Sticky Gravity

Mario Larch, Leandro Navarro, and Dennis Novy

(This paper also appears as CAGE Discussion paper 800)

April 2026 No: 1609

Warwick Economics Research Papers

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

Sticky Gravity[*]

Mario Larch[†], Leandro Navarro[†], and Dennis Novy[§]

[†]University of Bayreuth
[§]University of Warwick

April 2026

Abstract

International trade flows show strong persistence over time. Standard static gravity models cannot rationalize this persistence and lack a micro-foundation for including lagged trade flows as a determinant of current trade. We develop a structural dynamic gravity framework in which persistence arises from firms' sluggish adjustment of destination-specific prices, analogous to sticky prices in macroeconomics but operating at the bilateral level. The model delivers a gravity equation with lagged trade flows as a structural feature rather than an ad hoc add-on. We propose a novel estimation approach for dynamic gravity models that explicitly accounts for persistence. Empirically, we show that ignoring persistence can lead standard gravity estimates to substantially understate the effects of trade policy changes. As an application, we find that the estimated trade impact of regional trade agreements can increase by 30 percent or more once persistence is taken into account.

JEL Classification: E31, F13, F14, F41
Keywords: Dynamic gravity, persistence, sluggish price adjustment, sticky prices, RTA, trade costs

[*]We thank Swati Dhingra, Hanwei Huang, Nuno Limão, Amrei Stammann, Thomas Zylkin as well as conference and seminar participants at CEPII, the European Trade Study Group conference 2023, the City University of Hong Kong Conference on International Economics 2024, the Contributions of Jeff Bergstrand Conference 2024, the Göttingen Workshop on International Economic Relations 2024, LACEA-LAMES Montevideo 2024, the London School of Economics, the Midwest International Trade Conference Fall 2023, the TRISTAN Workshop at the University of Bayreuth 2024, and the University of Trier for helpful comments. The usual disclaimer applies. Novy gratefully acknowledges research support from the Economic and Social Research Council (ESRC grant ES/Z504701/1). Calculations were performed using the festus cluster of the Bayreuth Centre for High Performance Computing (https://www.bzhpc.uni-bayreuth.de), funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 523317330. Mario Larch: Department of Law, Business & Economics, University of Bayreuth, Bayreuth, Germany; Centre d'Études Prospectives et d'Informations Internationales (CEPII), Paris, France; ifo Institute and CESifo Research Network, Munich, Germany; Nottingham Centre for Research on Globalisation and Economic Policy (GEP), Nottingham, UK; mario.larch@uni-bayreuth.de. Leandro Navarro: Department of Law, Business & Economics, University of Bayreuth, Germany; leandro.navarro@uni-bayreuth.de. Dennis Novy: Department of Economics, University of Warwick, Coventry CV4 7AL, United Kingdom; CEPR; CEP/LSE; CESifo; CAGE; d.novy@warwick.ac.uk.

1 Introduction

The gravity framework is one of the most popular empirical models in economics and the workhorse for empirical research and policy analysis in international trade. If lagged trade flows are included as a regressor in an otherwise standard gravity specification, they are not only highly significant but also boost the explanatory power of the regression.[1] Bilateral trade flows therefore appear to be persistent. However, such empirical evidence of persistence in trade flows does not sit easily with the theoretical foundations of the gravity equation. In the standard gravity framework, lagged trade flows from previous periods should not matter as an explanatory variable. As the standard gravity framework is a static model, there is no theoretical justification for including past trade flows.

The empirical importance of lagged trade flows hints at the relevance of shocks that leave a dynamic bilateral footprint. Such shocks are not absorbed by time-invariant bilateral fixed effects. Neither can they be explained by unilateral shocks (such as productivity or demand) at the level of exporters or importers. Those shocks would be absorbed by exporter-time and importer-time fixed effects.

In this paper, we incorporate trade flow persistence into a dynamic gravity framework. We set up a dynamic model of international trade under monopolistic competition. In our model, firms face sticky prices in destination markets similar to Calvo (1983) but at the bilateral level. That is, in each period, some firms can optimally adjust their prices in the destination markets, whereas other firms cannot. This setup leads to a gravity equation with bilateral persistence. We show that not taking into account persistence can have substantial implications for the evaluation of trade policies.

In the steady state, i.e., in the absence of shocks, our model nests the static structural gravity framework as in Anderson and van Wincoop (2003). But in response to a trade cost shock, bilateral trade is characterized by two types of trade flows: those by firms setting prices flexibly and those by firms subject to price stickiness. We show that the two types of trade flows can be represented by a two-part gravity equation: one part related to flexible-price flows, which follows the standard gravity form, and the other part related to sticky-price flows, which has lagged trade flows as an independent variable. Using the two-part gravity equation, we obtain estimates of the degree of stickiness in the data, and recover the implied steady-state elasticity of trade cost shocks. In a nutshell, we provide a micro-foundation for why lagged trade values enter an otherwise standard gravity equation.

To motivate our approach, we argue that the adjustment to shocks and trade policies, such as the conclusion of trade agreements or sanctions, takes time. This adjustment affects their impact on trade flows and must be taken into account for proper quantification. In our empirical application, we obtain a smaller elasticity of trade flows with respect to trade cost shocks in the transition phase when stickiness is still present in the system, and a stronger trade elasticity in the new steady state when stickiness has dissipated. The difference between the transition and steady-state elasticities implies a difference between the short-run and long-run impacts of trade cost shocks on trade flows,

[1] See Appendix A for details.

thus improving our understanding and quantification of trade policy changes. Intuitively, this means firms take time to fully react to trade policy changes, as reflected in the sluggish adjustment of prices and trade flows.

An additional novelty of our approach is that we can estimate the Calvo parameter that governs the degree of consumer price stickiness in the system in response to shocks. That is, we can estimate the share of firms that adjust their prices in each period from aggregate bilateral trade data without information on prices or unit values. We simulate the model to illustrate the dynamic impact of trade cost shocks on trade flows, prices and welfare, and we analyze the performance of our methodology to recover the Calvo parameter.

In our empirical analysis, we confront the model with trade data for OECD countries at monthly frequency. First, we document the importance of persistence in bilateral trade flows. Lagged trade flows are highly significant, particularly at monthly frequency, but less so at longer intervals such as quarterly or six-month frequencies. Second, we find a high degree of price stickiness of 95% at monthly frequency. As one would expect, we find that stickiness declines systematically with lower data frequency, falling to 84% at six-month frequency. We also show why the coefficient on lagged trade flows in standard gravity equations (i.e., those without a micro-foundation for dynamic behavior) is below unity and falls at longer horizons. These estimates align with the model's prediction of gradual adjustment: transition trade elasticities are weaker than steady-state elasticities, and the gap narrows with longer time horizons as more firms are able to adjust.

We then apply the framework to quantify the trade effects of specific regional trade agreements (RTAs). For each agreement, we use our estimation procedure to recover the implied price stickiness parameter and the corresponding steady-state trade cost elasticity. We compare these steady-state elasticities to those obtained from a standard (static) gravity specification and analyze their relationship. A central determinant of their difference is the time elapsed since entry into force: as an agreement matures and trade flows approach a new steady state, the scope for further adjustment shrinks. We also find that holding other factors constant, deeper preferential tariff reductions are associated with stronger adjustment over time in our dynamic framework.

Related literature. Most of the literature in trade focuses on static and long-term equilibria and pays little attention to transition effects. Our aim is to explicitly account for transition effects and the persistence of trade flows over time. We develop a theory that leads to a dynamic gravity equation, i.e., a gravity equation in which lagged trade flows emerge as an explanatory variable for current trade flows. Lagged trade flows will not appear in gravity models as long as their dynamics work via *country-specific* channels such as output and expenditure. By contrast, to generate dynamics with lagged trade flows we need *bilateral* variables that slowly adjust over time. For that purpose, we tap into the large macroeconomics literature on sticky prices. But unlike in a closed economy, price stickiness in our setting is origin-destination specific and thus materializes at the bilateral level.[2]

[2] As a rare example of price stickiness in the trade literature, Rodríguez-Clare, Ulate, and Vásquez (2026) incorporate downward nominal wage rigidities into a dynamic trade and migration model. We focus on stickiness at the bilateral level.

Theory-consistent dynamic trade models typically lead to a *standard* gravity equation. In such a gravity equation, bilateral dynamics do not play a role (for prominent examples, see Eaton, Kortum, Neiman, and Romalis, 2016; Caliendo, Dvorkin, and Parro, 2019; Anderson, Larch, and Yotov, 2020). Only a few papers have explored the role of lagged trade flows. As an early contribution, Eichengreen and Irwin (1998) establish the statistical relevance of lagged trade flows, followed in more depth by Jung (2012) and Olivero and Yotov (2012).[3] We differ in that we introduce price stickiness at the bilateral level, and we work with data at various frequencies.

Cheng and Wall (2005) advocate larger intervals between time periods to allow for sluggish adjustment processes. Baier and Bergstrand (2007) introduce "phased-in" free trade agreement effects, assuming that trade cost reductions occur in a staggered manner over time. This approach imposes dynamics onto the trade cost specification, and due to price flexibility, firms fully adjust to trade cost changes every period. This is inconsistent with evidence on sticky consumer prices, and differs from our approach where not all firms adjust prices every period.

We also relate to the literature that focuses on sunk costs as a mechanism for hysteresis and sluggish adjustment (see Alessandria, Arkolakis, and Ruhl, 2021 for a survey, and Carballo, Handley, and Limão, 2022; Egger, Foellmi, Schetter, and Torun, 2025).[4] Option values typically imply asymmetric adjustments to trade cost shocks depending on whether shocks are negative or positive. Anderson and Yotov (2020) suggest a short-run gravity model based on the idea of bilateral 'marketing capital' (following Arkolakis, 2010 and consistent with Chaney, 2014). Anderson and Yotov (2022) build on this framework to investigate the international elasticity puzzle.[5] However, as they explain, the bilateral dynamics are not micro-founded and introduced in an ad hoc manner.[6] Our framework complements these studies, but it implies symmetric adjustment to trade cost shocks.

The literature on international macroeconomics has extensively studied the pass-through of exchange rates to prices and infrequent price adjustment.[7] Gopinath and Rigobon (2008) study import price rigidity and low exchange rate pass-through to US import prices, finding heterogeneous price duration across goods according to the Rauch (1999) classification. Whereas the median price duration for goods under organized exchange is 1.2 months and 3.3 months for goods under reference-priced markets, it reaches 14.2 months for differentiated goods. Our approach is consistent with infrequent price adjustment. We focus on the pass-through of bilateral trade cost shocks when

[3] In addition, Campbell (2010) shows the persistence of trade flows over the long run from 1870 to 2000. Other papers with dynamic trade specifications include de Nardis, De Santis, and Vicarelli (2008) and Martínez-Zarzoso, Nowak-Lehmann, and Horsewood (2009).

[4] Egger, Foellmi, Schetter, and Torun (2025) set up a dynamic framework in which trade flows are shaped by both contemporaneous and past trade frictions. While their approach leads to long-term persistence in trade flows, it is inconsistent with price stickiness as firms adjust prices flexibly.

[5] Consistent with our empirical results, Boehm, Levchenko, and Pandalai-Nayar (2023) find that tariff elasticities are smaller in magnitude over the short run than the long run.

[6] They write: "'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."

[7] Burstein and Gopinath (2014) provide a review of the literature. Auer, Burstein, and Lein (2021) extend the analysis to retail prices and consumer expenditure, also finding low pass-through of border price changes to retail prices.

firms face sticky prices in their destination markets, and we are interested in the implications for the dynamic behavior of trade flows.

In section 2, we introduce our theoretical model and derive the gravity equation with transition dynamics. In section 3, we derive the estimating equations for the transition dynamics implied by the theory. In section 4, we analyze the transition dynamics of our model with simulated data, and in section 5 we provide empirical evidence. In section 6, we conclude.

2 Theoretical model

We develop a micro-founded dynamic general equilibrium model that results in a gravity equation with lagged trade flows as an explanatory variable. To achieve this goal, we introduce *bilateral* price stickiness in a simple trade model.

We define a world economy with multiple countries and one differentiated final good. We assume CES preferences in country j in period t:

$$U_{j,t} = \left[\sum_{n \in N_{i,t} \forall i \in \mathcal{J}} [q_{ij,t}(n)]^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)}, \tag{1}$$

with $q_{ij,t}(n)$ denoting the quantity consumed in country j of the final good variety n from country i in period t. $N_{i,t}$ refers to the number of varieties in country i in period t, $\mathcal{J}$ is the set of countries in the world, and $\sigma > 1$ is the elasticity of substitution across varieties.

Each firm produces one final good variety under monopolistic competition, using domestic labor as the only production factor, with one unit of labor required per unit of output. At the beginning of each period, each firm observes the wage and trade costs to each destination market. The total labor force in country i in period t is exogenous and denoted by $L_{i,t}$.

For simplicity, we assume that each country has an exogenously given number of firms (i.e., $N_{i,t} = N_i \; \forall \; t$) that is sufficiently large and time-invariant. As we apply our estimation approach to aggregate bilateral trade data, we further assume that firms are homogeneous in productivity as in Krugman (1980), as we use aggregate trade where we do not observe firms directly.

A key feature of our model is that firms face rigidities when adjusting consumer prices across destination markets. We characterize consumer price rigidity in a simple and tractable way following Calvo (1983), but we apply the approach to the *bilateral* dimension. That is, we assume that in each period t, with exogenous probability $(1-\alpha)$, a firm of country i supplying variety n can adjust its consumer price $p_{ij,t}(n)$ for the destination market j, and with probability α the firm cannot adjust its consumer price. Therefore, a share α of firms do not adjust their prices (i.e., $N_i^s = \alpha N_i$ where N_i^s stands for the number of 'sticky' firms). The remaining share $(1-\alpha)$ of firms adjust their prices (i.e., $N_i^f = (1-\alpha)N_i$ where N_i^f stands for the number of 'flexible' price-adjusting firms). The parameter α thus represents the degree of consumer-price stickiness. Note that the ability to adjust consumer prices is idiosyncratic, i.e., firm-destination-specific, and independent across firms and destinations.

That is, a given firm may be able to adjust its price in one destination, but not in another.[8]

2.1 Firms' pricing decisions under flexible and sticky prices

As a benchmark, we first consider the standard environment in which all firms can flexibly adjust prices ($\alpha = 0$). The optimal producer price by country i firms is $p_{i,t} = \sigma/(\sigma-1)w_{i,t}$, where $w_{i,t}$ is the domestic wage. The optimal flexible consumer price in destination j is given by $\tilde{p}_{ij,t} = t_{ij,t}p_{i,t}$, where $t_{ij,t}$ represents the bilateral trade costs. We can think of this flexible-price environment as the steady state.

2.1.1 Prices of flexible (price-adjusting) firms

We now turn to the environment of sticky consumer prices ($0 < \alpha < 1$). We start by analyzing the pricing decision of firms that can flexibly adjust their consumer price during period t.

Price-adjusting firms—from now on, for simplicity, 'flexible' firms—in country i for destination j in period t set the optimal price $p^f_{ij,t} \equiv \arg\max_p V_{ij,t}(p|\mathcal{S}_t)$, characterized by the following Bellman equation:

$$V_{ij,t}(p|\mathcal{S}_t) = \Pi^f_{ij,t}(p|\mathcal{S}_t) + \delta\left[\alpha\mathbb{E}_t V_{ij,t+1}(p|\mathcal{S}_{t+1}) + (1-\alpha)\mathbb{E}_t V_{ij,t+1}(\mathcal{S}_{t+1})\right], \tag{2}$$

where $\Pi^f_{ij,t}$ denotes the profits of flexible firms in country i selling to destination j in period t, $\mathcal{S}_t$ denotes the state of the system (i.e., the world economy) in period t, and $\delta \in (0,1)$ is a discount factor. These firms anticipate that in period $t+1$, with probability α they will be stuck with the current price $p^f_{ij,t}$, and with probability $(1-\alpha)$ they will be able to adjust the consumer price. Note that due to the homogeneity of firms, every price-adjusting firm of country i sets the same consumer price in destination j in period t.

The expectations in the Bellman equation allow for different types of information sets or belief formation, including when firms can perfectly predict future general equilibrium effects on the entire future equilibrium path of a given shock. It is possible to show that $p^f_{ij,t} \to \tilde{p}_{ij,t}$ (with $\tilde{p}_{ij,t}$ denoting the optimal flexible consumer price) for the following cases. First, when $\delta \to 0$, flexible firms set a price that maximizes current profits, neglecting the effects on expected future profits. Second, when the economy achieves a new steady state (i.e., $\mathcal{S}_t = \mathcal{S}_{t+1}$), firms set the optimal flexible consumer price. Third, when firms have limited information or beliefs about the current state $\mathcal{S}_t$, they cannot form expectations about future general equilibrium adjustments. And finally, when $\alpha \to 0$, we are in the steady-state benchmark where all firms set prices flexibly in every period.

In general terms, we express the price set by price-adjusting firms with the help of a shifter as:

$$p^f_{ij,t} = \kappa^f_{ij,t}\tilde{p}_{ij,t}, \tag{3}$$

[8] We also assume that trade costs are sufficiently high such that price differences across destinations cannot be arbitraged away. That is, a firm cannot sell its goods to one destination where it can adjust its price and then resell to other destinations where it cannot adjust.

where $\kappa^f_{ij,t}$ is a shifter defined by expected future rigidities. That is, $\kappa^f_{ij,t}$ represents a deviation from the optimal flexible consumer price $\tilde{p}_{ij,t}$ due to possible restrictions on adjusting future prices. Note that $p^f_{ij,t} \to \tilde{p}_{ij,t}$ when $\kappa^f_{ij,t} \to 1$.

2.1.2 Prices of sticky (non-adjusting) firms

We now turn to firms from country i producing varieties $n \in N^s_i$ that are unable to adjust their prices in destination j in period t. The Calvo structure implies that these firms charge different prices depending on the time elapsed since their last price adjustment, which is firm-destination-specific. For example, in a given destination market, one firm might have set a new consumer price one period ago, and another firm might have set a new consumer price two periods ago.

We denote as $p^s_{ij,t}(n)$ the consumer price in destination j in period t of the variety produced by sticky firm n from country i. As in Calvo (1983), by the law of large numbers, the average price of sticky firms is $p^s_{ij,t} \equiv \bar{p}_{ij,t-1}$, where $\bar{p}_{ij,t-1}$ refers to the average price set by all firms from country i in destination j in period $t-1$.

Analogous to the notation above, we define $\kappa^s_{ij,t}$ as a shifter capturing the deviation of the average sticky price $p^s_{ij,t}$ from the optimal flexible consumer price $\tilde{p}_{ij,t}$:

$$p^s_{ij,t} = \kappa^s_{ij,t}\tilde{p}_{ij,t}. \tag{4}$$

It follows $p^s_{ij,t}/p^f_{ij,t} = \kappa^s_{ij,t}/\kappa^f_{ij,t} \equiv \kappa_{ij,t}$, where $\kappa_{ij,t}$ is the ratio between the average price set by sticky firms and the price set by flexible firms. Since sticky firms cannot adjust consumer prices after bilateral trade costs or wages change, they endogenously adjust their markups. Markups are heterogeneous, depending on the time that has elapsed since firms last adjusted their prices.

2.2 Aggregates

Given the symmetry of flexible firms and the heterogeneity of sticky firms, we can rewrite the preferences (1) as:

$$U_{j,t} = \left[\sum_{i \in \mathcal{J}} \left(N^f_i [q^f_{ij,t}]^{(\sigma-1)/\sigma} + \sum_{n \in N^s_i} [q^s_{ij,t}(n)]^{(\sigma-1)/\sigma} \right) \right]^{\sigma/(\sigma-1)}, \tag{5}$$

with the budget constraint given by

$$\sum_{i \in \mathcal{J}} \left(N^f_i p^f_{ij,t} q^f_{ij,t} + \sum_{n \in N^s_i} p^s_{ij,t}(n) q^s_{ij,t}(n) \right) = E_{j,t}, \tag{6}$$

where $q^f_{ij,t}$ denotes the quantity of varieties $n \in N^f_i$ from flexible firms in country i, $q^s_{ij,t}(n)$ refers to the quantities of varieties $n \in N^s_i$ from sticky firms in country i, and $E_{j,t}$ represents total expenditure.

Total expenditure in destination j in period t on varieties produced by flexible firms from country

i is:

$$X^f_{ij,t} = N^f_i \left(\frac{p^f_{ij,t}}{P_{j,t}} \right)^{1-\sigma} E_{j,t}, \tag{7}$$

where equation (7) uses the property that all flexible firms in country i set the same price. Total expenditure in destination j in period t on varieties produced by sticky firms from country i is:

$$X^s_{ij,t} = \left(\frac{P^s_{ij,t}}{P_{j,t}} \right)^{1-\sigma} E_{j,t}, \tag{8}$$

where $P^s_{ij,t} = \left[\sum_{n \in N^s_i} [p^s_{ij,t}(n)]^{1-\sigma} \right]^{1/(1-\sigma)}$ is the corresponding sticky-variety price index. Finally, the overall price index in destination j in period t can be written as (see Appendix B for proofs):

$$P_{j,t} = \left[\sum_{i \in \mathcal{J}} \left(N^f_i [p^f_{ij,t}]^{1-\sigma} + [P^s_{ij,t}]^{1-\sigma} \right) \right]^{1/(1-\sigma)} . \tag{9}$$

Total expenditure and aggregate profits. Total expenditure in country j on goods produced in country i is given by $X_{ij,t} = X^f_{ij,t} + X^s_{ij,t}$. Substituting equations (7) and (8), we obtain:

$$X_{ij,t} = (1-\alpha) N_i \left(\frac{p^f_{ij,t}}{P_{j,t}} \right)^{1-\sigma} E_{j,t} + \left(\frac{P^s_{ij,t}}{P_{j,t}} \right)^{1-\sigma} E_{j,t}. \tag{10}$$

The corresponding aggregate profits for flexible firms are $\Pi^f_{ij,t} = X^f_{ij,t}/\sigma - FP_{i,t}N^f_i$, and for sticky firms $\Pi^s_{ij,t} = X^s_{ij,t}/\sigma - FP_{i,t}N^s_i$. Total aggregate profits are therefore $\Pi_{ij,t} = X_{ij,t}/\sigma - FP_{i,t}N_i$, with F referring to a fixed cost, denominated in country i's price index, that firms must pay every period to sell in each destination including the domestic market. We assume the same fixed costs F in all markets. By keeping the number of firms in each country fixed, we can set F to zero, implying that all firms active in country i export to each country j.

Balanced trade. We assume balanced trade, and total expenditure equals total output. From the market-clearing condition for goods in each country, we therefore have:

$$(p_{i,t})^{1-\sigma} = E_{i,t} \left\{ \sum_j \left[(1-\alpha) N_i + \sum_{n \in N^s_i} \left(\frac{\kappa^s_{ij,t}(n)}{\kappa^f_{ij,t}} \right)^{1-\sigma} \right] \left(\frac{\kappa^f_{ij,t} t_{ij,t}}{P_{j,t}} \right)^{1-\sigma} E_{j,t} \right\}^{-1} . \tag{11}$$

Finally, total income is given by labor income and aggregate profits: $E_{i,t} = w_{i,t}L_{i,t} + \sum_j \Pi_{ij,t}$, which can be expressed as $E_{i,t} = w_{i,t}L_{i,t} + \sum_j X_{ij,t}/\sigma - \mathfrak{J}FP_{i,t}N_i$, where $\mathfrak{J}$ refers to the total number of countries.

2.3 The gravity model for transition dynamics

As a reference point, we assume the world economy is initially in a steady state at $t = 0$ where all firms set their prices optimally (i.e., $p_{ij,0} = \tilde{p}_{ij,0}$) with trade costs $t_{ij,0}$. We assume the labor endowment and the number of firms are given and fixed at initial values for every subsequent period t. As shown in Appendix C.1, the initial steady state is characterized by a system of equations that corresponds to a standard gravity system.

From the pricing structure described above, the average price of firms from country i in destination market j in period t can be expressed as:

$$\bar{p}_{ij,t} = (1-\alpha)p^{f}_{ij,t} + \alpha p^{s}_{ij,t} \quad \Leftrightarrow \quad \bar{p}_{ij,t} = (1-\alpha)p^{f}_{ij,t} + \alpha \bar{p}_{ij,t-1}, \tag{12}$$

which is a weighted average of the price set by flexible firms in period t and the average price in period $t-1$.

We now consider a bilateral trade cost shock in period $t = 1$. In the transition to period 1, only $N^f_i = (1-\alpha)N_i$ firms can adjust their prices, while $N^s_i = \alpha N_i$ firms cannot adjust and keep their prices from period $t = 0$. Based on equation (10), we obtain:

$$X_{ij,1} = (1-\alpha)N_i \left(\frac{p^{f}_{ij,1}}{P_{j,1}}\right)^{1-\sigma} E_{j,1} + \alpha \frac{\hat{E}_{j,1}}{\hat{P}^{1-\sigma}_{j,1}} X_{ij,0}, \tag{13}$$

where $\hat{z}_t \equiv z_t/z_{t-1}$ represents the ratio of any variable in period t relative to $t-1$ (see Appendix C.2 for proofs). Equation (13) shows that $X_{ij,1}$ is a function of $X_{ij,0}$. Intuitively, the first part represents a standard gravity relationship for flexible firms, and the second part represents trade flows of sticky firms. But this equation is no longer multiplicative, and hence it no longer corresponds to the typical gravity equation familiar from the literature. Note that the special case of $\alpha = 0$ nests the standard gravity framework based on the Krugman (1980) model.

Equation (13) describes the transition dynamics in response to the trade cost shock one period after the initial steady state. However, this structure has several limitations. First, we want to analyze the dynamic path over *many* periods after the shock. Second, in empirical applications, there is never an 'initial period' defined as a pre-shock steady state because the economy is always in a transition path from some past shock. We therefore generalize the transition dynamics for any period t relative to period $t-1$, where period $t-1$ may not represent a steady-state equilibrium. Starting from equation (10) and using a first-order approximation of $P^s_{ij,t}$ around $p^s_{ij,t}$, we can write bilateral trade flows in period t as:

$$X_{ij,t} = (1-\alpha)N_i \left(\frac{p^{f}_{ij,t}}{P_{j,t}}\right)^{1-\sigma} E_{j,t} + \alpha N_i \left(\frac{\bar{p}_{ij,t-1}}{P_{j,t}}\right)^{1-\sigma} E_{j,t} + e_{ij,t}, \tag{14}$$

where $e_{ij,t}$ denotes an approximation error. This approximation error can be interpreted as a distortion that comes from the price heterogeneity across sticky firms from country i selling to destination market j in period t (see Appendix C.3.2 for proofs and an explicit expression for $e_{ij,t}$). The first term in equation (14) represents the component of the trade flow originating from flexible firms, and the second term refers to sticky firms. The higher the share of flexible firms $(1-\alpha)$, the higher the weight of the standard gravity term that does not involve lagged variables.

A drawback of equation (14) is that we do not observe the lagged price $\bar{p}_{ij,t-1}$. To address this problem, we approximate the unobservable term $\left(\bar{p}_{ij,t-1}/P_{j,t}\right)^{1-\sigma}$ with the observable lagged trade flow $X_{ij,t-1}$. Specifically, we use a first-order approximation of the bilateral price index $P_{ij,t-1}$ around the average bilateral price $\bar{p}_{ij,t-1}$ and compute $\left(\bar{p}_{ij,t-1}/P_{j,t-1}\right)^{1-\sigma} \approx X_{ij,t-1}/(N_i E_{j,t-1})$. Substituting into equation (14), we obtain an expression for bilateral trade flows in terms of one-period lagged trade flows:

$$X_{ij,t} \approx (1-\alpha)N_i \left(\frac{p^f_{ij,t}}{P_{j,t}}\right)^{1-\sigma} E_{j,t} + \alpha \frac{\hat{E}_{j,t}}{\hat{P}^{1-\sigma}_{j,t}} X_{ij,t-1}, \tag{15}$$

where the ratio of the variables $\hat{E}_{j,t}$ and $\hat{P}_{j,t}$ now appears in the second term on the right-hand side. See Appendix C.3 for proofs, an extension to higher-order lags, and the characterization of a higher-order lag decay function.

3 Deriving the estimating equation

To transform the model given in equation (15) into an estimating equation, we have to account for the fact that the relationship will not hold exactly in any sample. We therefore introduce two mutually independent multiplicative error terms, $\eta_{ij,t}$ and $\nu_{ij,t}$, as follows:[9]

$$X_{ij,t} = (1-\alpha)N_i \left(\frac{p^f_{ij,t}}{P_{j,t}}\right)^{1-\sigma} E_{j,t}\eta_{ij,t} + \alpha \frac{\hat{E}_{j,t}}{\hat{P}^{1-\sigma}_{j,t}} X_{ij,t-1}\nu_{ij,t}. \tag{16}$$

Assuming $\mathbb{E}[\eta_{ij,t}|\cdot] = 1$ and $\mathbb{E}[\nu_{ij,t}|\cdot] = 1$, and using equation (3), we can write the conditional expectation as:

$$\mathbb{E}\left[X_{ij,t}|\cdot\right] = (1-\alpha)N_i \left(\frac{\kappa^f_{ij,t} p_{i,t} t_{ij,t}}{P_{j,t}}\right)^{1-\sigma} E_{j,t} + \alpha \left(\frac{\hat{E}_{j,t}}{\hat{P}^{1-\sigma}_{j,t}}\right) X_{ij,t-1}. \tag{17}$$

Equation (17) represents a non-linear system. Exporter-specific and importer-specific terms could, in principle, be absorbed by corresponding exporter-time and importer-time fixed effects. But in practice, direct non-linear estimation of equation (17) is not feasible due to the high dimension of fixed effects. We therefore propose an alternative estimating procedure in three steps.

[9]For simplicity, we use an equality sign in equation (16) and the following equations rather than an approximate equality sign.

3.1 3-step estimation procedure

Assuming, for simplicity, small trade cost shocks or one of the four conditions described in section 2.1.1 for $\kappa^f_{ij,t} \to 1$, we obtain the following:

$$\begin{aligned}\mathbb{E}\left[X_{ij,t}|\cdot\right] =&(1-\alpha)N_i\left(\frac{p_{i,t}t_{ij,t}}{P_{j,t}}\right)^{1-\sigma}E_{j,t}+\alpha\frac{\hat{E}_{j,t}}{\hat{P}^{1-\sigma}_{j,t}}X_{ij,t-1}\\ =&(1-\alpha)\exp\left[(1-\sigma)\ln\left(t_{ij,t}\right)+\psi_{i,t}+\chi_{j,t}\right]+\alpha\exp\left[\ln\left(X_{ij,t-1}+\phi_{j,t}\right)\right], \end{aligned} \tag{18}$$

with $\exp(\psi_{i,t}) \equiv N_i p^{1-\sigma}_{i,t}$, $\exp(\chi_{j,t}) \equiv E_{j,t}/P^{1-\sigma}_{j,t}$ and $\exp(\phi_{j,t}) \equiv \hat{E}_{j,t}/\hat{P}^{1-\sigma}_{j,t}$. We now introduce the three steps of our estimation procedure.

Step 1: Estimation of flexible and sticky gravity parts. The conditional expectation $\mathbb{E}\left[X_{ij,t}|\cdot\right]$ can be split into two parts. The first part corresponds to the standard gravity equation with bilateral trade costs and exporter-time and importer-time fixed effects. Consistent with the literature, we define the trade cost function as $t_{ij,t} = dist^{\beta_d}_{ij}\beta^{RTA_{ij,t}}_r$, where $dist_{ij}$ is bilateral distance, and $RTA_{ij,t}$ is a dummy variable taking the value one if a regional trade agreement (RTA) is in place between exporter i and importer j at time t.

When we substitute the expression for $t_{ij,t}$, the flexible part of equation (18) becomes:

$$\mathbb{E}\left[X^f_{ij,t}|\cdot\right] = (1-\alpha)\exp\left(\beta_1 RTA_{ij,t}+\psi_{i,t}+\chi_{j,t}+\mu_{ij}\right), \tag{19}$$

where we also introduce bilateral, time-invariant country-pair fixed effects μ_{ij} to capture all observable and unobservable bilateral time-invariant trade cost components as suggested by Baier and Bergstrand (2007). Since trade flows of flexible firms $X^f_{ij,t}$ are unobserved, we cannot estimate this equation. However, by using the approximation $X^f_{ij,t} \approx (1-\alpha)X_{ij,t}$ (i.e., by assuming a small trade cost shock), we can express the conditional expectation in terms of observed total trade flows as:

$$\mathbb{E}\left[X_{ij,t}|\cdot\right] = \exp\left(\beta^f_1 RTA_{ij,t}+\psi_{i,t}+\chi_{j,t}+\mu_{ij}\right), \tag{20}$$

and we denote the fitted values based on these estimates as $\widehat{X}^f_{ij,t}$. Note that β^f_1 is generally not equal to β_1, as the latter could only be recovered if we observed trade flows from flexible firms.[10]

The second part corresponds to the sticky gravity component. We have:

$$\mathbb{E}\left[X^s_{ij,t}|\cdot\right] = \alpha\exp\left(\gamma_1\ln\left(X_{ij,t-1}\right)+\phi_{j,t}\right), \tag{21}$$

where γ_1 is the coefficient of lagged trade flows. As above, using the approximation $X^s_{ij,t} \approx \alpha X_{ij,t}$ for trade flows of sticky firms to express the conditional expectation in terms of observed total trade

[10]For further details on β^f_1, see section 3.2.

flows, we estimate:

$$\mathbb{E}\left[X_{ij,t}|\cdot\right] = \exp\left(\gamma_1 \ln\left(X_{ij,t-1}\right) + \phi_{j,t}\right). \tag{22}$$

Note that even though the lag of trade flows appears as a regressor on the right-hand side, specification (22) does not suffer from Nickell bias under standard dynamic panel assumptions due to our fixed effects structure (we provide details in Appendix D). We denote the fitted values based on these estimates as $\widehat{X}^s_{ij,t}$. Our model predicts $\gamma_1 = 1$.

Step 2: Estimation of α. We find the value of α that minimizes the mean squared prediction error:

$$\widehat{\alpha} \equiv \underset{\alpha\in[0,1)}{\operatorname{argmin}} \, MSE(\alpha) = \frac{1}{\#obs}\sum_{i,j,t}\left[X_{ij,t} - \left((1-\alpha)\widehat{X}^f_{ij,t} + \alpha\widehat{X}^s_{ij,t}\right)\right]^2. \tag{23}$$

We run a grid search. As α is bounded between zero and one, we calculate $MSE(\alpha)$ for different values of α between 0 and 1, and then search for $\widehat{\alpha}$ that corresponds to the minimum $MSE(\alpha)$.

Step 3: Recovery of steady-state parameters. We implement a non-linear estimation on the following expression:

$$X_{ij,t} = (1-\widehat{\alpha})\exp\left[\beta_1 RTA_{ij,t} + \widehat{\psi}_{i,t} + \widehat{\chi}_{j,t} + \widehat{\mu}_{ij}\right] + \widehat{\alpha}\exp\left[\ln(X_{ij,t-1}) + \widehat{\phi}_{j,t}\right], \tag{24}$$

where we plug in the estimates of the fixed effects $\widehat{\psi}_{i,t}$, $\widehat{\chi}_{j,t}$, $\widehat{\mu}_{ij}$ and $\widehat{\phi}_{j,t}$ from step 1, and the estimated $\widehat{\alpha}$ from step 2. Non-linear estimation then produces an estimate $\widehat{\beta}_1$ for the steady-state parameter $\beta_1 \equiv (1-\sigma)\ln(\beta_r)$.

3.2 Estimation bias in the standard gravity specification

What happens when we estimate the gravity equation assuming fully flexible prices ($\alpha = 0$) and hence omitting the second part of specification (17) that represents trade flows of sticky firms? We compute an expression equivalent to equation (17):

$$\mathbb{E}\left[X_{ij,t}|\cdot\right] = \left(t_{ij,t}\right)^{1-\sigma}\exp(\psi_{i,t} + \chi_{j,t})s_{ij,t}, \tag{25}$$

$$\text{with } s_{ij,t} \equiv (1-\alpha)(\kappa^f_{ij,t})^{1-\sigma} + \alpha\left(\frac{t_{ij,t-1}}{t_{ij,t}}\right)^{1-\sigma}\left(\frac{p_{i,t-1}}{p_{i,t}}\right)^{1-\sigma}\left[(1-\alpha)\kappa^f_{ij,t-1} + \alpha\kappa^s_{ij,t-1}\right]^{1-\sigma},$$

where $s_{ij,t}$ is defined as the *sluggish adjustment factor*. It represents an adjustment factor to the standard gravity specification resulting from the sluggish price adjustment. For $s_{ij,t} = 1$, the system is in a steady state, and therefore the model is defined by the standard gravity specification. However, in the transition phase the standard gravity specification suffers from an omitted variable bias since $s_{ij,t}$ is correlated with $t_{ij,t}$. We briefly discuss how standard gravity estimates are affected when the sluggish adjustment factor is omitted. We refer to Appendix C.4 for details and proofs.

Consider for simplicity the log-linearized version of specification (25) given by $\ln X_{ij,t} = (1-\sigma)\ln t_{ij,t} + \psi_{i,t} + \chi_{j,t} + \ln s_{ij,t}$. A standard gravity specification omits $\ln s_{ij,t}$ and therefore leads to a bias in the estimation of the trade cost elasticity. This omitted variable bias is given by $\mathbb{E}[\beta_1^f] = \beta_1 + \beta_2 \text{cov}\left[\ln t_{ij,t}, \ln s_{ij,t}\right] / \text{var}\left[\ln t_{ij,t}\right]$, where β_1^f refers to the expected value of the log trade cost coefficient in the standard gravity specification, $\beta_1 = 1-\sigma < 0$ refers to the true value of the log trade cost coefficient, and $\beta_2 = 1$ refers to the true value of the coefficient related to $\ln s_{ij,t}$. Hence, considering that $\text{cov}\left[\ln t_{ij,t}, \ln s_{ij,t}\right] > 0$, $\beta_1 = 1-\sigma$ and $\beta_2 = 1$, the standard specification leads to an upward bias in the estimated coefficient, i.e., $\mathbb{E}[\beta_1^f] > \beta_1$ (the coefficient is less negative). This implies that when we use transition phase data, the standard model estimates a lower increase in bilateral trade flows in response to a reduction in $t_{ij,t}$. Vice versa, it underestimates the reduction in bilateral trade flows in response to an increase in $t_{ij,t}$ (see Appendix C.4.2 for proofs).[11]

In other words, we find that trade cost coefficient estimates are biased towards zero in the transition phase. This is consistent with sluggish price adjustment after trade cost shocks as not all firms immediately react to shocks.

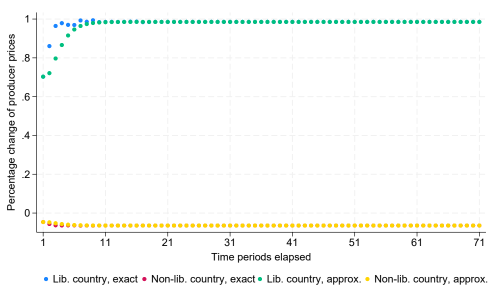
In section 4, we show how the estimates of the standard gravity model vary depending on the time elapsed since the shock, and how they converge to the steady-state elasticities using simulated data. Our model provides a structural relationship between the transition and steady-state elasticities. This allows us to recover the steady-state elasticity even when the economy is in a transition phase.

4 Implementation and simulation

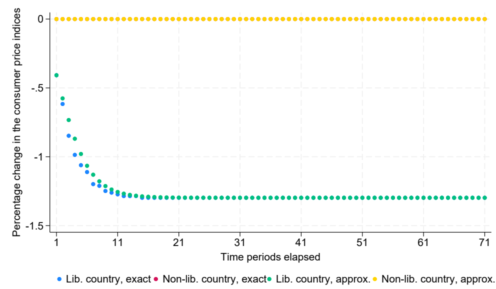
To demonstrate the mechanics of our model, we simulate data based on the following calibration: the number of countries is equal to 30 ($\mathfrak{J} = 30$), the number of firms is set to 100 ($N_i = 100$) for all countries, labor endowments for all countries are set to $L_i = 3000$, trade costs are set to $t_{ij,t} = 1.3$ for international trade flows (i.e., $i \neq j$) and to 1 for domestic flows (i.e., $i = j$) for all countries in the baseline. We set $\sigma = 5$, fixed costs to zero ($F = 0$), and we simulate for 71 periods. Hence, we have a completely symmetric setup in the baseline. The trade cost shock consists of a change in trade costs to 1 (i.e., full liberalization) between countries 1 and 2 in period 2, while all other trade costs remain constant. Only the share $(1-\alpha)$ of firms can adjust consumer prices, with α set to 0.7. Given our symmetric setup, in Figure 1 we show results for one liberalizing country (country 1 or 2) and one non-liberalizing country only.

Panel (a) of Figure 1 shows average producer prices. Average producer prices increase for liberalizing countries and decrease slightly for non-liberalizing countries. We implement two versions of our model: one where we simulate and trace price changes over time for each firm as described in our model (labeled "exact"), and one in which we approximate the price index of all sluggishly adjusting firms by the average price from the previous period (labeled "approximation").

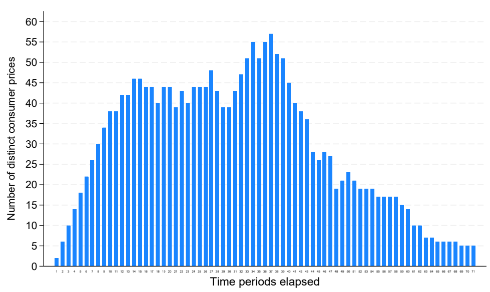
[11]We analyze the structure of the bias under different specifications of trade costs in Appendix C.4.3. In particular, we focus on the structure of the bias for the RTA elasticity when $t_{ij,t} = dist_{ij}^{\beta_d} \beta_r^{RTA_{ij,t}}$, which we later use in our empirical analysis in section 5.



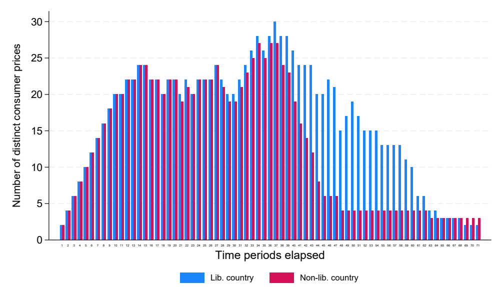
(a) Average producer prices



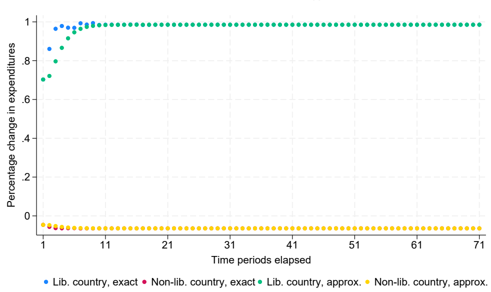
(b) Consumer price indices



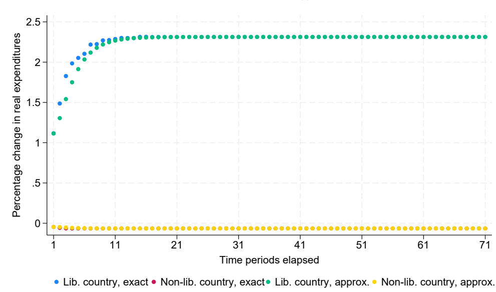
(c) Number of consumer prices (world)



(d) Number of consumer prices (countries)



(e) Expenditures



(f) Real expenditures

Figure 1: Transition dynamics in simulated data (see text for details)

The approximation has the advantage that we do not need to trace prices of each firm, and it is consistent with the use of aggregate data in our empirical application. As seen in Figure 1, the two solutions are very close. For the exact solution, we plot the number of distinct consumer prices for the whole world in panel (c), and for one liberalizing country and one non-liberalizing country in panel (d). We see that at the start, there are only two distinct consumer prices in the baseline steady state, but during the transition we see about 60 different consumer prices. Consumer prices in non-liberalizing countries converge faster than in liberalizing countries because of smaller effects in the former. In the transition, firms adjust markups when they cannot adjust consumer prices. Hence, producer prices show a similar pattern. The main difference is that each country has only one producer price in the initial and new steady states. The results are very similar when we simulate a scenario with 1,000 firms per country.

Panel (b) of Figure 1 shows consumer price indices, which decrease in the liberalizing countries and are hardly affected in the non-liberalizing countries. Together, increasing producer prices and falling consumer price indices in liberalizing countries reflect gains from trade. Panels (e) and (f) reinforce this message, showing increasing expenditures and real expenditures. The latter may be viewed as a measure of welfare. Overall, the results of this simulation illustrate the sluggish adjustment introduced by our sticky gravity model.

To demonstrate the validity of our suggested estimation approach, we now apply our estimation approach to the simulated data for different sample sizes (i.e., we include in the sample an increasing number of periods after the initial shock). In all specifications, the standard gravity estimation leads to an upward bias in the trade cost elasticity estimates (i.e., closer to zero). This bias fades as we include more periods and move farther away from the shock. We consistently recover our stickiness parameter and the steady-state trade cost elasticity close to their true values. We refer to Appendix E.1 for details, and in Appendix E.3 we illustrate how the standard gravity estimated elasticity converges to the true steady-state elasticity as time elapses since the initial trade cost shock.

To gain further insights, we next focus our analysis on simulated data with an explicit functional form for bilateral trade costs based on observable variables. This defines our baseline setting. The main difference from the previous scenario is that we now define $t_{ij,t} = dist_{ij}^{\beta_d} \beta_r^{RTA_{ij,t}}$ with $\beta_d = 0.25$ and $\beta_r = 0.9$. Distance $dist_{ij}$ is symmetric for each bilateral pair ij ($i \neq j$) and is drawn randomly between 2.5 and 5, whereas for domestic observations ($i = j$) it is drawn randomly between 1 and 2. We take a random draw of new RTA dummies in period 1, which leads to a share of 0.233 of bilateral relationships trading under an RTA. For further details on the setting, see Appendix E.2.

Table 1 reports the results of applying our procedure to the simulated data with this explicit functional form for bilateral trade costs. We provide estimates for samples with an increasing number of time periods since the RTA shock (10, 20 and 71 periods). In the top panel of the table, we estimate the two regressions in step 1 (flexible and sticky regressions) with PPML. See Appendix E.2 for alternative specifications. Columns (1), (3), and (5) report estimates for the short-run trade elasticity of an RTA from the standard (flexible) gravity specification. The table shows that these values increase in the number of periods included since the shock, implying that transition-phase

Table 1: Regression results

	Step 1					
	Sample: 10 periods		**Sample: 20 periods**		**Sample: 71 periods**	
	(1)	(2)	(3)	(4)	(5)	(6)
	Flex Reg	Sticky Reg	Flex Reg	Sticky Reg	Flex Reg	Sticky Reg
$RTA_{ij,t}$	0.328^{***} (0.00145)		0.374^{***} (0.000886)		0.408^{***} (0.000279)	
$\ln X_{ij,t-1}$		0.999^{***} (0.000612)		1.000^{***} (0.000326)		1.000^{***} (0.0000951)
FEs	it, jt, ij	jt	it, jt, ij	jt	it, jt, ij	jt
	Step 2					
Stickiness degree: $\widehat{\alpha}$	0.65		0.69		0.71	
	Step 3					
Est. Coef: $\widehat{\beta}_1$	0.371^{***} (0.000447)		0.401^{***} (0.000167)		0.417^{***} (0.0000280)	
Observations	9,900		18,900		64,800	

Notes: True coefficient $\beta_1 = (1-\sigma)\ln(\beta_r) = 0.421$. *Sample* refers to the number of periods after shock included. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

elasticities converge to the steady-state elasticity as the economy approaches the new steady state (the true value of the steady-state elasticity is 0.421). Columns (2), (4), and (6) show that the lagged trade coefficient is close or equal to unity in all specifications, which is consistent with our theoretical model.

Regarding step 2, the middle panel of Table 1 shows estimates of the α parameter close to the true value of 0.7, improving as we increase the number of periods since the shock. Regarding step 3, the bottom panel of the table shows that our approach delivers estimates of the latent steady-state RTA coefficient close to the true value of 0.421, with increasingly better results when we include more periods.

5 Empirical evidence

Before implementing our 3-step estimation procedure to analyze the effects of RTAs in section 5.2, we investigate how our estimates of α are affected by the frequency of the data (e.g., monthly, quarterly, half-yearly).

5.1 Analysis of stickiness estimates across different data frequencies

While most studies in the empirical trade literature rely on annual data, we briefly show how the estimated degree of price stickiness varies when using monthly, quarterly and half-yearly data. See

Appendix A.1 for data description.[12] At a frequency higher than yearly, we have a maximum of 44 quarters (2010Q1 to 2020Q4) and a maximum of 22 half-years (2010H1 to 2020H2) in our samples. We perform steps 1 and 2 of our procedure for each data frequency (see Table 2).

Table 2: Flexible-price and sticky-price gravity at different frequencies

	Step 1 (PPML)					
	Flexible	Sticky	Flexible	Sticky	Flexible	Sticky
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(dist_{ij})$	-0.8936***		-0.8937***		-0.8933***	
	(0.004)		(0.006)		(0.009)	
$\ln(X_{ij,t-1})$		0.9905***		0.9932***		0.9917***
		(0.001)		(0.001)		(0.001)
Frequency	Monthly	Monthly	Quarterly	Quarterly	Half-yearly	Half-yearly
FEs	it, jt	jt	it, jt	jt	it, jt	jt
Observations	163,104	163,104	53,097	53,097	25,696	25,696
Pseudo-R^2	0.922	0.992	0.924	0.995	0.924	0.995
	Step 2					
$\widehat{\alpha}$	0.95		0.90		0.84	

Notes: The dependent variable is $X_{ij,t}$, at different frequencies. PPML estimation. Standard errors are reported in parentheses. Columns (1), (3), and (5) are the flexible-price gravity specifications. Columns (2), (4), and (6) are the sticky-price gravity specifications. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The most relevant result is the comparison of stickiness estimates ($\widehat{\alpha}$'s) across frequencies in the bottom panel of Table 2. The estimates are $\hat{\alpha} = 0.95, 0.90, 0.84$ for monthly, quarterly and half-yearly frequencies, respectively. Two observations are noteworthy. First, the pattern in the estimates suggests a high degree of price stickiness at all frequencies. Second, as expected, stickiness decreases as we reduce the frequency of the data since firms have more time to adjust prices. In other words, prices become more flexible over longer horizons.

5.2 Application to specific RTAs

We apply our 3-step procedure described in Section 3.1 to different RTAs as follows. In step 1, we first estimate the standard gravity specification:

$$X_{ij,t} = \exp\left\{\beta_1^f RTA_{ij,t}^{specific} + \beta_2^f RTA_{ij,t}^{other} + \psi_{i,t} + \chi_{j,t} + \mu_{ij}\right\}, \tag{26}$$

where $RTA_{ij,t}^{specific}$ is a dummy variable that identifies the specific RTA under consideration, while $RTA_{ij,t}^{other}$ is a dummy variable that controls for all other RTAs in place. The data on RTAs come from Mario Larch's Regional Trade Agreements Database from Egger and Larch (2008), updated

[12]We drop monthly observations when adjacent observations are missing to construct a full quarter; for example, if January and February flows are available but March flows are missing, then we drop the former from the sample. We proceed analogously for aggregation to half-yearly and annual frequency.

version 20241028. As RTAs apply to all countries in the world, not only OECD countries, we use the ITPD-E-R02 of the USITC (Borchert, Larch, Shikher, and Yotov, 2021, 2022) as our trade flow data.[13] We complete step 1 by estimating the sticky gravity part:

$$X_{ij,t} = \exp\left\{\gamma_1 \ln\left(X_{ij,t-1}\right) + \phi_{j,t}\right\}. \tag{27}$$

Step 2 of our procedure recovers the parameter α by minimizing the mean squared prediction error as described in equation 23.

In step 3 we recover the respective steady-state parameters for $RTA_{ij,t}^{specific}$ and $RTA_{ij,t}^{other}$. Specifically, we solve the following minimization problem:

$$\min_{\beta_1,\beta_2}\left\{X_{ij,t} - (1-\hat{\alpha})\exp\left[\beta_1 RTA_{ij,t}^{specific} + \beta_2 RTA_{ij,t}^{other} + \hat{\psi}\right] - \hat{\alpha}\exp\left[\ln(X_{ij,t-1}) + \hat{\phi}_{j,t}\right]\right\}^2, \tag{28}$$

with $\hat{\psi} \equiv \hat{\psi}_{i,t} + \hat{\chi}_{j,t} + \hat{\mu}_{ij}$ denoting the estimated fixed effects of the standard gravity estimation in step 1, $\hat{\phi}_{j,t}$ representing the estimated fixed effects of the sticky gravity part in step 1, and $\hat{\alpha}$ is the estimated coefficient from step 2.

Before showing the results, i.e., identification of regularities or patterns in the estimated coefficients over the whole universe of RTAs included in the dataset, we discuss the estimated steady-state elasticities in more detail for four examples (see Table 3).

Table 3: Estimated short-run and long-run elasticities of RTAs: selected examples

Agreement	**RTA elasticities**			Elasticity ratio
	(1) Transition (T)	(2) Steady-State (S)	(3) S-T	(4) Ratio=S/T
Andean Community	0.62	0.64	0.02	1.04
EEA	0.11	0.15	0.04	1.38
NAFTA	0.14	0.15	0.01	1.06
NZL-SGP	0.24	0.31	0.07	1.30

Notes: The Andean Community includes Bolivia, Colombia, Ecuador and Peru and was concluded in 1988. The EEA refers to the European Economic Area Agreement, which began in 1994. NAFTA refers to the North American Free Trade Agreement between the USA, Canada and Mexico, which entered into force in 1994. NZL-SGP refers to the New Zealand-Singapore Agreement, which began in 2001. Column (1) reports the standard gravity estimated RTA coefficients for the specific agreements (i.e., $\widehat{\beta}_1^f RTA_{ij,t}^{specific}$), also referred to as transition-phase (T) elasticity. Column (2) reports the recovered coefficient from step 3 (i.e., $\widehat{\beta}_1 RTA_{ij,t}^{specific}$), also referred to as steady-state (S) elasticity. Column (3) reports the absolute difference between steady-state and transition-phase elasticities. Column (4) reports the ratio of elasticities (S/T).

Table 3 shows that the steady-state elasticity is larger than the transition elasticity for all four examples. But magnitudes vary. Comparing the entry into force of the agreements and the ratio of the steady-state and transition elasticities, we find that older agreements, such as the Andean Community and NAFTA, have lower ratios than newer ones, such as the agreement between New

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

Zealand and Singapore. This is consistent with our theory, which predicts that older agreements should be closer to their steady state, and therefore the adjustment implied by computing the latent steady-state elasticity in step 3 should be smaller. However, the EEA Agreement has a relatively large ratio even though it also started back in 1994. See Appendix F.1 for additional evidence on the elasticity ratio estimated for the EEA.

5.3 Summary of results using all RTAs

Table 4 summarizes the results using all RTAs. The first row of the table indicates that there are 438 RTAs in the sample, broken down into 27 customs unions (CU), 387 free trade agreements (FTA), and 24 partial scope agreements (PSA). Columns (5)-(8) show the fraction of elasticity ratios that exceed the value of 1, i.e., the fraction of ratios where an upward adjustment occurs so that the steady-state elasticity is larger than the transition elasticity (see Table 3). This upward adjustment occurs jn only 48% of all RTAs but in 56% of customs unions. This may suggest that the depth of agreement is important.

In the subsequent rows of the table, we therefore distinguish results by various conditions related to the obtained estimates, recognizing that many individual RTA coefficients are not significant or negative, contrary to the typical expectation. In particular, we report results for the subsample of agreements where we obtain positive first-step estimates (225), positive and significant first-step estimates (148), positive first- and third-step estimates (211), and positive and significant first- and third-step estimates (95), where we apply a significance level of 10%. For example, when both coefficients are positive and statistically significant (see the last row of Table 4), we obtain an average upward adjustment for 66% of all RTAs.

Table 4: All RTAs and elasticity ratios

	Number of RTAs				Share of elasticity ratio>1			
Condition on estimates	(1) All	(2) CU	(3) FTA	(4) PSA	(5) All	(6) CU	(7) FTA	(8) PSA
All	438	27	387	24	0.48	0.56	0.47	0.54
Positive first step	225	18	194	13	0.52	0.61	0.50	0.69
Positive and significant first step	148	14	123	11	0.53	0.57	0.51	0.64
Positive first and third steps	211	18	180	13	0.55	0.61	0.54	0.69
Positive and significant first and third steps	95	11	76	8	0.66	0.64	0.66	0.75

Notes: Columns (1)–(4) report the number of agreements by category. Columns (5)–(8) report the corresponding shares of elasticity ratios that exceed 1. Row conditions: "All" is based on all obtained estimates, "Positive first step" is based on the subsample where the first-step estimates are positive, "Positive and significant first step" is based on the subsample where the first-step estimates are positive and significant at the 10% level, "Positive first and third steps" is based on the subsample where the first- and third-step estimates are positive, and "Positive and significant first and third steps" is based on the subsample where the first- and third-step estimates are positive and significant at the 10% level.

We restrict our further analysis to FTAs, as this is the most common type of agreement in our sample. Our theory suggests that the adjustment will be proportional to the tariff reduction introduced by the agreement. We therefore compute a variable *dDIFF*. It denotes the average absolute reduction introduced by preferential tariffs relative to the MFN tariff among FTA members over the sample period since the entry into force of an agreement. We calculate *dDIFF* for each RTA. In other words, *dDIFF* represents the average preferential tariff reduction introduced by each RTA, where a larger tariff reduction represents a more negative *dDIFF* (i.e., a higher *dDIFF* in absolute value). We expect that a higher average tariff reduction induced by an FTA leads to a larger adjustment of the recovered long-run FTA effect in step 3 (i.e., a higher elasticity ratio). Moreover, theory predicts that the more recent the agreement, the higher the elasticity ratio, since economies are more likely to still be in the transition phase.

Focusing on the subsample of positive and significant first-step estimates as in the third row of Table 4 (123 FTAs), we plot in Figure 2 the correlation between the elasticity ratio and the average tariff reduction of FTAs as represented by *dDIFF*. In line with theory, we find a negative correlation. That is, larger tariff reductions are associated with a larger elasticity ratio and hence a larger upward adjustment. See Appendix F.2 for further evidence on the relationship between tariff reductions and the elasticity ratio.

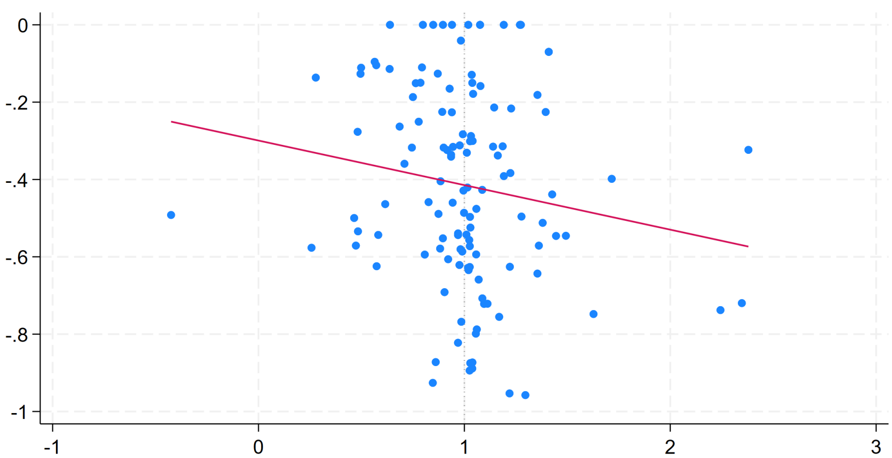


Figure 2: *dDIFF* vs. elasticity ratio. Subsample of positive and significant first-step estimates. The vertical axis shows *dDIFF* and the horizontal axis shows the elasticity ratio.

6 Conclusion

This paper develops and empirically implements a dynamic gravity framework that accounts for persistence in international trade flows. Unlike the static gravity model, our approach introduces sluggish price adjustment at the bilateral level, akin to Calvo pricing. This provides a micro-founded explanation for why lagged trade flows matter. We show that ignoring persistence leads to biased estimates of trade elasticities and mismeasurement of policy effects. By combining theory, simulations and empirical evidence, we demonstrate that the dynamic adjustment process is crucial to understanding the propagation of trade cost shocks over time.

To translate our theory into an estimating procedure, we propose three steps. First, we separately estimate the flexible-price and sticky-price components of the gravity equation. This allows us to identify transition-phase elasticities, and it underlines the importance of lagged trade flows. Second, we recover the stickiness parameter, which measures the share of firms unable to flexibly adjust prices. Third, we use these estimates to infer steady-state trade cost elasticities consistent with the underlying theory. This approach allows us to bridge the gap between transition-phase and steady-state effects even when the economy is still in transition.

Using OECD and ITPD-E-R02 trade data and applying the procedure to RTAs, we find strong evidence of price stickiness and persistence in trade flows. At monthly frequency, more than 90% of firms appear unable to adjust prices, with stickiness declining as the data frequency decreases. Consequently, standard gravity estimates understate the long-run impact of RTAs for a large share of RTAs: transition elasticities are systematically smaller than their steady-state counterparts, with older agreements, such as the Andean Community and NAFTA, showing smaller deviations than newer ones, such as the agreement between New Zealand and Singapore. In line with theory, we find that larger tariff reductions lead to larger differences between transition and steady-state elasticities. Our results imply that trade policy evaluations based on static models underestimate the eventual gains from integration, highlighting the importance of accounting for dynamic adjustment when quantifying the effects of trade policy.

References

ALESSANDRIA, G., C. ARKOLAKIS, AND K. J. RUHL (2021): "Firm Dynamics and Trade," *Annual Review of Economics*, 13, 253–280.

ANDERSON, J. E., M. LARCH, AND Y. V. YOTOV (2020): "Transitional Growth and Trade with Frictions: A Structural Estimation Framework," *Economic Journal*, 130(630), 1583–1607.

ANDERSON, J. E., AND E. VAN WINCOOP (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93(1), 170–192.

ANDERSON, J. E., AND Y. V. YOTOV (2020): "Short Run Gravity," *Journal of International Economics*, 126, 103341.

——— (2022): "Estimating Gravity from the Short to the Long Run: A Simple Solution to the 'International Elasticity Puzzle'," *CESifo Working Paper No. 10176*.

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

AUER, R., A. BURSTEIN, AND S. M. LEIN (2021): "Exchange Rates and Prices: Evidence from the 2015 Swiss Franc Appreciation," *American Economic Review*, 111(2), 652–686.

BAIER, S. L., AND J. H. BERGSTRAND (2007): "Do Free Trade Agreements Actually Increase Members' International Trade?," *Journal of International Economics*, 71(1), 72–95.

BOEHM, C. E., A. A. LEVCHENKO, AND N. PANDALAI-NAYAR (2023): "The Long and Short (Run) of Trade Elasticities," *American Economic Review*, 113(4), 861–905.

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

——— (2022): "The International Trade and Production Database for Estimation - Release 2 (ITPD-E-R02)," *USITC Working Paper 2022-07-A.*

Burstein, A., and G. Gopinath (2014): "International Prices and Exchange Rates," in *Handbook of International Economics*, vol. 4, pp. 391–451. Elsevier.

Caliendo, L., M. Dvorkin, and F. Parro (2019): "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 87(3), 741–835.

Calvo, G. A. (1983): "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, 12(3), 383–398.

Cameron, A., and P. Trivedi (2005): *Microeconometrics - Methods and Applications*. Cambridge University Press, Cambridge, United Kingdom.

Campbell, D. L. (2010): "History, Culture, and Trade: A Dynamic Gravity Approach," *EERI Research Paper Series No. 26/2010*.

Carballo, J., K. Handley, and N. Limão (2022): "Economic and Policy Uncertainty: Aggregate Export Dynamics and the Value of Agreements," *Journal of International Economics*, 139, 103661.

Chaney, T. (2014): "The Network Structure of International Trade," *American Economic Review*, 104(11), 3600–3634.

Cheng, I.-H., and H. J. Wall (2005): "Controlling for Heterogeneity in Gravity Models of Trade and Integration," *Federal Reserve Bank of St. Louis Review*, 87(1), 49–63.

de Nardis, S., R. De Santis, and C. Vicarelli (2008): "The Euro's Effects on Trade in a Dynamic Setting," *European Journal of Comparative Economics*, 5(1), 73–85.

Eaton, J., S. Kortum, B. Neiman, and J. Romalis (2016): "Trade and the Global Recession," *American Economic Review*, 106(11), 3401–3438.

Egger, P., R. Foellmi, U. Schetter, and D. Torun (2025): "Gravity with History: On Incumbency Effects in International Trade," *Journal of the European Economic Association*, 23(4), 1350–1396.

Egger, P., and M. Larch (2008): "Interdependent Preferential Trade Agreement Memberships: An Empirical Analysis," *Journal of International Economics*, 76(2), 384–399.

Eichengreen, B., and D. A. Irwin (1998): "The Role of History in Bilateral Trade Flows," in *The Regionalization of the World Economy*, pp. 33–62. University of Chicago Press.

Gopinath, G., and R. Rigobon (2008): "Sticky Borders," *Quarterly Journal of Economics*, 123(2), 531–575.

Jung, B. (2012): "Gradualism and Dynamic Trade Adjustment: Revisiting the Pro-Trade Effect of Free Trade Agreements," *Economics Letters*, 115(1), 63–66.

Krugman, P. R. (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70(5), 950–959.

Martínez-Zarzoso, I., D. F. Nowak-Lehmann, and N. Horsewood (2009): "Are Regional Trading Agreements Beneficial? Static and Dynamic Panel Gravity Models," *North American Journal of Economics and Finance*, 20(1), 46–65.

Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49(6), 1417–1426.

Olivero, M. P., and Y. V. Yotov (2012): "Dynamic Gravity: Endogenous Country Size and Asset Accumulation," *Canadian Journal of Economics*, 45(1), 64–92.

Rauch, J. E. (1999): "Networks versus Markets in International Trade," *Journal of International Economics*, 48(1), 7–35.

Rodríguez-Clare, A., M. Ulate, and J. P. Vásquez (2026): "Trade with Nominal Rigidities: Understanding the Unemployment and Welfare Effects of the China Shock," *Journal of Political Economy*, 134(2), 626–664.

APPENDIX

A Reduced-form empirical evidence

We first briefly describe the data in section A.1 and then introduce the reduced-form evidence on persistence in section A.2.

A.1 Data description

This appendix provides further details on the trade data. We use bilateral monthly import data for 37 OECD countries (i.e., all OECD countries as of 2023, except Belgium, for which data are not available). Excluding domestic trade, we have $37 \times 36 = 1332$ country pairs. Although some earlier observations are available for some countries, for most reporting countries, complete data are available only from 2010 onward. We therefore drop years before 2010. Our sample ends in 2020, comprising 11 years with 12 months, i.e., 132 periods.

We therefore have a theoretical maximum of $37 \times 36 \times 132 = 175,824$ bilateral trade flow observations. Due to missing observations, our sample comprises only 164,803 observations. The fraction of the maximum total available is, therefore, $164,803/175,824 = 0.937$. No observation is reported as a zero trade flow.

In the following, we list the number of trade flow observations by importing country in our sample. As explained above, there are 132 periods and 36 partners per country. Thus, we have a maximum of $132 \times 36 = 4752$ monthly trade observations per country.

- Australia 4644 (observations missing for 2010)
- Austria 3255 (no observations for 2018-2020, some missing for 2014 and 2017)
- Canada 4752
- Chile 4356 (some observations missing for 2020)
- Colombia 2724 (missing observations/patchy for 2010-2013, 2015-2016)
- Costa Rica 180 (only observations for 2018 and 2020)
- Czech Republic 4714 (some observations missing for 2017)
- Denmark 4752
- Estonia 4662 (some observations missing for 2010-2013, 2016, 2019)
- Finland 4752
- France 4320 (observations missing for 2020)
- Germany 4752
- Greece 4751 (one observation missing for 2015)
- Hungary 4752
- Iceland 4752
- Ireland 4752
- Israel 4716 (some observations missing for 2013)
- Italy 4752
- Japan 4752
- Latvia 4720 (a few observations missing, especially for 2010)
- Lithuania 4746 (some observations missing for 2010-2011)
- Luxembourg 4702 (some observations missing for 2010-2012)
- Mexico 4752
- Netherlands 4752
- New Zealand 4752

- Norway 4752
- Poland 4752
- Portugal 4745 (a few observations missing for 2016)
- Slovakia 4752
- Slovenia 4752
- South Korea 3024 (no observations for 2010-2012, 2020)
- Spain 4752
- Sweden 4752
- Switzerland 4752
- Turkey 4752
- United Kingdom 4752
- United States of America 4752

A.2 Reduced-form evidence for persistence

We use publicly available trade data from UN Comtrade, including OECD members (37 countries) and their respective bilateral imports at a monthly frequency from January 2010 to December 2020.[14] The data are at the aggregate country-level in nominal USD. In total, we have over 160,000 trade flow observations that we combine with gravity variables from the CEPII database, in particular, bilateral distance, a contiguity dummy, and a common language dummy. Appendix A.1 provides more details on the data.

To examine the persistence of gravity patterns, we explore the role of lagged trade flows. We run 'dynamic' gravity regressions where, apart from exporter-time fixed effects (it), importer-time fixed effects (jt), and country-pair fixed effects (ij), we include lagged trade flows as additional regressors According to the standard gravity model, the lagged regressor should not have any explanatory power. However, significant estimates of lagged trade flows may hint at dynamic effects.

What explains this persistence? There must be factors that vary bilaterally over time; that is, bilateral time-varying factors such as stickiness in consumer prices. It cannot be factors that only vary unilaterally by exporter and importer and over time, such as lagged i and lagged j variables, because these factors are captured by the respective fixed effects. Thus, we interpret the significance of lagged trade flows as evidence of dynamics at the bilateral level. Moreover, the table shows that the persistence monotonically declines towards zero as we increase the order of the lags.

Table A1 provides results showing a significant persistence in bilateral trade flows after controlling for the standard gravity specification.

[14]In 2023, 38 countries are members of the OECD; however, data for Belgium are missing, so we have a sample of 37 countries.

Table A1: Gravity with Lagged Dependent Variables

	(1)	(2)	(3)	(4)
$\ln(\text{trade}_{ij,t-1})$	0.5103***	0.3237***	0.2969***	0.2833***
	(0.021)	(0.011)	(0.012)	(0.012)
$\ln(\text{trade}_{ij,t-2})$		0.2044***	0.1666***	0.1606***
		(0.010)	(0.009)	(0.009)
$\ln(\text{trade}_{ij,t-3})$		0.1628***	0.1068***	0.0990***
		(0.008)	(0.008)	(0.008)
$\ln(\text{trade}_{ij,t-4})$			0.0724***	0.0546***
			(0.007)	(0.008)
$\ln(\text{trade}_{ij,t-5})$			0.0443***	0.0282***
			(0.007)	(0.007)
$\ln(\text{trade}_{ij,t-6})$			0.0597***	0.0297***
			(0.006)	(0.006)
$\ln(\text{trade}_{ij,t-7})$				0.0157*
				(0.009)
$\ln(\text{trade}_{ij,t-8})$				0.0111
				(0.008)
$\ln(\text{trade}_{ij,t-9})$				0.0301***
				(0.008)
$\ln(\text{trade}_{ij,t-10})$				0.0059
				(0.006)
$\ln(\text{trade}_{ij,t-11})$				0.0240***
				(0.007)
$\ln(\text{trade}_{ij,t-12})$				0.0427***
				(0.008)
FEs	it, jt, ij	it, jt, ij	it, jt, ij	it, jt, ij
Observations	163,104	159,778	155,213	146,133
R^2	0.975	0.978	0.979	0.979

Notes: The dependent variable is logarithmic trade, $\ln(\text{trade}_{ij,t})$, at monthly frequency. OLS estimation. Robust standard errors are reported in parentheses, clustered by country pairs. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B General model setup

The preferences of the representative consumer in any country j are defined in section 2. Distinguishing varieties produced by flexible (price-adjusting) firms from those produced by sticky firms, we can write the preferences as follows:

$$U_{j,t} = \left\{ \sum_i \left[\sum_{n \in N_i^f} \left[q_{ij,t}^f(n) \right]^{(\sigma-1)/\sigma} + \sum_{n \in N_i^s} \left[q_{ij,t}^s(n) \right]^{(\sigma-1)/\sigma} \right] \right\}^{\sigma/(\sigma-1)},$$

with $q_{ij,t}^f(n)$ and $q_{ij,t}^s(n)$ denoting the quantities consumed of a variety n from country i in market j and period t produced by a flexible or a sticky firm, respectively. All flexible firms in period t set the same price $p_{ij,t}^f$ given by equation (3), while firms not able to adjust their price in period t charge an average price $p_{ij,t}^s$ defined by equation (4), with heterogeneity in prices $p_{ij,t}^s(n)$ across $n \in N_i^s$ depending on the last adjustment period of each firm $n \in N_i^s$. Therefore, preferences can be expressed as defined in section 2.2.

The budget constraint is $\sum_i \left[\sum_{n \in N_i^f} p_{ij,t}^f(n) q_{ij,t}^f(n) + \sum_{n \in N_i^s} p_{ij,t}^s(n) q_{ij,t}^s(n) \right] = E_{j,t}$. Considering that $p_{ij,t}^f(n) = p_{ij,t}^f$ for all varieties $n \in N_i^f$, that is, price-adjusting firms in period t set the same price, it leads to the expression in section 2.2.

Bilateral price index. The price index in market j of varieties produced by firms from country i in period t is $P_{ij,t} = \left[\sum_{n \in N_i^f} [p_{ij,t}^f(n)]^{1-\sigma} + \sum_{n \in N_i^s} [p_{ij,t}^s(n)]^{1-\sigma} \right]^{1/(1-\sigma)}$. Considering that all flexible firms set the same price $p_{ij,t}^f$ and that the price index in market j and period t of the varieties produced by sticky firms from country i is given by:

$$P_{ij,t}^s \equiv \left[\sum_{n \in N_i^s} [p_{ij,t}^s(n)]^{1-\sigma} \right]^{1/(1-\sigma)}, \tag{A1}$$

we obtain the expression for the price index of varieties from country i in market j and period t, denoted as $P_{ij,t}$, defined in equation (9).

Firms' profits. The profit of a flexible firm from country i in a market j in period t is $\pi_{ij,t}^f = p_{ij,t}^f x_{ij,t}^f / \sigma - F P_{i,t}$, with $P_{i,t}$ denoting the price index in country i in period t, and $x_{ij,t}^f$ referring to the quantities produced by a flexible firm from country i for market j in period t. Similarly, the profit of a sticky firm from country i in a market j in period t selling a variety n at price $p_{ij,t}^s(n)$ is $\pi_{ij,t}^s(n) = p_{ij,t}^s(n) x_{ij,t}^s(n) / \sigma - F P_{i,t}$ for any $n \in N_i^s$, whereas the profit in period t of a sticky firm from country i selling in market j at the average price $p_{ij,t}^s$ is $\pi_{ij,t}^s = p_{ij,t}^s x_{ij,t}^s / \sigma - F P_{i,t}$ where $x_{ij,t}^s$ are the quantities produced by these firms for market j in period t.

Aggregation. The total expenditure in country j in goods produced in country i is $X_{ij,t} = X_{ij,t}^f + X_{ij,t}^s$. Substituting the terms on the right-hand side with expressions from equations (7) and (8) and using $N_i^f = (1-\alpha) N_i$, we compute equation (10).

C Transition dynamics

C.1 Initial conditions

The initial steady state is characterized by the following system of equations:

$$X_{ij,0} = \left(t_{ij,0}p_{i,0}/P_{j,0}\right)^{1-\sigma} N_i E_{j,0}, \tag{A2a}$$

$$(P_{j,0})^{1-\sigma} = \sum_i \left(t_{ij,0}p_{i,0}\right)^{1-\sigma} N_i, \tag{A2b}$$

$$(p_{i,0})^{1-\sigma} = \frac{E_{i,0}}{N_i \sum_j \left(t_{ij,0}p_{i,0}/P_{j,0}\right)^{1-\sigma} E_{j,0}}, \tag{A2c}$$

$$E_{i,0} = w_{i,0}L_i + \frac{N_i(p_{i,0})^{1-\sigma}}{\sigma} \sum_j \left(t_{ij,0}/P_{j,0}\right)^{1-\sigma} E_{j,0} - \mathfrak{J}FP_{i,0}N_i, \tag{A2d}$$

where $p_{ij,0} = \bar{p}_{ij,0} = t_{ij,0}p_{i,0}$. Note that the system of equations (A2) refers to a standard gravity system in the initial steady state, i.e., before the bilateral trade cost shock in $t = 1$.

C.2 One-period ahead transition dynamics relative to initial steady state

Given that $t = 0$ is defined as a steady state, all firms in country i charge the same prices $p_{ij,0}$, which implies $p_{ij,0} = \bar{p}_{ij,0}$. From equation (A2) and using $P_{ij,0} = (N_i)^{1/(1-\sigma)}\bar{p}_{ij,0}$, we have $X_{ij,0} = \left(P_{ij,0}/P_{j,0}\right)^{1-\sigma} E_{j,0}$. This leads to:

$$\left(\bar{p}_{ij,0}/P_{j,0}\right)^{1-\sigma} = X_{ij,0}/(N_i E_{j,0}). \tag{A3}$$

Based on equation (10), substituting with equation (A3), and denoting as $\hat{z}_t \equiv z_t/z_{t-1}$ the ratio of any variable in period t relative to $t-1$, we have:

$$X_{ij,1} = (1-\alpha)N_i(p^f_{ij,1}/P_{j,1})^{1-\sigma}E_{j,1} + \alpha X_{ij,0}\hat{E}_{j,1}/(\hat{P}_{j,1})^{1-\sigma}. \tag{A4}$$

C.3 Transition dynamics: Decay effect of higher-order lags

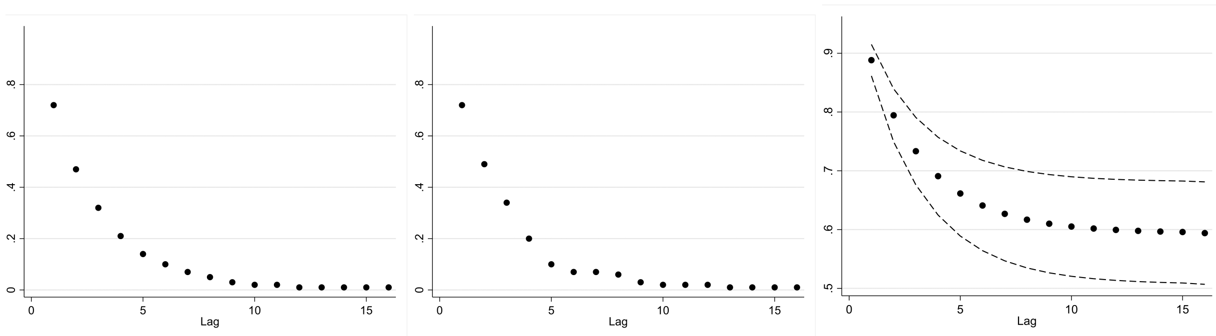
We extend our model by introducing higher-order lags and exploring the decay in their effects on bilateral trade flows. Considering that equation (15) holds for any period t, and approximating the unobservable $X^f_{ij,t-z} \approx (1-\alpha)X_{ij,t-z}$, we derive the generalized expression of the model with n lags:

$$\begin{aligned} X_{ij,t} \approx & (1-\alpha)N_i E_{j,t} \left(\frac{p^f_{ij,t}}{P_{j,t}}\right)^{1-\sigma} + \sum_{z=1}^{n-1} \alpha^z (1-\alpha)^2 X_{ij,t-z} \frac{E_{j,t}}{E_{j,t-z}} \left(\frac{P_{j,t-z}}{P_{j,t}}\right)^{1-\sigma} \\ & + \alpha^n \frac{E_{j,t}}{E_{j,t-n}} \left(\frac{P_{j,t-n}}{P_{j,t}}\right)^{1-\sigma} X_{ij,t-n}. \end{aligned} \tag{A5}$$

As $n \to \infty$, the last term on the right-hand side approaches zero and, ceteris paribus, the z-order lagged trade flows effect bilateral trade flows with a geometric decay at a rate α^z. See Appendix C.3.3 for proofs and further details, including an adjusted estimation procedure for a dynamic gravity model with higher-order lags.

C.3.1 Simulated data

Using simulated data, we analyze the performance of our procedure in estimating the decay patterns in trade flows as predicted by the theory. We consider the case where trade costs are permanently reduced from 1.3 to 1 in period $t = 1$ for trade flows between countries 1 and 2, as it provides the simplest framework to capture the transition path derived in our theoretical model.



(a) $\widehat{\alpha}$ constrained est. (b) $\widehat{\alpha}$ unconstrained est. (c) Lagged trade flow coef.

Figure A1: Decay effect relative to $t = 0$.

Figures A1a and A1b show $\widehat{\alpha}$ using simulated data for each period t relative to $t = 0$ (i.e., the initial steady state). In all specifications, we observe a geometric decay, as predicted by the theory. Figure A1c provides evidence on the persistence in trade flows, which is also consistent with the geometric decay predicted (ceteris paribus) by our theoretical model. Similar patterns are shown in the reduced-form estimates reported in Table A1 in Appendix A.2.

C.3.2 Proofs related to equation (14) and its approximation error

We compute a first-order Taylor approximation of $P^s_{ij,t}$ around the average price in the previous period (i.e., $\bar{p}_{ij,t-1}$), and considering $\sum_{n \in N^s_i} [p^s_{ij,t}(n) - \bar{p}_{ij,t-1}] = 0$, we have $P^s_{ij,t} \approx (N^s_i)^{\frac{1}{1-\sigma}} \bar{p}_{ij,t-1}$. Therefore, replacing this into equation (10), we obtain (14).

To obtain an explicit expression of the error from the first-order Taylor approximation above, we start from equation (10) and compute:

$$\begin{aligned} X_{ij,t} &= \left[(1-\alpha) N_i + \sum_{n \in N^s_i} \left(\frac{\kappa^s_{ij,t}(n)}{\kappa^f_{ij,t}} \right)^{1-\sigma} + \alpha N_i \left(\frac{\kappa^s_{ij,t}}{\kappa^f_{ij,t}} \right)^{1-\sigma} - \alpha N_i \left(\frac{\kappa^s_{ij,t}}{\kappa^f_{ij,t}} \right)^{1-\sigma} \right] \left(\frac{p^f_{ij,t}}{P_{j,t}} \right)^{1-\sigma} E_{j,t}, \\ &= (1-\alpha) N_i \left(\frac{p^f_{ij,t}}{P_{j,t}} \right)^{1-\sigma} E_{j,t} + \alpha N_i \left(\kappa_{ij,t} \right)^{1-\sigma} \left(\frac{p^f_{ij,t}}{P_{j,t}} \right)^{1-\sigma} E_{j,t} + e_{ij,t}, \end{aligned} \tag{A6}$$

with

$$e_{ij,t} \equiv \sum_{n \in N^s_i} \left([\kappa^s_{ij,t}(n)]^{1-\sigma} - [\kappa^s_{ij,t}]^{1-\sigma} \right) \left(\frac{t_{ij,t} p_{i,t}}{P_{j,t}} \right)^{1-\sigma} E_{j,t}. \tag{A7}$$

Replacing $\kappa_{ij,t} = \bar{p}_{ij,t-1} / p^f_{ij,t}$, we obtain equation (14).

Finally, to express bilateral trade flows in terms of observable lagged trade flows (instead of the unobservable average price $\bar{p}_{ij,t-1}$), we approximate $\left(\bar{p}_{ij,t-1} / P_{j,t-1} \right)^{1-\sigma} \approx X_{ij,t-1} / (N_i E_{j,t-1})$ and compute the bilateral trade flow for any period t in terms of the respective flows at $t-1$:

$$X_{ij,t} \approx (1-\alpha)N_i E_{j,t} \left(\frac{p^f_{ij,t}}{P_{j,t}}\right)^{1-\sigma} + \alpha \frac{E_{j,t}}{E_{j,t-1}} \left(\frac{P_{j,t-1}}{P_{j,t}}\right)^{1-\sigma} X_{ij,t-1} + e_{ij,t}. \tag{A8}$$

Defining $\hat{z}_t \equiv z_t / z_{t-1}$ for any variable z and excluding the error term $e_{ij,t}$, we yield equation (15).

C.3.3 Proofs for the model structure with higher-order lags

In this subsection, we discuss the model structure for two and three lags, and the generalization to n lags.

Model with two lags. Considering that equation (15) holds for any period t, we compute $X_{ij,t-1} \approx (1-\alpha)N_i E_{j,t-1}(p^f_{ij,t-1}/P_{j,t-1})^{1-\sigma} + \alpha \hat{E}_{j,t-1}/(\hat{P}_{j,t-1})^{1-\sigma} X_{ij,t-2}$. Replacing this expression in equation (15), we obtain:

$$X_{ij,t} \approx (1-\alpha)N_i E_{j,t} \left(\frac{p^f_{ij,t}}{P_{j,t}}\right)^{1-\sigma} + \alpha \frac{\hat{E}_{j,t}}{(\hat{P}_{j,t})^{1-\sigma}} (1-\alpha) X^f_{ij,t-1} + \alpha^2 \frac{E_{j,t}}{E_{j,t-2}} \left(\frac{P_{j,t-2}}{P_{j,t}}\right)^{1-\sigma} X_{ij,t-2}.$$

where we use $X^f_{ij,t-1} = N_i E_{j,t-1}(p^f_{ij,t-1}/P_{j,t-1})^{1-\sigma}$ (the right-hand side representing the standard gravity expression for bilateral trade flows in period $t-1$ when all firms set flexible prices). Using $X^f_{ij,t-1} \approx (1-\alpha)X_{ij,t-1}$ leads to the following:

$$X_{ij,t} \approx (1-\alpha)N_i E_{j,t} \left(\frac{p^f_{ij,t}}{P_{j,t}}\right)^{1-\sigma} + \alpha \frac{\hat{E}_{j,t}}{(\hat{P}_{j,t})^{1-\sigma}} (1-\alpha)^2 X_{ij,t-1} + \alpha^2 \frac{E_{j,t}}{E_{j,t-2}} \left(\frac{P_{j,t-2}}{P_{j,t}}\right)^{1-\sigma} X_{ij,t-2}.$$

Model with three lags. We extend the expression above to incorporate a third-order lag. We replace $X_{ij,t-2}$ with its expression as a function of $X_{ij,t-3}$, as before, we replace the unobservable bilateral flows for flexible firms by $X^f_{ij,t-z} = N_i E_{j,t-z}(p^f_{ij,t-z}/P_{j,t-z})^{1-\sigma}$ for $z = 1,2$, we approximate $X^f_{ij,t-z} \approx (1-\alpha)X_{ij,t-z}$, leading to:

$$\begin{aligned} X_{ij,t} \approx & (1-\alpha)N_i E_{j,t} \left(\frac{p^f_{ij,t}}{P_{j,t}}\right)^{1-\sigma} + \alpha \frac{\hat{E}_{j,t}}{(\hat{P}_{j,t})^{1-\sigma}} (1-\alpha)^2 X_{ij,t-1} \\ & + \alpha^2 \frac{E_{j,t}}{E_{j,t-2}} \left(\frac{P_{j,t-2}}{P_{j,t}}\right)^{1-\sigma} (1-\alpha)^2 X_{ij,t-2} + \alpha^3 \frac{E_{j,t}}{E_{j,t-3}} \left(\frac{P_{j,t-3}}{P_{j,t}}\right)^{1-\sigma} X_{ij,t-3}. \end{aligned}$$

Generalizing this for n-order lags, we obtain equation (A5) in section C.3.

C.4 Standard gravity bias

We derive equation (25) starting from the model specified in equation (17). Replacing $X_{ij,t-1}/(N_i E_{j,t-1})$ with $\left(\bar{p}_{ij,t-1}/P_{j,t-1}\right)^{1-\sigma}$, we recover the equivalent expression (14), and we compute:

$$
\begin{aligned}
X_{ij,t} =&(1-\alpha) N_i \left(t_{ij,t}\right)^{1-\sigma} (\kappa^f_{ij,t} p_{i,t}/P_{j,t})^{1-\sigma} E_{j,t} + \alpha N_i (\bar{p}_{ij,t-1}/P_{j,t})^{1-\sigma} E_{j,t}, \\
=&(1-\alpha) N_i \left(t_{ij,t}\right)^{1-\sigma} \left(\frac{\kappa^f_{ij,t} p_{i,t}}{P_{j,t}}\right)^{1-\sigma} E_{j,t} \\
&+ \alpha N_i \left(t_{ij,t-1}\right)^{1-\sigma} \left(\frac{p_{i,t-1}[(1-\alpha)\kappa^f_{ij,t-1} + \alpha \kappa^s_{ij,t-1}]}{P_{j,t}}\right)^{1-\sigma} E_{j,t},
\end{aligned} \tag{A9}
$$

where the last equation uses $\bar{p}_{ij,t-1} = t_{ij,t-1} p_{i,t-1}[(1-\alpha)\kappa^f_{ij,t-1} + \alpha\kappa^s_{ij,t-1}]$. We factor out common components to end up with the following expression:

$$
X_{ij,t} = N_i p_{i,t}^{1-\sigma} \frac{E_{j,t}}{P_{j,t}^{1-\sigma}} \left(t_{ij,t}\right)^{1-\sigma} \left\{ (1-\alpha) \left(\kappa^f_{ij,t}\right)^{1-\sigma} + \alpha \left[\frac{p_{i,t-1}}{p_{i,t}}[(1-\alpha)\kappa^f_{ij,t-1} + \alpha\kappa^s_{ij,t-1}]\right]^{1-\sigma} \right\},
$$

replacing with $\exp(\psi_{i,t}) \equiv N_i p_{i,t}^{1-\sigma}$, $\exp(\chi_{j,t}) \equiv E_{j,t}/P_{j,t}^{1-\sigma}$ and taking the conditional expectation, we obtain equation (25).

C.4.1 Analysis of the sluggish adjustment factor

To simplify the analysis, we assume that $\kappa^f_{ij,t} \approx 1$ for every period t. Consider a trade cost reduction $t_{ij,t-1} > t_{ij,t}$. When $\kappa^s_{ij,t-1} > 1$, for example due to the transition phase induced by a trade cost reduction, which may represent a progressive trade cost reduction from the implementation of an RTA, or when the effect of $\kappa^s_{ij,t-1}$ is relatively small relative to the current trade cost change (in the case the economy is still in the transition phase from a previous trade cost increase), we have $s_{ij,t} < 1$. Moreover, when trade cost reduction occurs before period $t-1$ (i.e., $t_{ij,t-1} = t_{ij,t}$) but the economy is still in the transition phase due to sluggish adjustment of prices, we have $\kappa^s_{ij,t-1} > 1$ and $p_{i,t-1} < p_{i,t}$—or $p_{i,t-1} \approx p_{i,t}$ when general equilibrium effects are relatively small, leading also to $s_{ij,t} < 1$. Therefore, in both situations, the increase in bilateral trade flows resulting from a reduction in trade costs is attenuated by the sluggish adjustment factor relative to the perfectly flexible world.

We now consider a trade cost increase, which leads to $t_{ij,t-1} < t_{ij,t}$. When $\kappa^s_{ij,t-1} < 1$ due to the transition phase from a previous trade cost increase, which can represent an escalation of a trade war, or when the effect $\kappa^s_{ij,t-1}$ is relatively small to the current trade cost change (if the economy is still in the transition phase from a previous trade cost reduction), we have $s_{ij,t} > 1$. Moreover, when the trade cost increase takes place before period $t-1$, but the system is still in the transition phase due to the sluggish adjustment of prices, we have $\kappa^s_{ij,t-1} < 1$ and $p_{i,t-1} > p_{i,t}$, or $p_{i,t-1} \approx p_{i,t}$ when the general equilibrium effects are relatively small, also leading to $s_{ij,t} > 1$. Therefore, in both situations, the reduction in bilateral trade flows due to the increase is attenuated by the sluggish adjustment factor.

C.4.2 Analysis of the trade cost bias in the standard gravity model

For simplicity, we analyze the bias in the trade cost coefficient when estimating a standard gravity regression that does not account for sluggish price adjustment in a log-linearized version of the model. In addition, we assume $\kappa^f_{ij,t} = 1$ for every period t. The log-linearized model is given by:

$$\ln X_{ij,t} \approx \psi_{i,t} + \chi_{j,t} + (1-\sigma)\ln t_{ij,t} + \ln s_{ij,t},$$

with $s_{ij,t}$ defined in equation (25). The standard gravity specification omits the term $\ln s_{ij,t}$ leading to a bias, as discussed in section 3.2. We compute:

$$\frac{\partial \ln s_{ij,t}}{\partial t_{ij,t}} = \frac{\alpha}{s_{ij,t}} \left(\frac{t_{ij,t-1}}{t_{ij,t}}\right)^{1-\sigma} \left(\frac{p_{i,t-1}}{p_{i,t}}\right)^{1-\sigma} \left[(1-\alpha) + \alpha\kappa^s_{ij,t-1}\right]^{1-\sigma} \frac{(\sigma-1)}{t_{ij,t}} > 0,$$

which shows that the sluggish adjustment factor is increasing in trade costs. This implies $\text{cov}(\ln t_{ij,t}, \ln s_{ij,t}) > 0$, leading to an upward bias in the estimated trade cost coefficient. In other words, the standard gravity model estimates a lower increase in bilateral trade flows during the transition phase after a reduction in bilateral trade costs. Hence, it underestimates the reduction in bilateral trade flows.

C.4.3 Standard gravity bias for RTA elasticity

To illustrate the bias structure in a commonly used specification of bilateral trade costs, we define $t_{ij,t} = dist_{ij}^{\beta_d} \beta_r^{RTA_{ij,t}}$ with $\beta_d > 0$ and $\beta_r \in (0,1)$ and $RTA_{ij,t}$ is a dummy variable that takes the value one if country i and j are members of a regional trade agreement in period t, and zero otherwise. The log-linearized model based on equation (25) follows as $\ln X_{ij,t} = \beta_d(1-\sigma)\ln dist_{ij} + (1-\sigma)\ln(\beta_r)RTA_{ij,t} + \psi_{i,t} + \chi_{j,t} + \ln s^{RTA}_{ij,t}$, with

$$s^{RTA}_{ij,t} \equiv (1-\alpha)(\kappa^f_{ij,t})^{1-\sigma} + \alpha \left(\frac{\beta_r^{RTA_{ij,t-1}}}{\beta_r^{RTA_{ij,t}}}\right)^{1-\sigma} \left(\frac{p_{i,t-1}}{p_{i,t}}\right)^{1-\sigma} \left[(1-\alpha)\kappa^f_{ij,t-1} + \alpha\kappa^s_{ij,t-1}\right]^{1-\sigma}. \quad \text{(A10)}$$

The standard specification leads to an omitted variable bias problem, in this case the expected value of the estimates of the RTA coefficient are given by $\mathbb{E}[\beta^f_1] = \beta_1 + \beta_2 \text{cov}\left[RTA_{ij,t}, \ln s^{RTA}_{ij,t}\right] / \text{var}\left[RTA_{ij,t}\right]$, where $\mathbb{E}[\beta^f_1]$ again refers to the expected value of the log trade cost coefficient in the standard specification, $\beta_1 = (1-\sigma)\ln\beta_r > 0$ refers to the true value of the RTA coefficient, and $\beta_2 = 1$ refers to the true value of the coefficient related to $\ln s^{RTA}_{ij,t}$. Hence, considering that $\text{cov}\left[RTA_{ij,t}, \ln s^{RTA}_{ij,t}\right] < 0$, $\beta_1 = (1-\sigma)\ln\beta_r > 0$ and $\beta_2 = 1$, this leads to a downward bias in the estimated coefficient for RTAs in the transition phase for the standard specification (i.e., $\beta^f_1 < \beta_1$). This implies that when data from the transition phase are used, the standard model estimates a lower increase in bilateral trade flows due to an RTA in the new steady state.

For illustration purposes, we consider the case of the implementation of an RTA in period t between countries i and j (i.e., $RTA_{ij,t} = 1$ and $RTA_{ij,t-1} = 0$). Regarding the sluggish adjustment factor, we have $(\beta_r^{RTA_{ij,t-1}}/\beta_r^{RTA_{ij,t}})^{1-\sigma} = (1/\beta_r)^{1-\sigma} < 1$, which implies that $s^{RTA}_{ij,t} < 1$. If the implementation of the RTA takes place before period t (i.e., $RTA_{ij,t} = RTA_{ij,t-1} = 1$) and the economy is still in the transition phase (under similar conditions as in the previous case), we have $s^{RTA}_{ij,t} < 1$. In both cases, the sluggish adjustment factor attenuates the positive effect of an RTA on bilateral trade flows.

Proofs. The log-linearized model is now given by:

$$\ln X_{ij,t} \approx \psi_{i,t} + \chi_{j,t} + \beta_d(1-\sigma)\ln dist_{ij} + (1-\sigma)\ln(\beta_r)RTA_{ij,t} + \ln s_{ij,t}^{RTA},$$

with $s_{ij,t}^{RTA}$ defined by equation (A10). For simplicity, we still assume that $\kappa_{ij,t}^{f} = 1$ for every period t. As before, the standard gravity specification omits the last term, leading to a bias in the steady-state estimate of the RTA coefficient given by $(1-\sigma)\ln(\beta_r)$. For simplicity, taking $RTA_{ij,t}$ as continuous, we compute:

$$\frac{\partial \ln s_{ij,t}^{RTA}}{\partial RTA_{ij,t}} = \frac{\alpha}{s_{ij,t}^{RTA}} \left(\frac{\beta_r^{RTA_{ij,t-1}}}{\beta_r^{RTA_{ij,t}}}\right)^{1-\sigma} \left(\frac{p_{i,t-1}}{p_{i,t}}\right)^{1-\sigma} \left[(1-\alpha) + \alpha\kappa_{ij,t-1}^{s}\right]^{1-\sigma} (\sigma-1)\ln\beta_r < 0,$$

which implies $\mathrm{cov}(RTA_{ij,t}, \ln s_{ij,t}^{RTA}) < 0$, leading to a downward bias in the estimated coefficient of an RTA in the transition phase in the standard gravity specification. This implies that the standard gravity model estimates a lower increase in bilateral trade flows during the transition phase after the conclusion of an RTA.

D Potential Nickell bias in the sticky-part gravity equation

We show that the Nickell bias (Nickell, 1981) that arises in standard dynamic panel models does not apply in our setting. We distinguish two cases, namely the case where γ_1 is treated as unknown and estimated, and the case where γ_1 is assumed to be a known (theory-consistent) parameter.

D.1 Assuming γ_1 to be an unknown parameter

Based on the conditional expectation for the sticky-part as provided in equation (22), we can state the sticky-part gravity equations as follows:

$$X_{ij,t} = \exp\left(\gamma_1 \ln\left(X_{ij,t-1}\right) + \phi_{j,t}\right) + \varepsilon_{ij,t}, \tag{A11}$$

or, when estimated in its log-linearized form, as follows:

$$\ln\left(X_{ij,t}\right) = \gamma_1 \ln\left(X_{ij,t-1}\right) + \phi_{j,t} + \epsilon_{ij,t}, \tag{A12}$$

where $\varepsilon_{ij,t}$ and $\epsilon_{ij,t}$ are remainder error terms. We first take a one period lag of Equation (A12):

$$\ln\left(X_{ij,t-1}\right) = \gamma_1 \ln\left(X_{ij,t-2}\right) + \phi_{j,t-1} + \epsilon_{ij,t-1}. \tag{A13}$$

Using Equation (A13) to replace $\ln\left(X_{ij,t-1}\right)$ in Equation (A12) leads to:

$$\ln\left(X_{ij,t}\right) = \gamma_1\left(\gamma_1 \ln\left(X_{ij,t-2}\right) + \phi_{j,t-1} + \epsilon_{ij,t-1}\right) + \phi_{j,t} + \epsilon_{ij,t}. \tag{A14}$$

Unlike the standard dynamic panel data case (see Cameron and Trivedi, 2005, for example), we have $\phi_{j,t-1}$ and $\phi_{j,t}$ in Equation (A14) (and not ϕ_j both times). $\phi_{j,t-1}$ and $\phi_{j,t}$ are orthogonal by construction and identified from the cross-section in each period but not across periods. Hence, $E[\epsilon_{ij,t}|\ln\left(X_{ij,t-1}\right), \phi_{j,t}] = 0$, i.e., so that the sequential exogeneity assumption is fulfilled, as only one lag of the dependent variable is needed to capture the dynamics. The standard argument for Nickell bias does not apply, because we have j fixed effects for every period. Therefore, we can estimate this equation using OLS as long as the standard assumptions (such as no serial correlation) hold.

Equivalently, taking a one-period lag of equation (A12), we have:

$$X_{ij,t-1} = \exp\left(\gamma_1 \ln\left(X_{ij,t-2}\right) + \phi_{j,t-1}\right) + \varepsilon_{ij,t-1}. \tag{A15}$$

Using equation (A15) to replace $X_{ij,t-1}$ in equation (A11) leads to:

$$X_{ij,t} = \exp\left(\gamma_1 \ln\left(\exp\left(\gamma_1 \ln\left(X_{ij,t-2}\right) + \phi_{j,t-1}\right) + \varepsilon_{ij,t-1}\right)\right) + \phi_{j,t} + \varepsilon_{ij,t}. \tag{A16}$$

As in the linear case, we see that Equation (A14) includes $\phi_{j,t-1}$ and $\phi_{j,t}$, which are orthogonal and identified only from the cross-section. Hence, the standard Nickell bias also does not apply.

D.2 Assuming γ_1 to be a known parameter

When we take γ_1 as a known parameter, i.e., restrict it to be one as suggested by theory, for example, we can reformulate Equation (A12) as follows:

$$\ln\left(X_{ij,t}\right) - \gamma_1 \ln\left(X_{ij,t-1}\right) = \phi_{j,t} + \epsilon_{ij,t}, \tag{A17}$$

i.e., we bring $\gamma_1 \ln \left(X_{ij,t-1}\right)$ to the left-hand side. In other words, we construct a new dependent variable given by $\ln \left(X_{ij,t}\right) - \gamma_1 \ln \left(X_{ij,t-1}\right)$. Then we are back in a standard panel setting.

Note that this also works with the multiplicative specification. Let us first assume a multiplicative error term. We start with Equation (A11) and re-write:

$$
\begin{aligned}
X_{ij,t} &= \exp\left(\gamma_1 \ln\left(X_{ij,t-1}\right) + \phi_{j,t}\right) \eta_{ij,t} \\
&= \exp\left(\gamma_1 \ln\left(X_{ij,t-1}\right)\right) \exp\left(\phi_{j,t}\right) \eta_{ij,t}.
\end{aligned}
$$

Now, we bring $\exp\left(\gamma_1 \ln\left(X_{ij,t-1}\right)\right)$ to the right-hand side:

$$
\begin{aligned}
X_{ij,t} \exp\left(-\gamma_1 \ln\left(X_{ij,t-1}\right)\right) &= \exp\left(\phi_{j,t}\right) \eta_{ij,t} \\
&= \exp\left(\phi_{j,t}\right) + \varepsilon_{ij,t},
\end{aligned}
$$

where we transformed the multiplicative error term into an additive error term using the relationship $\eta_{ij,t} = 1 + \varepsilon_{ij,t} / \exp\left(\phi_{j,t}\right)$. Hence, the new dependent variable is: $X_{ij,t} \exp\left(-\gamma_1 \ln\left(X_{ij,t-1}\right)\right)$ in the multiplicative case. If we start with an additive error, we have:

$$
\begin{aligned}
X_{ij,t} &= \exp\left(\gamma_1 \ln\left(X_{ij,t-1}\right) + \phi_{j,t}\right) + \varepsilon_{ij,t} \\
&= \exp\left(\gamma_1 \ln\left(X_{ij,t-1}\right)\right) \exp\left(\phi_{j,t}\right) + \varepsilon_{ij,t}.
\end{aligned}
$$

Now, we bring $\exp\left(\gamma_1 \ln\left(X_{ij,t-1}\right)\right)$to the right-hand side:

$$
X_{ij,t} \exp\left(-\gamma_1 \ln\left(X_{ij,t-1}\right)\right) = \exp\left(\phi_{j,t}\right) + \varepsilon_{ij,t} \exp\left(-\gamma_1 \ln\left(X_{ij,t-1}\right)\right).
$$

Hence, we obtain the same new dependent variable. The only difference from the case with multiplicative error is that the error term is now heteroskedastic, i.e., varies with $\exp\left(-\gamma_1 \ln\left(X_{ij,t-1}\right)\right)$.

E Simulated data

E.1 Trade cost shock

Tables A2 and A3 report the results of the baseline simulations of our procedure for a trade cost shock between country 1 and country 2 for PPML and log-linear specifications, respectively. In both cases, our procedure recovers a stickiness parameter close to the true value (i.e., $\alpha = 0.7$) and steady-state trade cost elasticities closer to the true value (i.e., $\beta_1 = -4$) compared to the standard gravity specification. As expected, the difference between the transition phase estimate and the steady-state estimate of the trade cost elasticity is reduced because the sample includes more periods after the shock.

Table A2: Regression results for trade cost specification: PPML model

	Step 1					
	Sample: 10 periods		**Sample: 20 periods**		**Sample: 71 periods**	
	(1)	(2)	(3)	(4)	(5)	(6)
	Flex Reg	Sticky Reg	Flex Reg	Sticky Reg	Flex Reg	Sticky Reg
$\ln(t_{ij,t})$	-3.191***		-3.596***		-3.889***	
	(0.131)		(0.0801)		(0.0247)	
$INTER_{ij}$	-0.212***		-0.106***		-0.0290***	
	(0.0344)		(0.0210)		(0.00648)	
$\ln X_{ij,t-1}$		1.001***		1.001***		1.000***
		(0.000336)		(0.000174)		(0.0000502)
FEs	it, jt	jt	it, jt	jt	it, jt	jt
	Step 2					
Stickiness degree: $\widehat{\alpha}$	0.65		0.71		0.73	
	Step 3					
Est. Coef: $\widehat{\beta}_1$	-3.655***		-3.896***		-3.980***	
	(0.00363)		(0.00178)		(0.000400)	
Observations	9,900		18,900		64,800	

Notes: *Sample* refers to the number of periods after shock included. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: Regression results for trade cost specification: log-linear model

	Step 1					
	Sample: 10 periods		Sample: 20 periods		Sample: 71 periods	
	(1)	(2)	(3)	(4)	(5)	(6)
	Flex Reg	Sticky Reg	Flex Reg	Sticky Reg	Flex Reg	Sticky Reg
$\ln(t_{ij,t})$	-3.141***		-3.556***		-3.875***	
	(0.00740)		(0.00491)		(0.00163)	
$INTER_{ij}$	-0.225***		-0.116***		-0.0328***	
	(0.00199)		(0.00133)		(0.000439)	
$\ln X_{ij,t-1}$		1.002***		1.001***		1.000***
		(0.000364)		(0.000189)		(0.0000550)
FEs	it, jt	jt	it, jt	jt	it, jt	jt
	Step 2					
Stickiness degree: $\widehat{\alpha}$	0.67		0.72		0.74	
	Step 3					
Est. Coef: $\widehat{\beta}_1$	-3.696***		-3.913***		-3.986***	
	(0.00485)		(0.00236)		(0.000490)	
Observations	9,900		18,900		64,800	

Notes: *Sample* refers to the number of periods after shock included. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

E.2 RTA shock

The simulated data consist of 30 countries. Each country has 100 firms and a labor endowment of 3000. The distance between a country pair ij with $i \neq j$ is drawn from a uniform distribution between 2.5 and 5, and later made symmetric by computing $(d_{ij} + d_{ji})/2$. The distance d_{ii} is drawn from a uniform distribution with lower bound 1 and upper bound 2. The dummy variable $INTER$ is defined as 1 when $i \neq j$, and zero otherwise; that is, it identifies international trade flows. The elasticity of substitution σ is set equal to 5. Fixed costs are normalized to zero (i.e., $F = 0$), and the price index of country 30 is set as the numéraire. In period 1, RTAs are formed. For RTAs, we draw integers from a uniform distribution with lower bound 1 and upper bound 2, assigning a 1 to the RTA dummy if we draw a 1, and zero else. Hence, while there are no RTAs in place in the baseline, a share of 0.233 of country pairs trade under an RTA in the counterfactual.

The RTA coefficient β_r is set equal to 0.9, which leads to $\beta_1 = (1 - \sigma)\ln(\beta_r) = 0.421$. The coefficient for distance β_d is set equal to 0.25, leading to $(1 - \sigma)\beta_d = -1$, and for the international border (INTER) we set β_i equal to 1.1, which implies that $(1 - \sigma)\ln(\beta_i) = -0.381$.

As in our baseline specification, Tables A4 and A5 report the results for PPML and log-linear specifications, respectively. Furthermore, when considering the introduction of RTAs, our procedure recovers the stickiness parameter close to the true value (i.e., $\alpha = 0.7$) and the long-run effect of RTAs closer to the true value (i.e., $\beta_1 = 0.421$) relative to the standard gravity specification. Again, the difference between the transition phase and the steady-state estimate of the effect of RTAs is reduced, as the sample includes more periods after the shock.

Table A4: Regression results: PPML model with distance

	Step 1					
	Sample: 10 periods		**Sample: 20 periods**		**Sample: 71 periods**	
	(1)	(2)	(3)	(4)	(5)	(6)
	Flex Reg	Sticky Reg	Flex Reg	Sticky Reg	Flex Reg	Sticky Reg
$\ln(dist_{ij})$	-1.000***		-1.000***		-1.000***	
	(0.00263)		(0.00173)		(0.000574)	
$INTER_{ij}$	-0.381***		-0.381***		-0.381***	
	(0.00262)		(0.00173)		(0.000577)	
RTA_{ijt}	0.328***		0.374***		0.408***	
	(0.00144)		(0.000883)		(0.000278)	
$\ln(X_{ij,t-1})$		0.999***		1.000***		1.000***
		(0.000612)		(0.000326)		(0.0000951)
FEs	*it, jt*	*jt*	*it, jt*	*jt*	*it, jt*	*jt*

	Step 2		
Stickiness degree: $\widehat{\alpha}$	0.65	0.69	0.71

	Step 3		
Est. Coef: $\widehat{(1-\sigma)\beta_d}$	-1.035***	-1.019***	-1.006***
	(0.000606)	(0.000230)	(0.0000394)
Est. Coef: $\widehat{\beta}_1$	0.393***	0.414***	0.421***
	(0.000364)	(0.000135)	(0.0000229)

Observations	9,900	18,900	64,800

Notes: *Sample* refers to the number of periods after shock included. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5: Regression results: log-linear model with distance

	Step 1					
	Sample: 10 periods		**Sample: 20 periods**		**Sample: 71 periods**	
	(1)	(2)	(3)	(4)	(5)	(6)
	Flex Reg	Sticky Reg	Flex Reg	Sticky Reg	Flex Reg	Sticky Reg
$\ln(dist_{ij})$	-1.000***		-1.000***		-1.000***	
	(0.00260)		(0.00170)		(0.000562)	
$INTER_{ij}$	-0.381***		-0.381***		-0.381***	
	(0.00300)		(0.00196)		(0.000649)	
RTA_{ijt}	0.326***		0.373***		0.408***	
	(0.000913)		(0.000583)		(0.000190)	
$\ln(X_{ij,t-1})$		1.009***		1.005***		1.001***
		(0.000800)		(0.000406)		(0.000116)
FEs	*it, jt*	*jt*	*it, jt*	*jt*	*it, jt*	*jt*

	Step 2		
Stickiness degree: $\widehat{\alpha}$	0.63	0.68	0.71

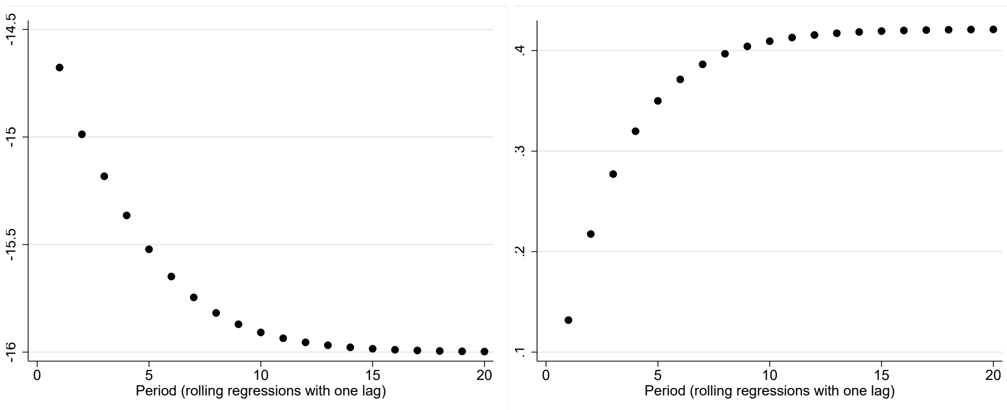
	Step 3		
Est. Coef: $\widehat{(1-\sigma)\beta_d}$	-0.890***	-0.925***	-0.976***
	(0.00112)	(0.000519)	(0.0000890)
Est. Coef: $\widehat{\beta}_1$	0.380***	0.409***	0.420***
	(0.000669)	(0.000303)	(0.0000515)

Observations	9,900	18,900	64,800

Notes: *Sample* refers to the number of periods after shock included. Standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

E.3 Simulated data: Estimation bias in steady-state coefficients

We illustrate the bias in standard gravity estimates for a permanent change in trade costs (i.e., in $t_{ij,t}$) for a random draw of countries. In this case, the sluggish adjustment factor is given by $s_{ij,t}$ as defined by equation (25), leading to an upward bias in the trade cost coefficient estimated by the standard gravity specification. This is illustrated in Figure A2a. Figure A2b shows the bias in the estimates for the steady-state elasticity of an RTA by the standard gravity model, as discussed in section 3.2.



(a) Trade cost coef. (steady state: -16) (b) RTA coef. (steady state: 0.421)

Figure A2: Standard gravity estimates

In both situations, we observe an attenuation of the shock in the initial periods. The estimated elasticities converge to their respective steady-state elasticities over time as firms adjust their consumer prices. We have performed a similar analysis for other types of shocks, showing similar patterns in the estimation of trade elasticities.

F Empirical analysis: application to RTAs

F.1 Additional evidence for the European Economic Area (EEA) agreement

We compute the yearly change in the average bilateral tariff among members of each RTA as follows. We use bilateral tariff data by industry from CEPII, compute the yearly change in bilateral tariffs, and aggregate the tariff changes at the country level, averaging across industries. Finally, we compute the average annual tariff change between the members of the agreement. In other words, the annual change in the average bilateral tariff between the RTA members is given by $dtariff_t \equiv pref_t - pref_{t-1}$, where $pref_t$ is the average preferential tariff between the RTA members in period t.

In Figure A3, we plot the yearly change in the average bilateral tariff between EEA members (bars) and the ratio (dots), calculated from 5 years before entry into force to each year on the horizontal axis. The ratio rises following large tariff reductions and gradually fades over time. After the initial tariff drop in 1995, the ratio reaches 1.1, implying a 10% upward adjustment relative to the standard gravity estimate, then declines during years with no significant tariff changes. New tariff cuts in 2004 and 2007 (driven by the entry of new members to the EEA) raise the ratio again, up to 1.2, with a more persistent effect than in the earlier period. We use the EEA to illustrate the link between tariff changes and the estimated adjustment ratio, without examining which features of the agreement drive this pattern.

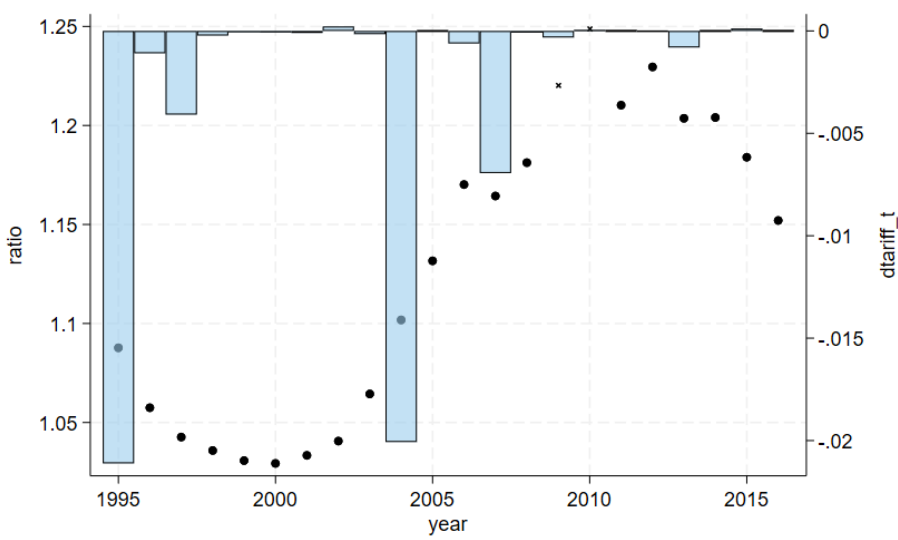


Figure A3: Ratio and tariff changes over time: EEA. Black dots report the recovered ratio if positive and significant first-step estimates, x reports the ratio when that condition is not fulfilled. Light blue columns show $dtariff_t$, reported on the right axis.

F.2 Ratio adjustment by tariff reduction: all FTAs

Here, we further examine the relationship between the adjustments implied by the recovered elasticity ratios and the bilateral preferential tariff reductions granted by FTAs relative to members' MFN tariffs. To do so, we split FTAs into five quantiles based on *dDIFF*, where Q1 includes agreements with the largest average preferential tariff reductions and Q5 those with the smallest.

The quantiles are constructed on the full FTA subsample with positive and significant first-step estimates.

Table A6 reports the mean elasticity ratio for two subsamples: FTAs with positive and significant first-step estimates (third row of Table 4) and FTAs with positive and significant first- and third-step estimates (last row of Table 4), by tariff-reduction quintile. The lower quintiles, especially Q1, exhibit greater tariff reductions and generally show larger adjustments in the mean ratio.

To explore heterogeneity, Table A7 further divides the statistics for FTAs signed before and after 2000. More recent agreements show a larger share of ratios above one, consistent with our theory that, all else equal, older agreements are more likely to be closer to their steady state, hence predicting a lower elasticity ratio. However, as shown for the EEA in Appendix F.1, older agreements with more recent tariff changes can still generate high ratios. We also compute the share of FTAs with a ratio greater than one (see columns *share ratio>1*). This share is highest in Q1 and, more importantly, substantially higher among more recent agreements.

Table A6: FTAs — Ratio by dDIFF quintiles

	positive and significant first-step			positive and significant first- and third-step		
	Mean *dDIFF*	Mean ratio	Number FTAs	Mean *dDIFF*	Mean ratio	Number FTAs
1	-0.776	1.180	25	-0.780	1.137	23
2	-0.566	0.931	25	-0.564	1.048	17
3	-0.415	1.033	24	-0.416	1.102	15
4	-0.245	0.979	25	-0.232	1.090	12
5	-0.058	0.890	24	-0.048	1.026	9

Table A7: Pre- and post-2000 FTAs — Ratio by dDIFF quintiles

	FTAs: pre 2000			FTAs: post 2000			all
Q	Mean ratio	share ratio>1	Number FTAs	Mean ratio	share ratio>1	Number FTAs	shr r>1
1	1.02	0.67	3	1.18	0.81	21	0.80
2	0.84	0.50	4	0.97	0.42	19	0.40
3	1.23	0.57	7	0.95	0.50	16	0.50
4	0.96	0.33	12	1.00	0.62	13	0.48
5	1.03	0.53	17	0.56	0.00	7	0.38

Notes: "share ratio>1" indicates the share of FTAs with an estimated elasticity ratio greater than one.