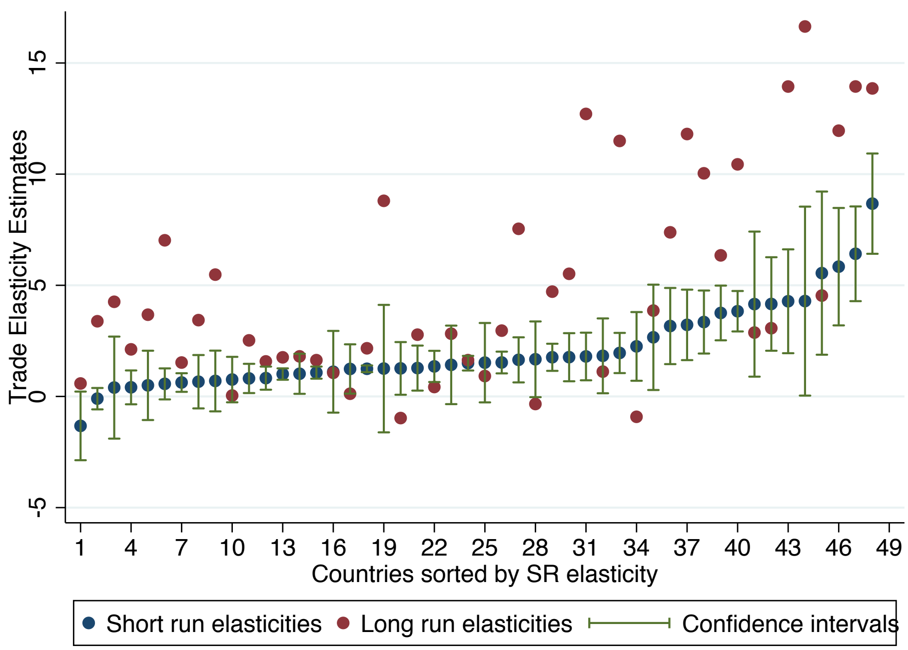


Note: This figure plots estimates of the short and the long run FTA elasticities. Both sets of elasticities are obtained from the estimates from specification (4) with the PPML estimator and with exporter-time, importer-time, and directional country-pair fixed effects, which are reported in Table 1. The 'blue' estimates, along with the corresponding 95% confidence intervals (in red), are obtained as the exponent of the FTA estimates from Table 1. The 'green' estimates, along with the corresponding 95% confidence intervals (in orange), are obtained as the exponent of the ratio between FTA and the tariff estimates from Table 1. See text for further details.

Two findings stand out from Figure 3. First, the evolution of the blue estimates, i.e., the total FTA effects, is consistent with the corresponding indexes from the literature, e.g., Egger et al. (2022). Specifically, we observe a gradual increase over time of the effects of FTAs, which becomes slower over time. Second, and more important for our purposes, we see that the evolution of the direct FTA elasticity (the green estimates) over time is relatively flat. The implication is that our theory explains quite well the evolution of the total FTA effects in our sample, and we do not find evidence that the 'direct elasticity' of trade with respect to FTAs varies over time.

To further demonstrate the robustness and flexibility of our methods, we also obtain short vs. long run elasticity estimates by country. To this end, we amend equation (4) to allow for country/importer-specific effects of tariffs. Even though the resulting specification becomes more demanding in terms of the number of parameters that we are after, we still estimate the short-run model with 3-year-pair fixed effects, which corresponds to the results from column (1) of Table 1, and we compare it with the long run model with pair fixed effects that do not vary over time, i.e., the one corresponding to column (4) of Table 1. Our results are reported in Figure 4, where we take the absolute value of the estimates and we sort them in an increasing order based on the estimates from the short run specification (i.e., the one with the 3-year-interval pair fixed effects). In addition to the point estimates, we also include the 95% confidence intervals from the short run specification.

Figure 4: Short vs. long run elasticities by country



Note: This figure plots estimates of the short and the long run trade elasticities. Both sets of elasticities are obtained from specification (4) with the PPML estimator and with exporter-time, importer-time, and directional country-pair fixed effects. The only difference between the two sets of elasticities is that the short run elasticities are obtained with pair fixed effects that vary by 3-year intervals, while the long run elasticities are obtained with pair fixed effects that do not vary over time. See text for further details.

Four main results stand out from Figure 4. First, most of the trade elasticities that we obtain are positive and statistically significant, as expected.[6] Second, the estimates

[6]To improve the clarity of Figure 1, we have dropped the elasticity estimates for three countries (Qatar, Kuwait, and Malta), because the corresponding short-run tariff estimates were very large in absolute value and, actually, positive.

of the trade elasticities vary within reasonable bounds. However, third, they are quite heterogeneous across the countries in our sample. This finding has potentially important implications for welfare analysis with the structural gravity model, which often are performed with a single trade elasticity estimate that is common across countries. Finally, consistent with the results from Table 1, and most important for our purposes, with relatively few exceptions, the long run trade elasticities that we obtain are larger than their short run counterparts.

6 Conclusion

We propose a simple and flexible econometric approach to estimate gravity models in the short vs. the long run. Our reduced-form specification is based on a theoretical setting that amends the standard Lucas-Prescott adjustment process to allow for non-linear rate of change of the adjustment parameter. This motivates the use of interval-pair fixed effects in an empirical specification that is representative of several alternative theories of bilateral adjustment in trade costs. We highlight the validity of our methods with an application that solves the 'international elasticity puzzle'. Specifically, we obtain trade elasticities that vary between 0.4 and 4.8 when we move from the short to the long run. We also use our results to test and confirm the theory of increasing marginal trading costs from Arkolakis (2010) and to explain the evolution over time of the FTA estimates in gravity models. Our estimates imply that a long-run equilibrium in our sample is reached in 16-17 years and that comparisons between trade elasticities from different studies should be conditional on the length of the time span of the corresponding estimating samples.

References

Agnosteva, Delina E., James E. Anderson, and Yoto V. Yotov, "Intra-national Trade Costs: Assaying Regional Frictions," *European Economic Review*, 2019, *112* (C), 32–50.

Anderson, James E., "The Specific Factors Continuum Model, with Implications for Globalization and Income Risk," *Journal of International Economics*, 2011, *85* (2), 174–185.

_ **and Eric van Wincoop**, "Borders, Trade and Welfare," *Brookings Trade Forum*, 2001, pp. 207–230.

_ **and Eric van Wincoop**, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 2003, *93* (1), 170–192.

_ **and Yoto V. Yotov**, "Terms of Trade and Global Efficiency Effects of Free Trade Agreements, 1990-2002," *Journal of International Economics*, 2016, *99*, 279–298.

_ **and Yoto Yotov**, "Short Run Gravity," *Journal of International Economics*, 2020, *126*, 103341.

_ **, Mario Larch, and Yoto V. Yotov**, "Transitional Growth and Trade with Frictions: A Structural Estimation Framework," *The Economic Journal*, 2020, *130* (630), 1583–1607.

Arkolakis, Costas, "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 2010, *118* (6), 1151–1199.

_ **, Arnaud Costinot, and Andrés Rodríguez-Clare**, "New Trade Models, Same Old Gains?," *American Economic Review*, 2012, *102* (1), 94–130.

Baier, Scott L. and Jeffrey H. Bergstrand, "Do Free Trade Agreements Actually Increase Members' International Trade?," *Journal of International Economics*, 2007, *71* (1), 72–95.

_ **, Yoto V. Yotov, and Thomas Zylkin**, "On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration," *Journal of International Economics*, 2019, *116* (C), 206–226.

Chaney, Thomas, "The Network Structure of International Trade," *American Economic Review*, 2014, *104* (11), 3600–34.

Crucini, Mario J. and J. Scott Davis, "Distribution capital and the short- and long-run import demand elasticity," *Journal of International Economics*, 2016, *100* (C), 203–219.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

_ **,** _ **, Brent Neiman, and John Romalis**, "Trade and the Global Recession," *American Economic Review*, 2016, *106* (11), 3401–3438.

Egger, Peter H. and Sergey Nigai, "Energy Demand and Trade in General Equilibrium," *Environmental and Resource Economics*, Feb 2015, *60* (2), 191–213.

_ **, Mario Larch, and Kevin E. Staub**, "Trade Preferences and Bilateral Trade in Goods and Services: A Structural Approach," *CEPR Working Paper No. 9051*, 2012.

Egger, Peter H., Mario Larch, and Yoto V. Yotov, "Gravity Estimations with Interval Data: Revisiting the Impact of Free Trade Agreements," *Economica*, January 2022, *89* (353), 44–61.

Fontagné, Lionel, Houssein Guimbard, and Gianluca Orefice, "Tariff-based product-level trade elasticities," *Journal of International Economics*, 2022, *137* (C).

_ **, Philippe Martin, and Gianluca Orefice**, "The international elasticity puzzle is worse than you think," *Journal of International Economics*, 2018, *115* (C), 115–129.

Giri, Rahul, Kei-Mu Yi, and Hakan Yilmazkuday, "Gains from trade: Does sectoral heterogeneity matter?," *Journal of International Economics*, 2021, *129* (C).

Head, Keith, Thierry Mayer, and John Ries, "The Erosion of Colonial Trade Linkages after Independence," *Journal of International Economics*, 2010, *81* (1), 1–14.

Lucas, Robert E. and Edward C. Prescott, "Investment Under Uncertainty," *Econometrica*, 1971, *39* (5), 659–681.

Nickell, Stephen J., "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 1981, *49* (6), 1417–26.

Roodman, David, "How To Do xtabond2: An Introduction to Difference and System GMM in Stata," *Stata Journal*, 2009, *9* (1), 86–136.

Ruhl, Kim J., "The International Elasticity Puzzle," unpublished manuscript, available for download at `http://www.kimjruhl.com/`, 2008.

Santos Silva, João M.C. and Silvana Tenreyro, "The Log of Gravity," *Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Simonovska, Ina and Michael Waugh, "The Elasticity of Trade: Estimates and Evidence," *Journal of International Economics*, 2014, *92* (1), 34–50.

Soderbery, Anson, "Estimating import supply and demand elasticities: Analysis and implications," *Journal of International Economics*, 2015, *96* (1), 1–17.

Yilmazkuday, Hakan, "Estimating the trade elasticity over time," *Economics Letters*, 2019, *183* (C), 1–1.

_ , "Understanding the international elasticity puzzle," *Journal of Macroeconomics*, 2019, *59* (C), 140–153.

Yotov, Yoto V., "On the role of domestic trade flows for estimating the gravity model of trade," *Contemporary Economic Policy*, February 2022.

Appendix: Short run gravity (Anderson and Yotov, 2020)

This Appendix is not intended for publication. It is only included for the convenience of the reader. In what follows, we have just copied and pasted the derivation of the short-run gravity model of Anderson and Yotov (2020) (see Proposition 1 below), which appears as equation (1) in the main text of the current paper.

An origin region i produces and ships a product to potentially many destinations j. Distribution on each link requires labor and capital in variable proportions, multiplicatively amplified by iceberg trade and production cost factors t_{ij}. t_{ij}s differ across origin-destination pairs according to bilateral geographical features such as distance and borders and other familiar variables in the gravity literature. Cost-minimizing allocation of resources implies that bilateral trade destinations are imperfect substitutes. One factor is variable in the short run – labor that can be freely allocated across destination markets – while capacity is fixed in the short run in each network link.

The model represents trade from each origin that requires destination-specific "marketing capital" that is committed (sunk) before the allocation of labor. "Marketing capital" is left vague to encompass both network connections (links between counter-parties that are inherently specialized and fixed in the short run) and physical capital particularized to serve a particular destination. The idea of a retail network in Crucini and Davis (2016) is similar.

Bilateral production and trade from origin i to destination j is formally modeled as a Cobb-Douglas function of labor L_{ij} and capital K_{ij}: $x_{ij} = (1/t_{ij})K_{ij}^{1-\alpha}L_{ij}^{\alpha}$. x_{ij} is delivered product. $t_{ij} > 1$ is the iceberg-melting parameter from origin i to destination j, a penalty imposed by 'nature' relative to the frictionless benchmark $t = 1$. t_{ij} also incorporates a productivity penalty in the usual sense that would apply to all destinations j uniformly. In the empirical application, t_{ij} will include tariffs and the effects of free trade agreements. With all inputs variable (in the long run), the production function has constant returns to scale. In the short run with K_{ij} fixed, decreasing returns dominate. For now, all firms are identical in any origin, aggregating to an industry with the representative firm's characteristics.

(Essentially the same short run gravity model extends to incorporate fixed infrastructure at each location and time, adding location-time-specific productivity shifters controlled for econometrically with location-time fixed effects.)

For the origin sector as a whole, labor supply L_i is drawn from a national labor market in an amount satisfying the value of marginal product condition at the national wage rate w_i. The labor market constraint on short run allocation across destinations is $L_i = \sum_{j=0}^{n} L_{ij}$. Destination $j = n$ is at the extensive margin, determined outside the model. L_i is efficiently allocated across destination activities with value of marginal product equal to the common wage: $w_i L_{ij} = \alpha(p_{ij}/t_{ij})x_{ij}$. Delivered price p_{ij} in competitive equilibrium covers costs: $w_i L_{ij} + r_{ij}K_{ij} = (p_{ij}/t_{ij})x_{ij}$ where r_{ij} is the realized (residual) return in origin i on the specific capital for delivery to destination j.

The representative firm in the origin sector effectively maximizes the value of delivered product to all destinations by efficiently allocating labor L_{ij} across destinations j:

$$\max_{L_{ij}} \sum_{j=0}^{n} \frac{p_{ij}}{t_{ij}} K_{ij}^{1-\alpha} L_{ij}^{\alpha} \mid \sum_{j} L_{ij} \leq L_i.$$

The value function works out to be[7]

$$Y_i = L_i^{\alpha} K_i^{1-\alpha} \mathbf{P}_i, \tag{7}$$

[7] The value of sectoral production at delivered prices in the origin country is equal to its cost: $Y_i = \sum_j (p_{ij}/t_{ij})x_{ij} = \sum_j [w_i L_{ij} + r_{ij}K_{ij}]$. The efficient allocation of labor across destinations implies $w_i = \alpha(p_{ij}/t_{ij})L_{ij}^{\alpha-1}K_{ij}, \ \forall j$. Solve for L_{ij}, substitute into the sectoral labor market clearance condition, and solve for the willingness to pay for sectoral labor:

$$w_i = L_i^{\alpha-1}\alpha \left[\sum_j K_{ij}(p_{ij}/t_{ij})^{1/(1-\alpha)}\right]^{1-\alpha}.$$

Multiply and divide the right hand side by $K_i^{1-\alpha}$ and multiply both sides by L_i. The resulting wage bill is $w_i L_i = \alpha L_i^{\alpha} K_i^{1-\alpha} \mathbf{P}_i \Rightarrow$ the aggregate cost of production and delivery is the value function. The setup here applies the specific factors production model of Anderson (2011) to production *cum* distribution over destinations for a single sector.

where the joint product deflator $\mathbf{P}_i$ is

$$\mathbf{P}_i \equiv [\sum_0^n \lambda_{ij}(p_{ij}/t_{ij})^{1/(1-\alpha)}]^{1-\alpha}, \tag{8}$$

and $\lambda_{ij} = K_{ij}/K_i$ and $K_i = \sum_j K_{ij}$. The real activity of production and delivery to many destinations

$$R_i = L_i^\alpha K_i^{1-\alpha}$$

in (7) is multiplied by the joint product deflator or 'average net price' $\mathbf{P}_i$ that embeds efficient allocation of labor.

The equilibrium share of sales to each destination j, applying Hotelling's Lemma to (7) and (8), is

$$s_{ij} = \lambda_{ij} \left(\frac{p_{ij}/t_{ij}}{\mathbf{P}_i} \right)^{1/(1-\alpha)}. \tag{9}$$

The equilibrium delivered price p_{ij} from origin i to each destination market j is endogenously determined by the supply side forces described in (7)-(9) interacting with demand forces described by CES preferences or technology (in the case of intermediate goods) over goods differentiated by place of origin.

Gravity modeling covers distribution from many origins to many destinations. The demand side of the model assumes a CES expenditure share for goods from origin i in destination j given by

$$\frac{X_{ij}}{E_j} = \left(\frac{\beta_i p_{ij}}{P_j} \right)^{1-\sigma}. \tag{10}$$

Here, X_{ij} denotes the bilateral flow at end user prices, E_j denotes the total expenditure in destination j on goods from all origins serving it, β_i is a distribution parameter of the CES preferences/technology, σ is the elasticity of substitution, and P_j is the CES price index for destination j.

The market clearing condition for positive bilateral trade from seller i (henceforth a subscript denoting origin) to j is

$$s_{ij} Y_i = X_{ij}.$$

Using (9) for s_{ij} and (10) for X_{ij} in the market clearing condition to solve for the equilibrium price p_{ij} yields:

$$p_{ij} = \left[\frac{E_j P_j^{\sigma-1} \beta_i^{1-\sigma} [t_{ij} \mathbf{P}_i]^\eta}{Y_i \lambda_{ij}} \right]^{1/(\eta+\sigma-1)}, \tag{11}$$

where $\eta = 1/(1-\alpha) > 1$ is the supply elasticity. The short run equilibrium price in origin-destination pair in (11) is an intuitive constant elasticity function of demand shifters $E_j P_j^{\sigma-1}$, supply shifters Y_i and $\mathbf{P}_i$, and the exogenous bilateral friction components in t_{ij} and the contemporaneously exogenous bilateral capacity λ_{ij}. The intuitive notion of a bilateral trade cost corresponds to p_{ij}/p_{i0}, so bilateral trade cost is endogenous.

Incidence of trade costs to buyers is given by the buyers' incidence elasticity $\partial \ln p_{ij}/\partial \ln t_{ij} = \eta/(\eta + \sigma - 1) \equiv \rho$. The incidence elasticity ρ (dropping "buyers" for brevity) plays a key role in the gravity representation of the model. ρ has a deep micro-foundation as a combination of the demand elasticity parameter σ and the supply elasticity parameter η, itself microfounded in the equilibrium of distribution based on the Cobb-Douglas specific factors model. ρ is increasing in η and decreasing in σ.[8]

Gravity in Short Run and Long Run

The gravity representation of short run equilibrium trade derives from exploiting the properties of market clearance globally, embedding the bilateral market clearance (11) along with the budget constraints that are part of the CES demand system. The global market clearing condition for Y_i implies multilateral resistances for sellers. Substitute (11) into (10), then multiply by E_j and sum over j to obtain the global demand for Y_i. Collect the terms for Y_i on the left hand side of the market clearance condition and simplify the exponents

[8]Equation (11) can, in principle, account for substantial variation in prices across time and space. Rents to the sector specific factor similarly vary. Rents are competitive in the model, so pricing to market behavior in the usual sense is not implied.

$1-\rho(\sigma-1)/\eta=\rho$ on Y_i and $(\sigma-1)/(\eta+\sigma-1)=1-\rho$ on λ_{ij}. The result is

$$Y_i^{\rho}=[\beta_i\mathbf{P}_i]^{\rho(1-\sigma)}\sum_j[E_jP_j^{\sigma-1}]^{\rho}t_{ij}^{\rho(1-\sigma)}\lambda_{ij}^{1-\rho}.$$

Divide both sides by Y^{ρ}. The result is

$$\left[\frac{Y_i}{Y}\right]^{\rho}=[\beta_i\mathbf{P}_i\Pi_i]^{\rho(1-\sigma)}\Rightarrow\frac{Y_i}{Y}=[\beta_i\mathbf{P}_i\Pi_i]^{1-\sigma}, \tag{12}$$

where outward multilateral resistance is

$$\Pi_i^{(1-\sigma)\rho}=\sum_j\left(\frac{E_j}{Y}\right)^{\rho}\left(\frac{t_{ij}}{P_j}\right)^{(1-\sigma)\rho}\lambda_{ij}^{1-\rho}. \tag{13}$$

The left hand side of (12) is recognized as a CES share equation for a hypothetical world buyer on the world market, with a world market price index for all goods equal to 1. Short run multilateral resistance in (13) is a CES function of bilateral relative trade costs t_{ij}/P_j, where the elasticity of short run substitution is $(1-\sigma)\rho$. Π_i is homogeneous of degree one in $\{t_{ij}\}$ for given $\{P_j\}$.

The gravity representation of trade also requires using a relationship between the buyers' price index (an implication of the budget constraint of the CES demand system) and relative trade costs. Substitute (11) in the CES price definition $P_j^{1-\sigma}=\sum_i[\beta_ip_{ij}]^{1-\sigma}$. Then use (12) to substitute for $[\beta_i\mathbf{P}_i]^{1-\sigma}$ in the resulting equation. After simplification this gives the short run price index as

$$P_j^{(1-\sigma)\rho}=\left(\frac{E_j}{Y}\right)^{-(1-\rho)}\sum_i\frac{Y_i}{Y}\left(\frac{t_{ij}}{\Pi_i}\right)^{(1-\sigma)\rho}\lambda_{ij}^{1-\rho}.$$

Buyers' price index P_j is the product of a size effect $[E_j/Y]^{-(1-\rho)}$ and buyers multilateral resistance:

$$\widetilde{P}_j^{(1-\sigma)\rho}=\sum_i\frac{Y_i}{Y}\left(\frac{t_{ij}}{\Pi_i}\right)^{(1-\sigma)\rho}\lambda_{ij}^{1-\rho},\ \forall j. \tag{14}$$

$\widetilde{P}_j$ is also interpreted as the buyers' short run overall incidence of trade costs. Simplifying the CES price index, $P_j = [E_j/Y]^{(1-\rho)/(\sigma-1)\rho}\widetilde{P}_j$. Higher relative demand E_j/Y raises P_j given $\widetilde{P}_j$ due to fixed capacities $\{\lambda_{ij}Y_i\}$. In long run gravity, as effectively $\eta \to \infty$, $\rho \to 1$ and the buyers' market size effect vanishes from the price index P_j.

Use $P_j^{(1-\sigma)\rho} = [E_j/Y]^{-(1-\rho)}\widetilde{P}_j^{(1-\sigma)\rho}$ in sellers' multilateral resistance (13) to yield the more intuitive equivalent form

$$\Pi_i^{(1-\sigma)\rho} = \sum_j \frac{E_j}{Y}\left(\frac{t_{ij}}{\widetilde{P}_j}\right)^{(1-\sigma)\rho} \lambda_{ij}^{1-\rho},\ \forall i. \tag{15}$$

The final step in deriving short run gravity is to substitute the right hand side of (11) for p_{ij} in (10) and use (12) to substitute for $[\beta_i \mathbf{P}_i]^{1-\sigma}$ in the resulting expression. After simplification using incidence elasticity $\rho = \eta/(\eta + \sigma - 1)$, this gives:[9]

Proposition 1: Short Run Gravity. *Short run gravity trade flows are given by:*

$$X_{ij} = \frac{Y_i E_j}{Y}\left[\frac{t_{ij}}{\Pi_i \widetilde{P}_j}\right]^{(1-\sigma)\rho} \lambda_{ij}^{1-\rho}. \tag{16}$$

where the multilateral resistances Π_i, $\widetilde{P}_j$ *are given by (14)-(15).*

The first term on the right hand side of (16) is the frictionless benchmark flow at given sales Y_i and expenditure E_j. The middle term is the familiar effect of gravity frictions, the ratio of bilateral to the product of buyers' and sellers' multilateral resistance. The difference is that the short run trade elasticity is reduced in absolute value to $(1-\sigma)\rho$. The last term $\lambda_{ij}^{1-\rho}$ captures the effect of investment in capacity on link ij. Dividing both sides of (16) by

[9] $X_{ij} = E_j(\beta_i p_{ij})^{1-\sigma} = E_j^{\rho}[t_{ij}/P_j]^{(1-\sigma)\rho}\lambda_{ij}^{1-\rho}H_i$ where $H_i = [\beta_i \mathbf{P}_i]^{(1-\sigma)\rho}Y_i^{1-\rho}$. Substitute $[Y_i/Y\Pi_i^{1-\sigma}]^{\rho}$ for $[\beta_i \mathbf{P}_i]^{(1-\sigma)\rho}$ in H_i and replace $P_j^{(1-\sigma)\rho}$ with $[E_j/Y]^{-(1-\rho)}\widetilde{P}_j^{(1-\sigma)\rho}$. Rearranging the result yields equation (16).

the frictionless benchmark, size adjusted trade is

$$\frac{X_{ij}}{Y_i E_j/Y} = \left[\frac{t_{ij}}{\Pi_i \widetilde{P}_j}\right]^{(1-\sigma)\rho} \lambda_{ij}^{1-\rho},$$

a geometric weighted average of long run gravity frictional displacement and short run link capacity allocation.

Intuition about short run gravity system (14)-(16) is aided by considering an equiproportionate change in all bilateral trade costs t_{ij}: $t_{ij}^1 = \mu t_{ij}^0$. Intuitively, bilateral trade flows should be unchanged because no relative price changes. Checking the system (14)-(15), $\{\widetilde{P}_z, \Pi_i\}$ are homogeneous of degree 1/2 in $\{t_{ij}\}$, hence buyers' and sellers' multilateral resistances change by $\mu^{1/2}$ so indeed bilateral trade flows are constant. As with long run gravity, system (14)-(15) solves for multilateral resistances up to a normalization. Multilateral resistances retain their interpretation as sellers' and buyers' incidence of trade costs.

Over time the short run allocation of destination specific capital $\{\lambda_{ij}\}$ presumably moves toward the long run efficient level. (Whether the short run allocation is efficient or not depends on initial conditions and adjustment costs that are outside the model.) The long run efficient allocation matches the long run demand pattern, so that the short run gravity equation (16) approaches the long run gravity equation, intuitively equivalent to $\rho \to 1$.

The preceding derivation of (16) combined with (13) and (14) uses for simplicity the Armington CES/endowments setup of Anderson and van Wincoop (2003), but essentially the same short run gravity structure emerges when the supply structure (more realistically) includes an extensive margin in addition to the intensive margin above. The composite supply elasticity combines intensive margin elasticity η above with an extensive margin elasticity based on the dispersion parameter θ of the Pareto productivity distribution. The composite supply elasticity is $\tilde{\eta} = \eta(1 + \theta - \eta) < \eta$ for the realistic case $\theta < \eta$, the necessary and sufficient condition for the average productivity of firms serving a destination to be rising in net sellers' price p_{ij}/t_{ij}. The incidence elasticity becomes $\tilde{\rho} = \tilde{\eta}/(\tilde{\eta} + \sigma - 1) < \rho$ for $\theta < \eta$.

The extensive margin parameter θ plays a role below in discussing the implications of the estimated $\hat{\rho}$ for the underlying structural parameters. Short run gravity also encompasses the Eaton and Kortum (2002) interpretation of gravity based on heterogeneous productivity draws in a Ricardian model. Thus, the short run gravity model developed here extends to the wide class of models described in Arkolakis et al. (2012).