# Trade Persistence Heterogeneity[*]


Mariarosaria Comunale[†]
*International Monetary Fund*

Justas Dainauskas[‡]
*London School of Economics and Political Science*

Povilas Lastauskas[§]
*Queen Mary University of London and Vilnius University*



**Abstract**

International trade flows are volatile and imbalanced. We develop a theory of learning by importing, which generates autocorrelated bilateral trade flows that are heterogeneous across different country pairs. Our framework gives rise to a dynamic gravity equation that nests popular alternatives in the literature, namely a homogeneous parameter dynamic version with zero aggregate trade imbalances and a popular static gravity model. Not only does our model improve accuracy of trade flows predictions, but it is also consistent with the empirically relevant declines and rapid recoveries of trade flows in response to shocks, thereby escaping what we call the "trade persistence puzzle". We also show that small asymmetries in the learning-by-importing externality across countries are sufficient to create bilateral and multilateral trade imbalances endogenously, which are important both theoretically and empirically.



**JEL Classification:** C23, F14, F41, F62.

**Keywords:** Dynamic Gravity Equation; Learning-by-Importing; Trade Persistence; Trade Imbalance; Common Shocks; Parameter Heterogeneity.



[*]For useful comments and suggestions, we thank Dennis Novy, Yoto V. Yotov, Lu Han, Piyush Panigrahi, Alessandro Barattieri, Marinos Tsigas, Vanessa Gunnella, Alessia Paccagnini, Artūras Juodis, Milda Norkutė, Ziran (Josh) Ding, Ovidijus Stauskas, Tamara Gurevich, Chris Papageorgiou, Juan Passadore, Rui Mano as well as participants of the Trade Expert Network Workshop at the European Central Bank, seminar series at the U.S. International Trade Commission, Econometric Society European and African Meetings, Royal Economic Society Annual Conference, IAAE Annual Conference, EEA-ESEM Congress, RESDM internal seminar series at the International Monetary Fund, Midwest Economic Theory and International Trade Conference, and the Bank of Lithuania seminar series. For their help with data and/or estimation, we thank Jan Ditzen, Michele Mancini, and Ingrida Šimkutė. Disclaimer: The views expressed in this article are those of the authors and do not necessarily represent those of the Bank of Lithuania or the Eurosystem and should not be attributed to the International Monetary Fund, its Executive Board, or its management. Lastauskas is grateful for support from Iceland, Liechtenstein and Norway through EEA grants (project no. S-BMT-21-8 (LT08-2-LMT-K-01-073)) under a grant agreement with the Research Council of Lithuania (LMTLT).



[†]Address at: `mcomunale@imf.org`, International Monetary Fund, Research Department, 700 19th St NW, Washington, DC 20431, United States. `Formerly`: Applied Macroeconomic Research Unit (TMTS), Bank of Lithuania, Totoriu Ìš 4, LT-01121, Vilnius, Lithuania; Faculty of Economics and Business Administration, Vilnius University, Vilnius, Lithuania.

[‡]Address at: `j.dainauskas@lse.ac.uk`, Department of Economics, London School of Economics and Political Science, 32 Lincoln's Inn Fields, London WC2A 3PH, United Kingdom.

[§]Address at: `p.lastauskas@cantab.net`, Queen Mary University of London, Department of Business Analytics and Applied Economics, London, United Kingdom and Faculty of Economics and Business Administration, Vilnius University, Saulėtekio al. 9, 2nd Building, LT-10222 Vilnius, Lithuania.

# 1 Introduction

Trade adjustments at the extensive margin are infrequent. The patterns of "who trades with whom" are regional and take decades to adapt (Eichengreen and Irwin (1998)). But at the intensive margin, it takes much less time for the value of "how much is traded" among those who partner up to adjust in response to shocks. And when hit with common shocks, trade between some countries slows down more suddenly and more forcefully than others.

In this paper, we study the transitional trade adjustment dynamics at the intensive margin and their variation across different trade partners using the ubiquitous gravity equation. The standard gravity approach predicts bilateral trade flows as a function of country size and a multitude of other observable and unobservable factors that measure trade frictions at the bilateral and multilateral levels (Anderson and van Wincoop (2003); Head and Mayer (2014)). It is celebrated for its simplicity and success of accounting for the long-term distribution of global trade flows based on the countrywide economic, geographic, institutional, demographic, and historical characteristics.

However, we recognise three hurdles of fitting the timing and the variation of trade adjustment dynamics for different trade partners in the short-to-medium term. First, Isaac Netwon's universal law of gravitation and its trade counterpart are both static, which suggests immediate, as opposed to transitional, trade adjustments that are generally more abrupt than those that we observe at annual or higher frequencies. Second, the inference is typically drawn using pooled coefficient estimators that average trade elasticities across all trade partners and brushes with broad strokes over their idiosyncrasies. And third, gravity equations typically feature fixed effects that reflect both time-invariant trade resistance as well as the state of the world economy over time. But even if the shocks at different points in time are worldwide, the exposure to and the transmission of those shocks across the globe is likely to be country-specific due to differences in their state of development, institutions, openness, patterns of specialisation, and other structural characteristics.

Building on the widely-established panel data modelling techniques pioneered by Pesaran and Smith (1995), Pesaran (2006), and Chudik and Pesaran (2015), we estimate an augmented version of the gravity equation that: (i) incorporates dynamics; (ii) identifies heterogeneous gravity equation coefficients at a bilateral level; and (iii) accounts for common shocks with heterogeneous exposure. In doing so, we use widely-available data from the Penn World Tables (Feenstra et al. (2015)) and International Monetary Fund's (IMF) Direction of Trade Statistics (DOTS) for 39 developed and emerging market economies (i.e., 741 country pairs) over the period of 1950-2014. The metric that we use to summarize trade adjustment dynamics is the (bilateral) trade *persistence* coefficient, which measures the magnitude of trade flow autocorrelation controlling for all of their determinants.

Our key empirical findings and their wider implications are as follows. First, we find that the trade persistence coefficients between nearly all trade partners are significantly different from zero and unity. This suggests that shocks have a lasting, but not permanent, effect on trade flows and rejects the static gravity model specification at annual frequency. Second, compared with a variation of the standard gravity equation that simply adds lagged trade flows, and depending on the specifics of how we account for structural heterogeneity, the trade persistence coefficient roughly halves from 0.9 to around 0.35-0.55. And third, the estimates of the trade persistence coefficients for any exporter (i.e., the source country) vary significantly across different importers (i.e., the destination country). Specifically, we estimate the standard deviation of the trade persistence coefficients across different exporters to be around 0.15. Altogether, our empirical results caution against a "one size fits all" approach that only takes notice of trade persistence and turns a blind eye to structural heterogeneity. Because without accounting for their interaction comprehensively, the dynamic gravity equation not only overstates trade persistence for all trade partners systematically, but easy fixes, such as adding more data or lags, do not correct for these biases.

We control for two types of structural heterogeneity in our empirical analysis, each playing an equally significant role in our results. First, it is well-known that pooled, aggregated, or cross-sectional estimates of the *static* gravity equation coefficients give unbiased mean estimates. But in the context of *dynamic* gravity equations, it is not yet recognised that pooling and aggregating give inconsistent and potentially misleading estimates of the coefficients (see Pesaran and Smith (1995)).[1] Instead, to obtain consistent estimates we: (i) exploit the relatively long time dimension of our panel data; (ii) estimate the gravity equation for each country pair individually and recover heterogeneous coefficients at the bilateral level; and (iii) if we choose to compare our results with the pooled estimates, we average the bilateral coefficients across all country pairs (i.e., the Mean Group (MG) estimator). Second, because some countries are "small" and others are "large", unfavourable shocks that hit the large trade partners of different small countries diminish bilateral trade flows not only directly between large-small country pairs, but also indirectly and heterogeneously between small-small country pairs depending on their exposure (e.g., Gopinath et al. (2020)). However, the standard gravity equation neglects the differential impact of third-country effects on bilateral trade flows that are endogenous to the regressors, such as size, which causes the gravity equation coefficient estimates to be biased and inconsistent. Following the methodology of Pesaran (2006), we therefore specify a gravity equation with a multi-factor error structure that accounts for the presence of these so-called Common Correlated Effects (CCE) and mitigates the bias asymptotically.[2]

Our baseline gravity model estimation combines both CCE and MG (i.e., CCEMG developed by Chudik and Pesaran (2015)). In addition, to assess the robustness of our empirical results and to provide direct comparisons using the same data set, we implement and highlight the key differences of 13 additional variations of the dynamic gravity model specification and estimation, each replicating different related approaches of dynamic heterogeneous panel data modelling techniques. Moreover, we conduct a "horse race" comparison of how well each gravity model specification fits the data using: (i) in-sample forecasts for each country pair individually; and (ii) Root Mean Square Errors (RMSEs) for the global trade flows of all country pairs. What we learn is that our preferred CCEMG approach outperforms not only CCEP, which incorporates CCE and pools the regression coefficients, or MG, which only retains coefficient heterogeneity, but also all other specifications considered in this paper. Incidentally, CCEMG also predicts the lowest cross-sectionally averaged trade persistence coefficient of 0.35 compared to all other approaches considered in this paper. We reach a conclusion in our empirical analysis that trade persistence heterogeneity plays a significant role in terms of characterising the transitional trade adjustment dynamics both for individual country pairs and in the context of a cross-section of country pairs.

[1] The problem arises, because when regressors are serially correlated, incorrectly ignoring coefficient heterogeneity induces serial correlation in the disturbance, which generates inconsistent estimates in models with lagged dependent variables, even if the time dimension is large.

[2] The basic idea is to approximate common shocks by means of cross-sectionally averaged regressors and to estimate their effect on bilateral trade flows with a heterogeneous factor loading, such that as the number for country pairs goes to infinity, the differential impact of the unobserved third-country effects are eliminated.

A potential concern with our empirical strategy is that the current theoretical trade literature does not offer any mechanism that would justify trade persistence heterogeneity.[3] We do not attempt to develop a trade theory that fully explains the foundations of our empirical findings. However, we do present a simple and tractable theoretical framework that rationalises trade persistence heterogeneity in reduced-form. We then use that framework to derive a theoretically-consistent empirical specification of the dynamic gravity equation.

Our theoretical model incorporates a "learning-by-importing" externality into an otherwise standard trade resistance framework developed by Anderson and van Wincoop (2003). Explicitly or implicitly, the concept of learning-by-importing appears in a large body of the trade literature (e.g., Ethier (1982); Grossman and Helpman (1995); Keller (2004); Acharya and Keller (2009); Amiti and Konings (2007); Elliott et al. (2016); Halpern et al. (2015); Zhang (2017) among others). It generally refers to the exporter productivity gains from importing offshore intermediate inputs. The more exporters import, the more efficient they become. What makes trade persistent and heterogeneous across different country pairs in our theory is: (i) it takes time to internalise the productivity gains, such that exporter efficiency depends on lagged bilateral trade flows; and (ii) the capacity to internalise those gains depends on a parameter that reflects, among other things, bilateral infrastructure. Due to differences in the resulting bilateral efficiencies, the model generates both short-run and long-run multilateral trade imbalances that, as is standard, are mediated by international trade of assets (e.g., Dekle et al. (2008); Dix-Carneiro et al. (2021)). Consequently, in addition to country size and third-country effects, our empirical gravity equations feature an additional theoretically-motivated regressor that captures the importer country trade imbalance.

We are not the first to acknowledge the inter-temporal dependence of trade flows. The prominent dynamic gravity equation in the literature relies on the neo-classical theory of capital accumulation (e.g., Yotov and Olivero (2012); Alvarez (2017); and Anderson et al. (2020)), a fundamental source of persistence in all areas of macroeconomics. The key outcome of this theory, is the prediction that the trade persistence coefficient simply corresponds to the annual share of undepreciated capital stock. Looking at the empirical estimates of the capital depreciation rate, however, shows very little variation across countries and an average magnitude of around 10% (see IMF (2015)). This suggests an average trade persistence coefficient of around 0.9, which is at odds with our empirical results. We only manage to estimate a trade persistence coefficient of a similar magnitude from a dynamic gravity equation that dismisses all procedures that account for structural heterogeneity of different trade partnerships as is currently considered to be the norm. But for reasons discussed above, we advocate the opposite in order to avoid misleading inference.

The rest of this paper is organized as follows. In Section 2 we describe the primitives of our theoretical model and derive theoretically-consistent gravity equation. Section 3 contains the description of the data and the methodology of all panel regression techniques applied in this paper. We then present the coefficient estimates of all dynamic gravity equation specifications and compare them with those in the existing literature. Sections 4 and Section 5 presents prediction performance of our gravity equation and compares it to the alternatives. Finally, Section 6 summarizes and concludes, whereas Online Appendix collects all technical details and other supporting material.

[3] This includes the foremost theories, such as Eaton and Kortum (2002), Melitz (2003), Chaney (2008), and others. In fact, Arkolakis et al. (2012) demonstrate that they all generalise to a wide class of models with very different micro-foundations, but very similar aggregate outcomes.

## 2 Theoretical Model

The key ingredients of our theoretical model are as follows.[4] The world economy evolves over discrete time $t = \{0, 1, 2, ...\}$. There are many countries $i, j \in n = \{1, 2, ..., N\}$. Each country is populated by two types of interacting agents: (i) a representative consumer; and (ii) producers. There are two sequential stages of production in each country: (i) wholesale; and (ii) distribution. The wholesale sector in each country is populated by a unit mass of firms $\omega \in [0, 1]$. Each firm produces a unique variety using inelastically supplied labor subject to exogenous productivity shocks. All wholesale varieties are tradable, but consumers can only purchase composite goods from the distributors, who import all individual wholesale varieties from a particular country and aggregate them into composite bundles. Trade in wholesale varieties is bound by Samuelson's 'iceberg costs', which are not internalised by the distributors and instead fully passed onto the consumers. But distributors are subject to 'learning-by-importing' positive externality, such that they become more efficient at aggregating wholesale varieties from a particular country the more they imported from that country in the past, thereby recouping a fraction of the iceberg costs incurred *en route*.

### 2.1 Supply Side

**Technology.** Wholesale varieties from country $i$ sold to country $j$ are produced using *Ricardian* technology $m_{ij,t}(\omega) = z_{i,t}h_{ij,t}(\omega)$, where $z_{i,t}$ is stochastic labour productivity and $h_{ij,t}(\omega)$ are hours of labor. Wholesale varieties are imperfect substitutes. The distributor aggregates the amounts $m_{ij,t}(\omega)$ of wholesale varieties $\omega$ imported from source country $i$ into an infinitely-divisible composite good $x_{ij,t}$ consumed in destination $j$. The aggregation technology is CES:

$$x_{ij,t} = e_{ij,t}m_{ij,t}, \tag{2.1}$$

where

$$m_{ij,t} = \left[\int_0^1 m_{ij,t}(\omega)^{1-1/\eta} d\omega\right]^{1/(1-1/\eta)}, \tag{2.2}$$

such that $\eta > 1$ is a constant and $e_{ij,t} > 0$ reflects distributor productivity (i.e. 'efficiency').

**Learning-by-Importing.** We posit that distributor productivity is endogenous – distributors become more efficient at aggregating wholesale varieties originating from country $i$ the more they imported from that country in the past (i.e., learning-by-importing). Formally,

$$e_{ij,t} = x_{ij,t-1}^{\chi_{ij}}, \tag{2.3}$$

such that the marginal productivity gains are characterised by $\chi_{ij} = \partial \ln e_{ij,t} / \partial \ln x_{ij,t-1} > 0$.[5] In the standard CES technology, efficiency $e_{ij,t}$ is normalised to unity and distributors produce zero real value added, such that $\lim_{\chi_{ij} \to 0} x_{ij,t} = m_{ij,t}$. But more generally, the real value added of the distributor is strictly non-negative, such that $x_{ij,t} \geq m_{ij,t}$ when $\chi_{ij} \geq 0$.

[4] For full derivations and additional technical details refer to Appendix A.

[5] We assume multiplicative efficiency, because the standard gravity relationship $X_{ji} = \frac{Y_i Y_j}{d_{ji}}$ is multiplicative, where $Y_i$ and $Y_j$ are aggregate incomes and $X_{ji}$ are the bilateral trade flows. With additive efficiency, the analytical benchmark is not CES, but Translog (e.g., Novy (2013)).

**Iceberg Costs.** Our model features standard iceberg costs at the intensive margin $d_{ij} - 1 > 0$. We can therefore think of the distributor value added as a positive externality that (partially) abates these costs. In particular, let $P_{ij,t}(\omega) > 0$ denote the 'import' price of $m_{ij,t}(\omega)$. With iceberg costs, but otherwise efficient international arbitrage forces, we have $P_{ij,t}(\omega) = d_{ij} P_{ii,t}(\omega)$, where $P_{ii,t}(\omega)$ is the 'export' price. Then assuming that the distributor is perfectly competitive, it can easily be shown that the aggregate price index of the wholesale varieties (i.e., the price of the aggregate bundle $m_{ij,t}$) is given by

$$P_{ij,t} = \left[ \int_0^1 P_{ij,t}(\omega)^{1-\eta} d\omega \right]^{1/(1-\eta)}. \tag{2.4}$$

By the same token, under perfect competition, the break-even price of the composite good $x_{ij,t}$, henceforth denoted as $\tilde{P}_{ij,t}$, is proportional to $P_{ij,t}$, such that altogether we have

$$\tilde{P}_{ij,t} = \frac{d_{ij} P_{ii,t}}{e_{i,t}}. \tag{2.5}$$

Clearly, the cost of the composite good $x_{ij,t}$ at present is decreasing in the efficiency $e_{i,t}$, which in turn comes with greater trade volumes realised in the past (i.e., larger $x_{ij,t-1}$).

## 2.2 Demand Side

**Preferences.** Each destination $j \in n$ is populated by a representative consumer, who derives utility from consumption of the composite goods from each source country:

$$c_{j,t} = \left[ \sum_{i=1}^{N} x_{ij,t}^{1-1/\eta} \right]^{1/(1-1/\eta)}. \tag{2.6}$$

The aggregate cost of living is henceforth denoted as $P_{j,t} > 0$ (i.e., the price of $c_{j,t}$). In addition, the representative household is infinitely-lived with present discounted value of utility given by

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \ln\left(c_{j,t}\right), \tag{2.7}$$

where $\mathbb{E}_0$ denotes rational expectations and $\beta \in (0,1)$ is the time preference parameter.[6]

**Feasibility Constraint.** The representative consumer is subject to an indefinite sequence of budget constraints:

$$c_{j,t} + \mathbb{E}_t[\zeta_{j,t,t+1} b_{j,t+1}] \leq b_{j,t} + w_{j,t} h_j, \tag{2.8}$$

where $b_{j,t}$ is the net real stock of internationally-traded one-period bonds, $\zeta_{j,t,t+1}$ is the real price of the one-period bond (i.e., pricing kernel), $w_{j,t} = W_{j,t}/P_{j,t}$ is the real hourly wage rate, and $h_j = \sum_{j=1}^{N} \int_0^1 h_{ij,t}(\omega) d\omega$ are the inelastically supplied hours of labor.

[6] For simplicity, we assume that the international financial markets are complete. We use compact notation that suppresses the state of nature, which is a conventional approach in the business cycle literature. The expectations operators implicitly represent probability-weighted averages across all states of nature. See Schmidt-Grohé and Uribe (2017) pp. 92-93 for a more thorough description.

## 2.3 Closing the Model

Let $\bar{P}_{i,t}$ denote the GDP deflator in source country $i$, such that $Y_{i,t} = \bar{P}_{i,t} y_{i,t}$ is nominal GDP. Similarly, let $M_{ij,t} = P_{ij,t} m_{ij,t}$ denote the gross nominal trade flows.[7] Then the goods market for source country $i$ clears when

$$Y_{i,t} = \sum_{j=1}^{N} M_{ij,t}. \tag{2.9}$$

Similarly, we can write $C_{i,t} = P_{i,t} c_{i,t} = \sum_{j=1}^{N} X_{ji,t}$, such that the aggregate trade balance is the difference between total sales of wholesale varieties and total purchases of the composite goods: $NX_{i,t} = Y_{i,t} - C_{i,t}$. If we then define $\xi_{i,t} = NX_{i,t}/C_{i,t}$ as the trade balance share of consumption expenditure, it can easily be shown that aggregate consumption is proportional to aggregate income:

$$C_{i,t} = \Xi_{i,t} Y_{i,t}, \tag{2.10}$$

where $\Xi_{i,t} = 1/(1+\xi_{i,t}) \lesseqgtr 1$. As is now standard in the macro-trade literature, our model features short-run and long-run trade imbalances. The short-run trade imbalances emerge due to consumption smoothing of the representative households who trade riskless bonds internationally in the context of complete asset markets similar to Dix-Carneiro et al. (2021). Though not immediately apparent, asymmetries in learning-by-doing across countries (i.e., if $\chi_{ij} \neq \chi_{ji}$) give rise to long-run cross-country differences in import intensities.However, our model features a stationary equilibrium (i.e., a steady state), such that long-run trade imbalances are constant and sustained without violating the transversality condition as in Dekle et al. (2008) and Schmidt-Grohé and Uribe (2017).

## 2.4 General Equilibrium

General equilibrium is a set of dynamic processes of allocations $\{m_{ij,t}(\omega), x_{ij,t}, c_{i,t}, b_{i,t+1}, y_{i,t}\}_{t=0}^{\infty}$ and prices $\{P_{ij,t}(\omega), \tilde{P}_{ij,t}, \bar{P}_{ij,t}, \zeta_{i,t,t+1}, W_{i,t}\}_{t=0}^{\infty}$ conditional on parameters $\{d_{ij}, \chi_{ij}, \eta, \beta\}$, a *numéraire* chosen arbitrarily, and exogenous shocks to productivity $\{z_{i,t}\}_{t=0}^{\infty}$ for all countries $i, j \in n$ and varieties $\omega \in [0,1]$. It further satisfies: (i) utility-maximizing behavior of the representative households; (ii) goods and labour market clearing conditions; and (iii) all firms in each sector – wholesale and distribution – break-even. A brief summary of the equilibrium conditions is as follows.

**Consumption Smoothing.** Representative consumers choose $c_{j,t}$ and the net stock of bonds $b_{j,t}$ to maximize utility (2.7) subject to a sequence of budget constraints (2.8), taking the price of bonds $\zeta_{j,t,t+1}$ and the wage bill $w_{j,t} h_j$ as given. Assuming that $\Lambda_{j,0} > 0$ and $\Lambda_{i,0} > 0$ are well-defined initial conditions, the first-order conditions lead to the standard Euler equation and the perfect consumption risk sharing relationships, respectively, that summarise consumption smoothing:

$$1 = \beta \mathbb{E}_t \left[ \frac{c_{j,t}}{\zeta_{j,t,t+1} c_{j,t+1}} \right], \tag{2.11}$$

$$\Lambda_{j,0} \zeta_{j,0,t} = \Lambda_{i,0} \zeta_{i,0,t}. \tag{2.12}$$

[7] Incidentally, the gross trade flows $M_{ij,t} = P_{ij,t} m_{ij,t}$ in our model are identical to $X_{ij,t} = \tilde{P}_{ij,t} x_{ij,t}$. To see this, recall that the distributor technology implies $x_{ij,t} = e_{ij,t} m_{ij,t}$ and the break-even price index is $\tilde{P}_{ij,t} = P_{ij,t}/e_{ij,t}$. Hence, calculating nominal GDP using either the gross output of the distribution sector $X_{ij,t}$ or the wholesale sector

**Composite Consumption.** The representative consumer chooses the amount of composite goods to purchase from the distributor $x_{ij,t}$ to minimize the total expenditure $\sum_{i=1}^{N} \tilde{P}_{ij,t}x_{ij,t}$, subject to the available budget of $P_{j,t}c_{j,t}$ and subject to CES preferences (2.6). The first order conditions characterise the optimal demand for composite goods:

$$x_{ij,t} = c_{j,t} \left( \frac{\tilde{P}_{ij,t}}{P_{j,t}} \right)^{-\eta}. \tag{2.13}$$

**Wholesale Imports.** The distributor chooses the amount of wholesale varieties to purchase $m_{ij,t}(\omega)$ to minimize the total expenditure on wholesale imports $\int_0^1 P_{ij,t}(\omega)m_{ij,t}(\omega)d\omega$ subject to the aggregate cost of producing a composite good $\tilde{P}_{ij,t}x_{ij,t}$ and CES technology (2.2). The first order conditions characterise the optimal demand for wholesale varieties:

$$m_{ij,t}(\omega) = \left[ \frac{P_{ij,t}(\omega)}{P_{ij,t}} \right]^{-\eta} \frac{x_{ij,t}}{e_{ij,t}}. \tag{2.14}$$

This concludes the description of the primitives. We now turn to the description of how the equilibrium conditions of our theoretical model can be used to derive a dynamic gravity equation.

### 2.5 Dynamic Gravity Equation

The workhorse model used to predict bilateral trade flows across space and time is the *gravity* equation. To derive the gravity equation in our model, we follow the footsteps of Anderson and van Wincoop (2003), who are the first to micro-found a gravity equation in an *Armington*-type general equilibrium model similar to ours. The key difference in our model is the presence of learning-by-importing externality in sourcing wholesale varieties, which introduces lagged trade flows as a relevant determinant of contemporaneous trade flows. By contrast, the standard gravity equation is static. The key analytical similarities and differences are highlighted in the following results.

**Proposition 1.** *Let $\theta_{i,t} = Y_{i,t}/Y_t$ measure the relative size of the source country, where the size of the entire world economy is measured as $Y_t = \sum_{j=1}^{N} Y_{j,t}$. Then the share of the source country aggregate income relative to the world income is given by*

$$\theta_{i,t} = (\Phi_{i,t}P_{ii,t})^{1-\eta}, \tag{2.15}$$

*where $P_{ii,t} = W_{i,t}/z_{i,t}$ stands for the 'export' price, whereas*

$$\Phi_{i,t} = \left[ \sum_{j=1}^{N} \theta_{j,t}\Xi_{j,t} \left( \frac{d_{ij}x_{ij,t-1}^{-\chi_{ij}}}{P_{j,t}} \right)^{1-\eta} \right]^{1/(1-\eta)} \tag{2.16}$$

*is the so-called 'outward' multilateral resistance (i.e., source country's market access abroad).*

*Proof.* See Appendix B.1. □

$M_{ij,t}$ gives identical results. The only discernible difference comes from real value added (i.e. $m_{ij,t}$ and $x_{ij,t}$).

**Lemma 1.** *The gravity equation is dynamic when there is learning-by-importing, such that $\chi_{ij} \in (0,1)$ for all $i \in n \setminus j$. And when learning-by-importing is asymmetric across countries, such that $\chi_{ij} \neq \chi_{ji}$ for all $i \in n \setminus j$, and/or the iceberg costs are asymmetric, such that $d_{ij} \neq d_{ji} > 1$ for all $i \in n \setminus j$, the gravity equation features the multilateral trade imbalances:*

$$A_{ij,t} = \Xi_{j,t} \left[ \frac{d_{ij}}{\Phi_{i,t} P_{j,t}} \right]^{1-\eta} \prod_{s=1}^{S} \left( \frac{d_{ij} Y_{i,t-s}^{\eta/(1-\eta)}}{\Phi_{i,t-s} A_{ij,t-s} Y_{j,t-s} Y_{t-s}^{\eta/(1-\eta)}} \right)^{\chi_{ij}^{s}(1-\eta)}, \tag{2.17}$$

*where $A_{ij,t} := (X_{ij,t} Y_t)/(Y_{i,t} Y_{j,t})$ measures the 'size-adjusted' bilateral trade flows.*

*Proof.* See Appendix B.2. □

Intuitively, a favourable productivity shock increases the contemporaneous source country exports to the rest of the world. At the same time, it increases imports to the booming source country as it becomes larger. Even if the productivity shock is transitory, learning-by-importing externalities generate a lasting reduction in the relative size of the iceberg costs internalised by the distributors in the source and destination countries. As such, it takes time for the spell of elevated trade flows to fade after the incidence of the shock. Asymmetries in shock exposure and learning-by-importing externalities create wedges in trade intensities across countries, which are in equilibrium mediated by international trade of assets and ultimately result in trade imbalances.

Taking logs on both sides of (2.17) gives the following log-linear specification:

$$\begin{aligned}
\ln A_{ij,t} = {} & \underbrace{(\eta-1)\sum_{s=1}^{S}\chi_{ij}^{s}\ln A_{ij,t-s}}_{\text{size-adjusted bilateral trade flows}} + \underbrace{\ln \Xi_{j,t}}_{\text{destination trade imbalance}} \\
& \underbrace{-(\eta-1)\sum_{s=0}^{S}\chi_{ij}^{s}\ln d_{ij}}_{\text{bilateral resist.}} + \underbrace{(\eta-1)\sum_{s=0}^{S}\chi_{ij}^{s}\ln \Phi_{i,t-s}}_{\text{source outward mult.'l resist.}} + \underbrace{(\eta-1)\ln P_{j,t}}_{\text{destination inward mult.'l resist.}} \\
& \underbrace{-\eta\sum_{s=1}^{S}\chi_{ij}^{s}\ln Y_{t-s}}_{\text{world size}} + \underbrace{\eta\sum_{s=1}^{S}\chi_{ij}^{s}\ln Y_{i,t-s}}_{\text{source size}} + \underbrace{(\eta-1)\sum_{s=1}^{S}\chi_{ij}^{s}\ln Y_{j,t-s}}_{\text{destination size}}.
\end{aligned} \tag{2.18}$$

According to the dynamic gravity equation above, trade intensifies (i.e., $A_{ij,t}$ increases) with: (i) lagged trade flows $A_{ij,t-s}$; (ii) destination country's trade deficit $\Xi_{ij,t}$; (iii) (lagged) source and destination country multilateral resistance $\Phi_{i,t-s}$ and $P_{j,t-s}$, respectively; (iv) and lagged source and destination country size $Y_{i,t-s}$ and $Y_{j,t-s}$, respectively. By the same token, trade slows down (i.e., $A_{ij,t}$ falls) with: (i) iceberg costs $d_{ij}$; and (ii) the lagged size of the world economy $Y_{t-s}$.

**Corollary 1.** *Without learning-by-importing, such that $\chi_{ij} \to 0$ for all $i \in n \setminus j$, the gravity equation is static à la Anderson and van Wincoop (2003):*

$$\lim_{\chi_{ij}\to 0\,\forall\, i\in n\setminus j} \ln A_{ij,t} = (1-\eta)\left[\ln d_{ij} - \ln \Phi_{i,t} - \ln P_{j,t}\right], \tag{2.19}$$

*since $\lim_{\chi_{ij}\to 0\,\forall\, i\in n\setminus j} \ln \Xi_{j,t} = 0$ assuming that iceberg costs are symmetrical, such that $d_{ij} = d_{ji}$, which additionally implies that $\lim_{\chi_{ij}\to 0\,\forall\, i\in n\setminus j} \Phi_{i,t} = P_{i,t}$.*

**Corollary 2.** *With symmetrical learning-by-importing and iceberg costs across countries, such that $d_{ij} = d_{ji}$ and $\chi_{ij} \to \chi > 0$ for all $i \in n \setminus j$, the gravity equation is dynamic, but all bilateral trade flows are balanced, such that*

$$\begin{aligned}
\lim_{\chi_{ij} \to \chi \, \forall \, i \in n \setminus j} \ln A_{ij,t} =& (\eta - 1) \sum_{s=1}^{S} \chi^s \ln A_{ij,t-s} \\
& - (\eta - 1) \sum_{s=0}^{S} \chi^s \ln d_{ij} + (\eta - 1) \sum_{s=0}^{S} \chi^s \ln \Phi_{i,t-s} + (\eta - 1) \ln P_{j,t} \\
& - \eta \sum_{s=1}^{S} \chi^s \ln Y_{t-s} + \eta \sum_{s=1}^{S} \chi^s \ln Y_{i,t-1} + (\eta - 1) \sum_{s=1}^{S} \chi^s \ln Y_{j,t-1}.
\end{aligned} \tag{2.20}$$

*since* $\lim_{\chi_{ij} \to \chi \, \forall \, i \in n \setminus j} \ln \Xi_{j,t} = 0$ *under the assumption that* $d_{ij} = d_{ji}$*.*

There are several features of the learning-by-importing-augmented gravity equation that stand out compared to the alternatives in the existing literature. First, it features heterogeneous parameters. Consequently, common shocks lead to heterogeneous adjustment of bilateral trade flows. Second, learning-by-importing generates on average greater, dynamic, and heterogeneous coefficient next to the bilateral iceberg costs $d_{ij}$, commonly known as the "trade elasticity" (i.e., $(\eta - 1) \sum_{s=1}^{S} \chi_{ij}$).[8] While we do not discuss or quantify the resulting welfare gains from trade in this paper, it nonetheless hints at the fact that the *Armington* elasticity $\eta$ is not necessarily a catch-all parameter as argued by Arkolakis et al. (2012) for instance. Third, learning-by-importing generate a dynamic outward multilateral trade resistance term $\Phi_{i,t-s}$.[9]. By contrast, the static gravity equation contains only the contemporaneous multilateral trade resistance $\Phi_{i,t}$ (see Corollary 1). This leads to a fundamentally different transmission of shocks as they reflect contemporaneous and lagged unobserved common factors that would not be captured by time fixed effects. And fourth, differences in bilateral efficiencies due to learning-by-importing generate endogenous multilateral trade imbalances.[10]

[8] Fieler (2011), Novy (2013), and Carrere et al. (2020) consider alternative theories of heterogeneous trade elasticities in the static environments. Boehm et al. (2020) and Alessandria et al. (2021) present estimates of homogeneous dynamic trade elasticities. However, learning-by-importing offers a framework with both dynamics and heterogeneity.

[9] This result is consistent with a number of studies that explore the persistence of trade costs during the period of hyper-globalization (e.g., Anderson and van Wincoop (2004); Baldwin and Taglioni (2006); Disdier and Head (2008); Zwinkels and Beugelsdijk (2010); Head and Mayer (2014))

[10] Davis and Weinstein (2002) and Dekle et al. (2008) assume that trade imbalances are constant and exogenous.

# 3 Empirical Analysis

In this section of the paper, we describe the data and the methodology used to estimate the learning-by-importing-augmented gravity equation, summarized in (2.18). The theory of learning-by-importing requires to account for a lagged dependent variable, time-invariant heterogeneity in parameters, and a dynamic structure of unobservable factors (dynamic multilateral resistance terms) in the empirical model. We start by defining empirical counterparts of theoretical variables and describing the data. We then move on to the mapping of the theoretical model into an empirical setup. Just like our theory nests static and dynamic alternatives of gravity models, similarly our empirical model nests various estimators, depending on the assumptions on parameters and the error structure. Once we establish the baseline coefficient estimates, we discuss the resulting implications with reference to competing theories in the existing trade literature. We refer the reader to Appendix E for an in-depth discussion on the advantages and disadvantages of several different methodologies in the existing panel data literature and more details for the case of our preferred baseline model specification.

## 3.1 Data

The data used to estimate the dynamic gravity equation are displayed in Table 1. All time series are mapped directly to the variables in the theory of learning-by-importing. The value of size-adjusted bilateral trade flows ($\text{FLOW}_{ij,t}$) represents the dependent variable, explicitly defined in equation (2.17). Consistent with the theory, multilateral trade imbalance ($\text{TB}_{j,t}$) is measured as the reciprocal of the gross net export share in private consumption expenditure. As is usual in the trade literature, aggregate income in the source country ($\text{GDP}_{i,t}$), destination country ($\text{GDP}_{j,t}$), and world economy ($\text{GDP}_t$) is measured by the nominal gross domestic product in each location.

**Table 1: Data Description**

| Variable height | Data | Description | Measurement Units |
|---|---|---|---|
| $\ln(A_{ij,t})$ | $\text{FLOW}_{ij,t}$ | Size-Adjusted Bilateral Trade Flows | U.S. dollars, Millions |
| $\ln(\Xi_{j,t})$ | $\text{TB}_{j,t}$ | Multilateral Trade Imbalance | Gross Share, Percent |
| $\ln(Y_{i,t})$ | $\text{GDP}_{i,t}$ | Source Country Aggregate Income | U.S. dollars, Millions |
| $\ln(Y_{j,t})$ | $\text{GDP}_{j,t}$ | Destination Country Aggregate Income | U.S. dollars, Millions |
| $\ln(Y_t)$ | $\text{GDP}_t$ | World Aggregate Income | U.S. dollars, Millions |

*Data Sources*: Penn World Tables 9.1 by Feenstra et al. (2015), IMF Direction of Trade Statistics (DOTS) Database, World Bank Database.

*Data Coverage*: the data cover the period 1950-2014 for 39 countries including both advanced and emerging markets, namely: Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, China, Colombia, Cyprus, Denmark, Egypt, Finland, France, Germany, Greece, Iceland, India, Ireland, Israel, Italy, Japan, Luxembourg, Mexico, Morocco, the Netherlands, New Zealand, Norway, Peru, Philippines, Portugal, South Africa, Spain, Sweden, Switzerland, Turkey, Great Britain, the United States, and Venezuela.

### 3.2 Methodology

The empirical adaptation of the learning-by-importing-augmented gravity equation (2.17) is a large $N$ and large $T$ panel regression model. Specifically, we capture temporal variation over $t = 1, 2, ..., T$ and also spatial variation across the source country $i = 1, 2, ..., N$ and the destination country $j = 1, 2, ..., N-1$, such that $j \neq i$. Formally, our theoretical gravity equation (2.18), letting $S = 1$, is mapped into the following estimating model:

$$\ln A_{ij,t} = \beta_0 + \mathbf{x}'_{ij,t}\boldsymbol{\beta}_{ij} + u_{ij,t}, \tag{3.1}$$

$$u_{ij,t} = \boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t + \varepsilon_{ij,t}, \tag{3.2}$$

$$\mathbf{x}_{ij,t} = \boldsymbol{\gamma}'_{ij}\boldsymbol{\phi}_t + \boldsymbol{\nu}_{ij,t}, \tag{3.3}$$

where $\ln A_{ij,t} := \text{FLOW}_{ij,t}$ are the size-adjusted trade flows on a logarithmic scale, $\boldsymbol{\beta}_{ij} = [\beta_{1ij}, \beta_{2ij}, ..., \beta_{5ij}]'$ is a $5 \times 1$ vector of coefficients, $\mathbf{x}_{ij,t} = [\text{FLOW}_{ij,t-1}, \text{TB}_{j,t}, \text{GDP}_{i,t-1}, \text{GDP}_{j,t-1}, \text{GDP}_{t-1}]'$ is a $5 \times 1$ vector of all common and country-specific observable variables. The empirical counterparts of the theory-implied variables are $\ln(\Xi_{j,t}) := \text{TB}_{j,t}$, $\ln Y_{i,t-1} := \text{GDP}_{i,t-1}$, $\ln Y_{j,t-1} := \text{GDP}_{j,t-1}$, $\ln Y_{t-1} := \text{GDP}_{t-1}$ while $\boldsymbol{\phi}_t$ and $\boldsymbol{\lambda}_{ij}$, $\boldsymbol{\gamma}_{ij}$ represent some configuration of the unobservable vector of dynamic common factors (inward and outward multilateral resistances in our case) and country-pair-specific vectors of parameters (known as factor loadings), respectively. Effectively, our approach extends the interactive fixed effects representation of Bai (2009) into a three-dimensional data structure. We focus on the log-linear representation because there are no known econometric tools that could be used for the dynamic (i.e., when the lagged dependent variable is part of regressors), heterogeneous (i.e. when parameters are trade pair specific), and time-varying unobserved terms driven nonlinear panel model.[11] We further elaborate on the theory and empirical representation in Appendix C.

It is well-known from the seminal contribution by Anderson and van Wincoop (2003) that controlling for the multilateral resistance terms is crucial for the empirical gravity applications. The most popular choice is the use of fixed effects. Econometrically, that is nothing else but the use of the multi-way error components, as in our error term (3.2), to control for the unobservable terms. In the static framework, they are directional (source and destination) fixed effects whereas in the panel data they usually include country-time fixed effects. To see how we complement that tradition more clearly, we can expand the error structure as follows:

$$\begin{aligned}\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t &= \begin{bmatrix} \lambda_{1ij} & \lambda_{2ij} & \lambda_{3ij} & \lambda_{4ij} \end{bmatrix} \begin{bmatrix} 1 \\ \phi_{j,t} \\ \phi_{i,t} \\ \phi_{i,t-1} \end{bmatrix} \\ &= \begin{bmatrix} -(1+\chi_{ij})(\eta-1) & (\eta-1) & (\eta-1) & \chi_{ij}(\eta-1) \end{bmatrix} \begin{bmatrix} \ln d_{ij} \\ \ln P_{j,t} \\ \ln \Phi_{i,t} \\ \ln \Phi_{i,t-1} \end{bmatrix}.\end{aligned} \tag{3.4}$$

[11] The Poisson model, which turned out be an indispensable empirical tool for the static gravity specifications, cannot be invoked in this case since averaging, dealing with lagged trade values, time-varying unobserved factors,

More precisely, our key identifying assumption is the interactive representation of unobservable static and dynamic terms, as given above. We have thus mapped our heterogeneous parameters and unobservable trade costs, bilateral, multilateral and time-varying, into, what is known as, heterogeneous factor loadings $\boldsymbol{\lambda}_{ij}$, and factors $\boldsymbol{\phi}_t$.[12] In other words, we have subsumed all trade costs, bilateral, multilateral and time-varying, within the term $\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t$ as neither are fully observable. The literature sometimes uses proxies for, say, bilateral trade costs by employing distance, bilateral trade treaties or trade agreements, the standard empirical approach, however, remains to be the use of fixed effects, which can be cast as a part of the multi-way error term. Finally, the error terms $\varepsilon_{ij,t}$ and $\boldsymbol{\nu}_{ij,t}$ are assumed to be independently distributed of each other, uncorrelated with the unobservable unobserved factors, and uncorrelated across country pairs.

Notice that country-time fixed effects can arise as a special case (3.4): letting $\lambda_{ij}$ be trade pair invariant and removing dynamics from the $\boldsymbol{\phi}_t$ (i.e., abstracting from $\ln \Phi_{i,t-1}$), we are back to homogeneous country-time fixed effects. They postulate that each trade pair gets impacted by $\ln P_{j,t}$ and $\ln \Phi_{i,t}$, empirically captured by country-time fixed effect, with the identical intensity. That is not compatible with our theory, which stipulates country-pair specific intensity stemming from the unobservable time-varying terms. Luckily, our model encompasses many competing estimators, which are nested as a special case of equations (3.1), (3.2), and (3.3) by: (i) choosing an estimator of $\boldsymbol{\beta}_{ij}$; and (ii) imposing restrictions on the inner product of $\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t$. We discuss each estimation strategy discussed in detail in Section E. In short, the configuration of $\boldsymbol{\phi}_t = [1, 1, \tau_t]'$ and $\boldsymbol{\lambda}_{ij} = [\lambda_i, \lambda_j, 1]'$ gives rise to the traditional fixed effects (FE) error structure of Feenstra (2016), usually applied to static gravity models, adjusted for the time fixed effect, namely $u_{ij,t} = \lambda_i + \lambda_j + \phi_t + \varepsilon_{ij,t}$, where $\lambda_i$ and $\lambda_j$ are the country fixed effects and $\phi_t$ is the time fixed effect. However, FE pools the regressions coefficients by averaging $\mathbf{x}_{ij,t}$ across all source and destination countries.[13] By contrast, the so-called mean group (MG) estimator ignores unobserved heterogeneity in the error term, but it preserves parameter heterogeneity between different country pairs.[14]

A more flexible alternative, coined by us as the hybrid fixed effects (HFE), combines the FE and the MG approaches, by imposing $\boldsymbol{\phi}_t = [1, 1]'$ and $\boldsymbol{\lambda}_{ij} = [\lambda_i, \lambda_j]'$. Consequently, the only difference between the MG and the HFE approaches is that the error structure is now given by $u_{ij,t} = \lambda_i + \lambda_j + \varepsilon_{ij,t}$, while the regression coefficients $\boldsymbol{\beta}_{ij}$ are estimated using the standard MG approach. By contrast, the other common variations of the FE approach adopted in the literature rely exclusively on the pooled coefficient estimator as does the conventional FE approach, but their

---

and heterogeneous parameters are substantially harder in the nonlinear setup (as is Poisson) rather than in the linear environment.

[12]With a slight abuse of notation, we merged a lagged factor and subsumed country-specific dynamic factors within the expression of $\boldsymbol{\phi}_t$ to save on space. Notice that factor loadings and factors are not separately uniquely identified without additional restrictions due to rotational indeterminacy. For our purposes, however, $\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t$ will act as nuisance and we will not be interested in the identification of separate components. Our econometric approach takes care of the fact that $\boldsymbol{\phi}_t$ includes dynamic components and factors are constructed using country $i$ and $j$ variables for the trade flow between $i$ and $j$, as summarized in (3.4).

[13]The FE estimator is such that $\boldsymbol{\beta} = (\bar{\mathbf{x}}'_t\bar{\mathbf{x}}_t)^{-1}\bar{\mathbf{x}}'_t\bar{\mathrm{y}}_t$ is homogeneous for all $i, j \in n$, where $\bar{\mathbf{x}}_t = 1/\bar{N}\sum_{ij=1}^{\bar{N}} \mathbf{x}_{ij,t}$ and $\bar{\mathrm{y}}_t = 1/\bar{N}\sum_{ij=1}^{\bar{N}} \mathrm{y}_{ij,t}$ denote cross-sectional averages, while $\bar{N} = (N-1)N$ measures the number of unique country pairs (i.e., dyads).

[14]More technically, the MG estimator nullifies the inner product $\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t = 0$, ignoring unobserved heterogeneity in the error term (see (3.2) and (3.4)), but it allows parameters to vary between different country pairs, such that $\boldsymbol{\beta}_{ij} = (\mathbf{x}'_{ij,t}\mathbf{x}_{ij,t})^{-1}\mathbf{x}'_{ij,t}\mathrm{y}_{ij,t}$ and the inference is drawn from the cross-sectional average of the MG coefficient estimates $\boldsymbol{\beta} = 1/\bar{N}\sum_{ij=1}^{\bar{N}} \boldsymbol{\beta}_{ij}$.

differences arise from the specification of the multi-way error structure. Specifically, we consider three additional FE alternatives: FE2 imposes $\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t = \phi_{i,t} + \phi_{j,t}$, i.e., country-specific time fixed effects, imposing unitary factor loadings, as is often done in the gravity applications with panel data (see Feenstra (2016) or Head and Mayer (2014)), FE3 imposes pair-specific time-invariant fixed effect and common time fixed effect, $\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t = \lambda_{ij} + \phi_t$, and FE4 imposes pair-specific fixed effect and country-specific time fixed effects, $\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t = \phi_{i,t} + \phi_{j,t} + \lambda_{ij}$.[15] To summarize, FE2 controls for country-time fixed effects, FE3 for trade pair and (homogeneous) time fixed effects, FE4 for trade pair and country-time fixed effects. All of them arise as particular cases of our proposed model.

Our preferred, and baseline, empirical model, called the common correlated effects mean group (CCEMG) approach, is the most flexible version, subsuming alternative popular versions in the literature as special cases. It is a more robust version of the principal components estimator that works well for dynamic panel models with large time and cross-sectional dimensions. Instead of postulating identical parameters as do country-time fixed effects, the procedure allows for trade pair specific parameters. And to capture common time trends, as in time fixed effects estimator, it uses common variation in dependent and independent variables, similarly to the principal components estimator. Using our notation, the CCEMG imposes $\boldsymbol{\phi}_t = [1, 1, \mathbf{z}'_t]'$ and $\boldsymbol{\lambda}_{ij} = [\lambda_i, \lambda_j, \boldsymbol{\lambda}'_{ij}]'$, where $\mathbf{z}_t = [\bar{\mathrm{y}}_t, \bar{\mathbf{x}}'_t]'$ is the vector of unobserved common factors (capturing theoretical multilateral resistance terms), multiplied by the trade pair specific parameters. The unobserved factors are proxied by the cross-sectional averages of the dependent and independent variables pair-wise: $\bar{\boldsymbol{y}}_{\boldsymbol{i}} = 1/N \sum_{ij=1}^{N} \boldsymbol{y}_{ij}$ and $\bar{\boldsymbol{x}}_{\boldsymbol{i}} = 1/N \sum_{ij=1}^{N} \boldsymbol{x}_{ij}$. The cross-sectional averages include the pair of interest $ij$, making them also a representation of global factors (see Chudik et al. (2011) among others). With the large time and cross-sectional dimension, we can uncover common components by taking simple averages of included variables relying on the law of large numbers (see Pesaran (2006) for the seminal contribution justifying such an approach). Importantly, our dynamic factor structure is well-accommodated by including lagged cross-sectional averages, as deliberated by Chudik and Pesaran (2015).[16] What

[15] As covered in the Appendix E.2, another estimator we consider, and one of the alternatives to our baseline empirical model, is the so-called AMG estimator as in Eberhardt and Teal (2013). It sets $\boldsymbol{\phi}_t = [1, 1, \hat{\phi}_t]'$ and $\boldsymbol{\lambda}_{ij} = [\lambda_i, \lambda_j, \lambda_{ij}]'$, where $\hat{\phi}_t$ are the pre-estimated time fixed effects from the standard FE regression model. Notice that $\lambda_{ij}$ is restricted to equal unity in the FE approach, such that the time fixed effects exert a homogeneous factor loading across all country pairs, but AMG relaxes this assumption, such that the error structure is given by $u_{ij,t} = \lambda_i + \lambda_j + \lambda_{ij}\hat{\phi}_t + \epsilon_{ij,t}$ and the time fixed effects exert a heterogeneous response for each country pair. The AMG estimator of $\boldsymbol{\beta}_{ij}$ after the pre-estimation is analogous to the MG approach. The main drawback of AMG resides precisely in having an homogeneous first step, inconsistent to the fact that our panel regression model is inherently dynamic with heterogeneous coefficients.

[16] Analogous to Pesaran (2006), equation (3.3) justifies the use of cross-sectional averages of $\mathbf{x}_{ij,t}$ to proxy the unknown factors in $u_{ij,t}$. This is because ordinary least squares applied to equations (3.1) and (3.2) generally delivers biased and inconsistent coefficient estimates whenever the unobservable common factors $\boldsymbol{\phi}_t$ are correlated with the regressors $\mathbf{x}_{ij,t}$. And this constraint generally binds, since the learning-by-importing-augmented gravity equation (2.18) predicts that bilateral trade flows depend on the inward and outward multilateral resistance, which in turn are functions of bilateral trade flows to and from all trade partners, respectively (see equation (2.16)). However, we explicitly control for the endogeneity between the unobservable common factors and the regressors by assuming that the regressors are generated by equation (3.3). The regressors are thus projected onto their cross-sectionally weighted averages, which renders the coefficient estimates consistent under general assumptions set out by Pesaran (2006) and Chudik and Pesaran (2015). In particular, a sufficient number of lagged values of cross-sectional averages must also be included. Despite similar growth rates of countries and time periods within each country panel (trade partners over time for a fixed source economy), we apply "half-panel" jackknife correction to reduce a bias in the persistence parameter for the aggregate model for the trade pairs. This correction performs very well in substantially smaller samples than ours (see Monte Carlo evidence in Chudik and Pesaran (2015)).

is more, as recent literature shows, the estimator is robust even if the time dimension is limited and the nature of unobserved common shocks (factors) is either deterministic or stochastic (Westerlund (2018); Westerlund et al. (2019)).

Once again, it is important to stress that the interactive fixed effect approach complements and generalizes country time fixed effects estimator (e.g., FE2 or FE4), which dominate the existing literature, and they emerge as a special case in our setting. This result will re-emerge when comparing model performance. In other words, instead of adding country-specific time fixed effects with unitary parameters, we can allow for a richer structure, predicted by our theory, suggesting heterogeneous loadings, as summarized in (3.4). In addition to the more flexible treatment of dynamic unobserved components, capturing trade resistance terms, we also explore the importance of parameter heterogeneity in our inference using the so-called common correlated effects pooled (CCEP) approach. Panel gravity models ignore structural heterogeneities due to limited time dimension as well as lack of theory, pointing to key structural differences among economies. The CCEP approach adopts an identical error structure to the CCEMG estimator, but applies the pooled (homogeneous) regression coefficient estimator analogous to the FE approach. And finally, we consider the construction of the unobserved dynamic terms by exploiting the gravity structure, i.e., only using contemporaneous and lagged trade flows to proxy for the common shocks albeit acting with the trade pair specific intensity. We call such an estimator as common correlated effects restricted (CCEMGR) approach.[17] This means that the error structure of the CCEMG, CCEMGR, and CCEP approaches is given by $u_{ij,t} = \lambda_i + \lambda_j + \mathbf{z}_t'\boldsymbol{\lambda}_{ij} + \varepsilon_{ij,t}$, which distinguishes between unobservable country-specific time-invariant heterogeneity captured by $\lambda_i$ and $\lambda_j$ as well as unobservable time-varying heterogeneity specific to each country pair captured by the inner product of $\mathbf{z}_t'\boldsymbol{\lambda}_{ij}$. Our strategy thus provides a neat way to nest various theoretical models into a unified empirical framework for the panel data.

### 3.3 Coefficient Estimates

Consider the coefficient estimates of the learning-by-importing-augmented gravity equation presented in Table 2. Each column displays the values of the coefficient estimates that are obtained using one of the different techniques described above and covered in detail in Appendix E.[18] Our preferred baseline model specification, the CCEMG, is presented in column (1). It incorporates the time-invariant heterogeneity specific to each country pair, controls for the dynamic structure of the unobservable common factors, and also estimates the regression coefficients for each country pair in order to retain the cross-sectional parameter heterogeneity. The ordering of the other columns follows this reasoning: estimated homogeneous coefficients and no common factors (FE), heterogeneous coefficients and no common factors (MG), homogeneous coefficients but includes common

[17] It simply removes $\bar{\mathbf{x}}_t$ from the proxied unobserved common factors $\mathbf{z}_t$ and replaces it with $[\bar{y}_t, \bar{y}_{t-1}]'$.

[18] We do not cover Poisson regressions in the main text as they are not developed for heterogeneous parameters, dynamic unobserved factors and cannot account for zero trade flows entering as explanatory variables. What is more, Poisson model cannot be nested in the system (3.1)-(3.3) due to the inherent nonlinear structure. However, to be aligned with the literature, the additional results generated using other variations of the FE and Poisson pseudo-maximum-likelihood (PPML) approaches are relegated to Table 4 in Appendix E.4. The PPML1-4 have the same characteristics for the fixed effects as in FE1-4. In all cases, sticking to the existing Poisson model, zero lagged trade flows are dropped, thereby losing the key advantage of the PPML, parameters are homogeneous, and dynamic unobserved trends are ignored, all in violation of our theory-predicted gravity model.

factors (CCEP), a restricted version of the common factor where only trade flows data are used (CCEMGR).[19,20]

For direct comparability reasons, Table 2 displays only the pooled or the cross-sectionally averaged coefficient estimates, but we demonstrate and discuss the extent of parameter heterogeneity of our preferred baseline model specification in Section 4.2. The first line of Table 2 presents the coefficient estimates associated with the lagged dependent variable ($\text{FLOW}_{ij,t-1}$), which we define as the "trade persistence coefficient" (i.e., pooled or averaged $\beta_{1ij}$ in equation (3.1)). First, the trade persistence coefficient is significantly different from zero and unity for all seven different techniques. This implies that following a random shock, trade flows generally revert back to the trend gradually rather than instantaneously as is implied by the static gravity equation due to Anderson and van Wincoop (2003). Second, and more importantly, our estimates demonstrate a remarkable difference between the standard FE approach, which generates a pooled trade persistence coefficient estimate of 0.91 that is homogeneous for all country pairs (see column (2) in Table 2), and all other techniques that retain estimated parameter heterogeneity and/or incorporate some measure of the unobservable common factors. This implies that following a random shock, trade flows revert back to the trend at a considerably faster rate than suggested by the neo-classical gravity equation pioneered by Yotov and Olivero (2012); Anderson et al. (2020). We draw particular attention to the value of the trade persistence coefficient because our theory identifies the heterogeneity of learning-by-importing specific to each country pair $\chi_{ij}$ from the trade persistence coefficient $\beta_{1ij} = \chi_{ij}(\eta - 1)$. Specifically, $\beta_1$ is mapped directly to $1/\bar{N}\sum_{ij=1}^{\bar{N}} \chi_{ij}(\eta - 1)$ in equation (2.18), where $\eta > 0$ and $\bar{N} = N(N-1)$ measures the total number of unique country pairs in our sample.

What is the relative importance of controlling for the unobservable common factors and retaining parameter heterogeneity? If we expend (retain) the unobservable common factors, but retain (expend) parameter heterogeneity following the MG (CCEP) approach, the trade persistence coefficient estimate is equal to 0.54 (0.37) (see columns (3) and (4) in Table 2). Recall that following the FE approach, which expends unobservable common factors and applies the pooled coefficient estimator, the trade persistence coefficient estimate is 0.91. By contrast, our preferred CCEMG approach that retains parameter heterogeneity and incorporates the unobservable common factors generates a cross-sectionally averaged trade persistence coefficient estimate of 0.35. The trade persistence therefore disappears when controlling for the unobservable common factors and/or retaining parameter heterogeneity, since they both lead to a significantly lower trade persistence coefficient than predicted by the conventional FE approach. It means that in the presence of common, albeit acting at a pair-specific intensity, shocks, trade adjusts considerably more abruptly than would be predicted by standard techniques and in line with the empirical evidence in Appendix D.

The second line of Table 2 presents the coefficient estimates associated with the multilateral trade imbalance in the destination country ($\text{TB}_{j,t}$), which we define as the "trade imbalance coefficient" (i.e., pooled or averaged $\beta_{2ij}$ in equation (3.1)). As shown in equations (2.17) and (2.18), multilateral trade imbalance enters the dynamic gravity equation with unitary elasticity, such that

[19] The results presented for the full sample are broadly confirmed if we look at pre-Global Financial Crisis (GFC). The CCEMG coefficients is slightly smaller in that case, i.e., 0.23 compared to 0.90 for the FE. The full set of checks pre-GFC is available upon request.

[20] For estimations and in-/out- of sample predictions, we also make use and follow Ditzen (2018) and Ditzen (2021).

**Table 2: Coefficient Estimates (Learning-by-Importing-Augmented Gravity Equation)**

| | CCEMG | FE | MG | CCEP | HFE | CCEMGR |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| VARIABLES | $\text{FLOW}_{ij,t}$ | $\text{FLOW}_{ij,t}$ | $\text{FLOW}_{ij,t}$ | $\text{FLOW}_{ij,t}$ | $\text{FLOW}_{ij,t}$ | $\text{FLOW}_{ij,t}$ |
| $\text{FLOW}_{ij,t-1}$ | 0.347*** | 0.907*** | 0.540*** | 0.374*** | 0.488*** | 0.460*** |
| | (0.00825) | (0.00451) | (0.00714) | (0.0161) | (0.00711) | (0.00730) |
| $\text{TB}_{j,t}$ | 0.975*** | 0.219*** | 0.793*** | 0.612*** | 0.865*** | 0.770*** |
| | (0.126) | (0.0279) | (0.0714) | (0.0801) | (0.103) | (0.0989) |
| $\text{GDP}_{i,t-1}$ | -0.312*** | -0.00174 | -0.189*** | -0.296*** | -0.232*** | -0.210*** |
| | (0.0778) | (0.00749) | (0.0304) | (0.0338) | (0.0283) | (0.0422) |
| $\text{GDP}_{j,t-1}$ | -0.117 | -0.0239*** | -0.188*** | -0.195*** | -0.134*** | -0.0634 |
| | (0.0954) | (0.00714) | (0.0253) | 0.0271 | (0.0306) | (0.0454) |
| $\text{GDP}_{t-1}$ | 0.228 | -0.0419 | 0.564*** | | 0.456*** | 0.594*** |
| | (0.201) | (0.0258) | (0.0243) | | (0.0996) | (0.160) |
| Time Fixed Effects | N | Y | N | N | N | N |
| Country/Country-Pair Fixed Effects | Y | Y | N | N | Y | Y |
| Unobservable Common Factors | Y | N | N | Y | N | Y |
| Constant | -0.628 | 2.035*** | | | | -4.471** |
| | (3.552) | (0.353) | | | | (2.033) |
| Observations | 70,579 | 70,604 | 70,591 | 70,526 | 70,579 | 70,579 |
| Number of pairs | 1,473 | 1,480 | 1,475 | 1,468 | 1,473 | 1,473 |
| Adj. R-squared | 0.79 | 0.90 | 0.75 | 0.84 | 0.74 | 0.77 |

*Note*: Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

and increase in the multilateral trade deficit in the destination country $j \in n \setminus i$ should *ceteris paribus* attract more trade flows from each source country $i = 1, 2, ..., N$. Our empirical estimates correspond precisely to the theoretical predictions of the model. Specifically, the trade imbalance coefficient estimate is equal to 0.98 using the CCEMG approach (and statistically not significantly different from one, just as predicted by the theory), which is our preferred baseline specification (see column (1) of Table 2). Moreover, we obtain a positive and statistically significant pooled or average trade imbalance coefficients using all seven estimation techniques presented in Table 2 (and Table 4 in Appendix for other popular estimation methods). Notice that the lowest trade imbalance coefficient estimate of 0.22 is obtained using the FE approach (see column (2) of Table 2). The remaining estimation techniques generate the trade imbalance coefficient estimates that are generally closer to the CCEMG approach compared to the FE approach.

The remaining lines of Table 2 present the (pooled or average) coefficient estimates associated with the lagged source country aggregate income ($\mathrm{GDP}_{i,t-1}$), lagged destination country aggregate income ($\mathrm{GDP}_{j,t-1}$), and lagged world aggregate income ($\mathrm{GDP}_{t-1}$). According to the learning-by-importing-augmented gravity equation depicted in equation (2.18), the size-adjusted bilateral trade flows ($\mathrm{FLOW}_{ij,t}$) are *ceteris paribus* increasing in $\mathrm{GDP}_{j,t-1}$ and $\mathrm{GDP}_{t-1}$, but decreasing in $\mathrm{GDP}_{i,t-1}$. Consistent with the theory, we do find evidence of a small, negative, and statistically significant coefficient estimate for $\mathrm{GDP}_{i,t-1}$ across the board. We also find that the coefficient estimate for $\mathrm{GDP}_{t-1}$ is positive and statistically significant, but only for some of the estimation techniques that retain parameter heterogeneity. However, contrary to the theoretical predictions, the coefficient estimates for $\mathrm{GDP}_{j,t-1}$ are generally negative, but statistically significant only in the absence of the unobservable common factors and the CCEP approach. As a consequence, our preferred benchmark model specification CCEMG delivers coefficient estimates that are the most theoretically consistent. And when a subset of the unobserved common factors are removed (see CCEMGR in column (6) of Table 2), the coefficient estimate for our single observable common factor $\mathrm{GDP}_{t-1}$ becomes relatively large, positive, and statistically significant. Therefore, only controlling for a full set of sources, which drive common shocks, make observable world output insignificant. In other words, it follows that both the observable and the unobservable common factors generally play an important role when drawing inference.

Finally, we conduct a battery of robustness checks in Appendix E.5. We explore the econometric issues due to the fact that both the bilateral trade flows (i.e., the dependent variable) and the multilateral trade balance (i.e., a regressor) are determined contemporaneously and simultaneously. We also explore the prediction that the trade imbalance coefficient is homogeneous and equals to unity for each country pair. By excluding the multilateral trade imbalance from the vector of regressors as well as re-specifying the size-adjusted trade flows in terms of consumption expenditure in the destination country, we find that the CCEMG approach remains more than twice lower than the FE approach. We therefore conclude that in the context of the learning-by-importing-augmented gravity equation, our results are robust to the aforementioned deviations from the baseline model.

### 3.4 Empirical Implications

Our opening results presented in Section 3.3 establish an important empirical stylized fact. Specifically, in general, traditional panel regression models, proposed by Feenstra (2016), Piermartini and Yotov (2016), Anderson and Yotov (2020) and others that are based solely on country-specific fixed effects, time fixed effects, and a pooled coefficient estimator, provide misleading inference when extended from a static to a dynamic gravity equation setting. In particular, traditional panel regression models generate an exceedingly upwardly-biased estimate of the trade persistence coefficient. There are two distinct sources of this upward bias. First, in keeping with Chudik and Pesaran (2015), omitted unobservable common factors cause the trade persistence coefficient estimates to be biased and inconsistent, because they ignore strong cross-sectional dependence of bilateral trade flows across different country pairs, which stems from the country-specific and dynamic structure of the multilateral trade resistance in equation (2.18). Second, in accordance with Pesaran and Smith (1995), the coefficient estimates of a dynamic panel regression equation are biased and inconsistent if the coefficient estimator neglects parameter heterogeneity. Based on the premise that the estimate of the trade persistence coefficient in our baseline model specification incorporates the unobservable common shocks and retains parameter heterogeneity, this section of the paper compares and contrasts the magnitude and the robustness of our coefficient estimates to those in the existing literature.

The most well-known existing theory of bilateral trade flow persistence due to Yotov and Olivero (2012) and Anderson et al. (2020) is based on the standard neo-classical capital accumulation equation. As per usual, the neo-classical theory introduces an infinitely-divisible measure of aggregate capital stock, which depreciates at a deterministic rate $\delta \in [0, 1]$ per every time period and requires investment into new capital stock in order to preserve the balanced growth path. The dynamics of the aggregate capital stock are then linked to the bilateral trade flows through a Cobb-Douglas production function and standard homothetic preferences across the domestic and foreign varieties from which a dynamic gravity equation is derived. The main advantage of the neo-classical theory of trade persistence is that it hinges on capital accumulation and exploits one of the most fundamental sources of dynamics in the real business cycle literature. However, the main disadvantage of the neo-classical theory is that it predicts a highly restrictive domain for the trade persistence coefficient that is at odds with the empirical evidence.

In particular, Yotov and Olivero (2012) show that the neo-classical theory predicts a trade persistence coefficient equivalent to $1 - \delta$ and estimate $\delta$, measuring the annual rate of capital depreciation, to be anywhere from 0.06 to 0.14. Conversely, IMF (2015) estimates that the value of $\delta$ lies in the interval of 0.04 and 0.1, depending on the time period and whether the country is advanced or developing. If we take the neo-classical theory at face value, it follows that an empirically plausible lower bound for the annual trade persistence coefficient is around 0.86. But the lower bound of 0.86 merely corresponds to some of our exceedingly upwardly-biased and inconsistent estimates that neglect parameter heterogeneity and exclude unobservable common factors (see column (2) in Table 2). Once we incorporate the pair-specific fixed effects and flexible time effects and refrain from the pooled coefficient estimator, the magnitude of the trade persistence coefficient shrinks by around 2-3 times. Specifically, our baseline model specification predicts a trade persistence coefficient of 0.35 (see column (1) in Table 2), which in the light of the neo-classical theory implies that 65% of the world capital stock depreciates every single year (i.e., up to 16 times more than the IMF (2015) estimates).

Despite how simple and elegant the neo-classical framework of trade persistence is, the striking discrepancy between theory and evidence suggests a more pragmatic view that capital accumulation forms only a subset, but perhaps not the core, of the trade persistence mechanism. We develop a theory of trade persistence based on learning-by-importing. The main advantages of our is that it presents not only a much more flexible identification for the domain of the trade persistence coefficient, but also allows for a heterogeneous magnitude across different country pairs. Specifically, our theory predicts a trade persistence coefficient equal to $\chi_{ij}(\eta - 1)$. In this theoretical identity, parameter $|1 - \eta|$ stands for trade elasticity, where $\eta > 0$ is the elasticity of substitution, the value of which generally ranges between 5 and 10 in the related literature (see Anderson and van Wincoop (2004) for evidence and Arkolakis et al. (2012) for the application). Conversely, parameter $\chi_{ij} > 0$ measures the intensity of learning-by-importing specific to any given country pair. If we take the values of $\eta$ from the literature and combine them with our CCEMG estimate for the trade persistence coefficient of 0.35, then our theoretical model predicts a lower (upper) bound for the learning-by-importing parameter to be 0.035 (0.07). This value is even lower than 0.1, which was originally assumed in the seminal contribution of Ravn et al. (2006).

However, contrary to the traditional predictions in the gravity literature, there exists some evidence that the trade elasticity $|1 - \eta|$ is in fact time-varying as opposed to constant over time as is traditionally considered to be the case (e.g., Anderson and van Wincoop (2003)). Specifically, Boehm et al. (2020) measure the short-run and the long-run trade elasticities by exploiting recurring exogenous tariff changes for identification purposes. The authors find substantially smaller values of trade elasticities equal to around 0.7 in the short-run and 1.75 in the long run in absolute value terms. Looking at this new evidence from the perspective of the learning-by-importing-augmented gravity equation indicates that trade persistence can be quite large in the short-run (i.e., $0.35/0.7 = 0.5$), but declines by around 2.5 times in the long-run (i.e., $0.35/1.75 = 0.2$). And since we analyze more than 60 years worth of data across advanced and developing economies with starkly different industrial structures, the time-variation and heterogeneity of the trade elasticities across countries is expected (see Imbs and Mejean (2017) for the cross-country evidence). Notice that the implied long-run persistence parameter of 0.2 is remarkably compatible with the relatively sharp and synchronized international trade flow adjustments in response to common shocks observed in the data (see more details in Appendix D).

# 4 Cross-Validation

We have thus far established that controlling for the unobservable common factors and retaining parameter heterogeneity specific to each country pair when estimating the learning-by-importing-augmented gravity equation leads to a significantly lower trade persistence coefficient than predicted using the conventional estimation techniques applied in the existing literature. The benefits of adopting our empirical approach are two-fold. First, our preferred empirical strategy is consistent with the theory of learning-by-importing, which predicts heterogeneous trade persistence coefficients across different country pairs and generates inward and outward multilateral resistance with lags that strongly correlate with foreign demand and foreign supply shocks. Second, unlike the static gravity equation due to Anderson (1979), Anderson and van Wincoop (2003), and Feenstra (2016), which predicts zero trade persistence, or the neo-classical gravity equation due to Yotov and Olivero (2012), Alvarez (2017), and Anderson et al. (2020), which predicts a trade persistence coefficient of around 0.8-0.9, our preferred estimation strategy delivers a cross-country average trade persistence coefficient equal to around 0.35, which is able to rationalize the sharp and synchronized international trade flow adjustments in response to common trade shocks.

In order to illustrate that our preferred estimation strategy, titled CCEMG, outperforms the leading rival empirical strategies, we will conduct a battery of cross-validation tests. First, we conduct a so-called "horse race" for the predictive performance of different empirical estimation strategies of our empirical model presented in equations (3.1)-(3.3). Second, we describe and visualize the extent of heterogeneity in key parameters of the model.

## 4.1 Prediction Performance "Horse Race"

### 4.1.1 In-Sample Performance

We start by Table 3 compares the Root Mean Square Errors (RMSE) calculated using the CCEMG, MG, CCEP, and FE methodologies (see Sections E and 3.2 for more details). The in-sample RMSEs are presented for the full data sample, the observed "normal times", and the observed "bad times", in order to compare different model performance inside and outside of time periods characterized by common shocks. Consistent with Kose et al. (2020), the "bad times" represent the global recession years, namely 1975, 1982, 1991, and 2009, while the "normal times" are all of the remaining years in our data sample that spans 1950-2014. The term $w = \{0, 1, 2, 3\}$ further indicates the length of the windows surrounding the recession years (i.e., number of years before and after global shocks that are included in the "bad times" sample in addition to the outlined recession years).

According to our RMSE calculations presented in Table 3, the CCEMG approach delivers the most accurate data fit not only throughout the entire data sample, but also during solely "normal times" or "bad times". Recall that the CCEMG approach controls for the unobservable dynamic common factors (dynamic multilateral resistance terms) and retains the parameter heterogeneity specific to each country pair. The runner-up methodologies are the MG approach, which retains the country-pair-specific parameter heterogeneity, but expends the unobservable common factors, and the CCEP approach, which controls for the unobservable common factors, but ignores the country-pair-specific parameter heterogeneity. The conventional FE approach, which expends the unobservable common factors and ignores the country-pair-specific parameter heterogeneity delivers the largest RMSE value and predicts the least accurate data fit. The reason why the performance of the FE approach is inferior to the MG, CCEP, and CCEMG approaches is because the latter all deliver a lower trade persistence coefficient than the FE approach (see Table 2). Allowing for the pair-specific fixed effects and source and destination time effects does not change conclusions as evidenced in Appendix E.6 Table 7. That said, all of the methodologies we consider perform marginally better during the "normal times" rather than the "bad times", among other reasons due to the fact that "bad times" occur considerably less frequently.

For the sake of robustness, we calculate the RMSEs for numerous other methodologies considered in this paper and generally reach the same outcome (see Table 7). While Tables 3 and 7 calculate

**Table 3: Root Mean Square Error**

| Method | Full Sample | "Bad Times" | | | | "Normal Times" | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $w = 0$ | $w = 1$ | $w = 2$ | $w = 3$ | $w = 0$ | $w = 1$ | $w = 2$ | $w = 3$ |
| CCEMG | **0.38** | **0.41** | **0.41** | **0.40** | **0.41** | **0.38** | **0.37** | **0.37** | **0.35** |
| MG | 0.45 | 0.52 | 0.51 | 0.49 | 0.45 | 0.44 | 0.43 | 0.41 | 0.40 |
| CCEP | 0.47 | 0.47 | 0.47 | 0.46 | 0.47 | 0.43 | 0.42 | 0.42 | 0.39 |
| FE | 0.55 | 0.65 | 0.63 | 0.60 | 0.59 | 0.54 | 0.52 | 0.51 | 0.49 |

*Note*: This figure presents the Root Mean Square Errors (RMSE) calculated using different methods of estimating the learning-by-importing-augmented gravity equation. The in-sample RMSEs are presented for the full data sample, the observed "normal times", and the observed "bad times" in order to compare different model performance inside and outside of time periods characterized by common trade shocks. Consistent with Kose et al. (2020), the "bad times" represent the global recession years, namely 1975, 1982, 1991, and 2009, while the "normal times" are all of the remaining years in our data sample that spans 1950-2014. The term $w = \{0, 1, 2, 3\}$ further indicates the length of the windows surrounding the recession years (i.e., number of years before and after global crises). The values in bold indicate the smallest RMSE.

RMSEs based on the "normal times" and "bad times" sub-samples, they nonetheless rely on the dynamic gravity equation coefficient estimates from the entire data sample. In order to ensure that our findings are robust, we also present Table 8 in Appendix E.6, which calculates both the RMSEs as well as the coefficient estimates based solely on the "normal times" and "bad times" sub-samples. Due to the limited number of time periods in the "bad times" sub-sample, not all methodologies can be successfully implemented, since the "mean group" techniques that retain parameter heterogeneity rely on a sufficiently large temporal dimension of the panel. However, the outcome regarding the superiority of the CCEMG approach generally holds (the only viable rival during "normal times" is PPML4 though CCEMG strictly dominates during global recessions).[21] Consequently, we conclude that incorporating the unobservable common factors and retaining parameter heterogeneity in dynamic gravity models generally outperforms the standard empirical approaches in the existing literature that tend to ignore both.

### 4.1.2 Out-of-Sample Performance

A critique of the prediction performance so far relates to the parametrization of different estimators. We thus repeat the above exercise for the training sample, using half of the sample over time (i.e., until 1982) and test how various estimators perform in the out-of-training-sample. The calculation of the out-of-sample predictions for MG, CCEP, and CCEMG is non-trivial, there are few assumptions and possible options that need to be considered.[22] The exercise is much simpler for homogeneous estimators without unobserved common factors. We find that the out-of-sample RMSEs for the FE estimates are slightly larger compared to the in-sample counterparts, both

[21] PPML-types of estimators are, however, not ideal in this context, see Appendix E

[22] In the case of CCEMG and CCEP, we create the cross-sectional averages first and then use them as simple observed regressors. We hence know both factors and their numbers (see Stauskas and Westerlund (2022)), as it is throughout the paper. This out-of-sample exercise also assumes that there are no breaks in the factor loadings and the number of factors remain the same. Moreover, standing how our coefficients are estimated, i.e., pair-wise first and country-wise second to come up with a single value, this requires great attention on how the predictions are made in the first step and incorporated to maintain the full pairs heterogeneity. There are two ways to do that in our case: a) using OLS at the pair level and predict the out-of-sample values, or b) using the coefficients of the pair-wise step (and country-wise as a check) multiplied by their coefficients to have predicted values, also out-of-sample. We plan to check these options carefully in the out-of-sample performance assessment.

overall (0.57) and in recession/normal times.[23]

### 4.2 Parameter Heterogeneity

Our results have thus far established the importance of retaining parameter heterogeneity across all country pairs, since it is one of the drivers of the trade persistence puzzle and a source of bias and inconsistency of parameters in dynamic gravity models (see Pesaran and Smith (1995) and our discussion in Sections 3.3 and 3.4). We now present Figure 1, which demonstrates the extent of the cross-country parameter heterogeneity in our sample as well as the associated country-specific uncertainty surrounding the coefficient estimates. Our focus is limited to the trade persistence and the trade imbalance coefficients calculated using the CCEMG approach described in Section 3.2 and further elaborated in Appendix E, which retains the country-pair-specific parameter heterogeneity and controls for the unobservable time-varying terms (common factors). For the sake of clarity and space, the coefficient estimates specific to each country pair are averaged across all destinations $j$ for each source country $i$, resulting in $N$ number of coefficient estimates that we report out of the total of $N(N-1)$ number of unique country pairs for which coefficient estimates exist. In general, we establish pervasive country-specific parameter heterogeneity clustered around the average coefficient estimates presented in Table 2 with few and far between outliers.

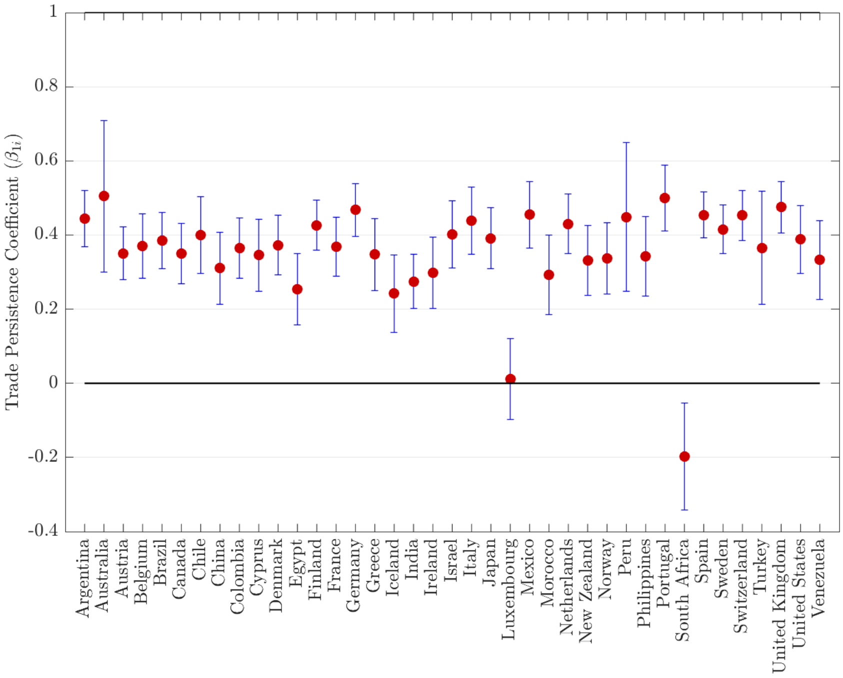
Figure 1 summarizes how hetereogeneous parameters are for each country in our sample. Unlike confidence interval, capturing the population parameter with pre-determined probability, the width of the shown interval visualizes how different coefficients are for different trade partners. Notice that nearly all of the trade persistence coefficients turn out to be positive and their value is scattered around the interval of -0.2 and 0.5 (see Figure 1a) compared to the cross-sectional average of 0.35 (see column (1) in Table 2). The country-specific estimates of the trade persistence coefficients contain two notable outliers, namely South Africa, where it is small, negative, and statistically significant, and Luxembourg, where it is not significantly different from zero.[24] The trade persistence coefficients in all other countries are significantly different from zero and unity. Conversely, the uncertainty of the estimated trade imbalance coefficients is considerably larger (see Figure 1b); namely, it ranges from around -2 in Peru to nearly 2.5 in Greece. Though some countries exhibit statistically insignificant trade imbalance coefficients, the majority of the trade imbalance coefficients are statistically significant (i.e., 26 out of 39) and clustered around the cross-sectional average unitary elasticity, in line with the theory of learning-by-importing.

Due to a relatively large number of coefficient estimates (i.e., $N = 39$), and the fact that the trade persistence coefficient estimate outliers are relatively small, the inference drawn from the cross-sectional average of the trade persistence coefficients is arguably not susceptible to the presence of those outliers. While there exist larger outliers of the trade imbalance coefficients, they are both positive outliers (e.g., Greece and Venezuela) as well as negative outliers (e.g., Cyprus, Peru, and South Africa). As a consequence, the inference drawn from the cross-sectional averages of the trade imbalance coefficients is largely unbiased by the presence of outliers. We also document a

[23] Or PPML4 performs marginally better out-of-sample than in-sample but only in normal times and overall (0.52).

[24] One possible explanation for the negative trade persistence coefficient we obtain in South Africa for the period of 1960-2014 is the abolishment of the authoritarian apartheid regime in the early 1990s. During that institutional transformation, South Africa opened up to trade with many new trade partners and at the same time shifted away from trade with the old trade partners, thereby causing bilateral trade persistence to break down.

(a) **Country-Specific Trade Persistence Coefficients**




(b) **Country-Specific Trade Imbalance Coefficients**

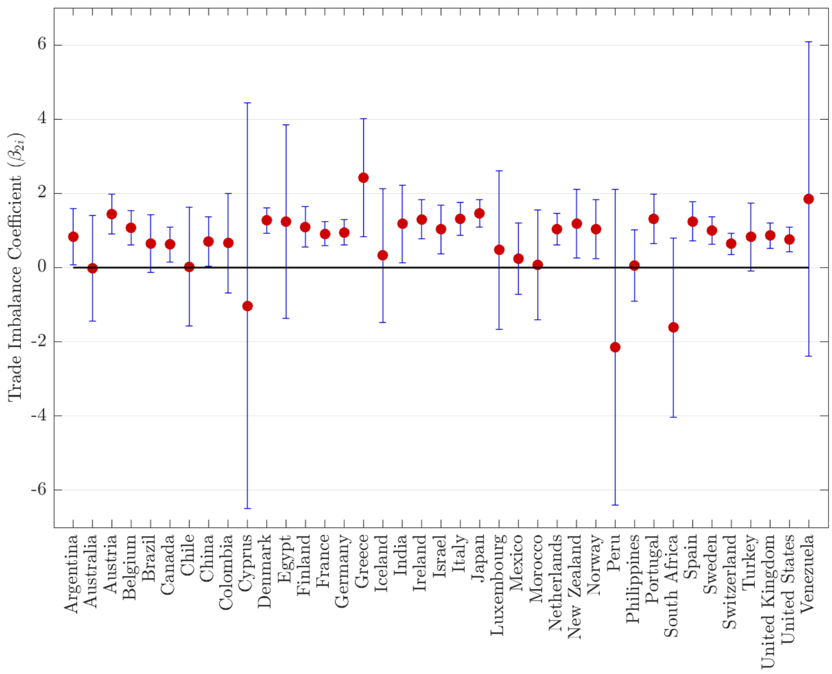



**Figure 1: Cross-Country Heterogeneity of Trade Persistence and Trade Imbalance Coefficient Estimates (CCEMG)**

*Notes:* Subplots (a) and (b) display the cross-country heterogeneity and the uncertainty surrounding the CCEMG coefficient estimates $\beta_{1i}$ and $\beta_{2i}$, respectively, where $i$ stands for the "source" country $i = 1, 2, ..., N$ (i.e., the market from which exports originate). The magnitude of the red dots is measured by the vertical distance and denotes the CCEMG coefficient estimates specific to each source country $i$. The names of the source countries are displayed on the horizontal axis. The country-specific CCEMG coefficient estimates are calculated as an average across all $N$ destinations indexed by $j$ from which the source country $i$ imports. The blue bars surrounding the CCEMG coefficient estimates are the 95% confidence intervals.

largely symmetric and fat-tailed distribution of the country-pair-specific trade imbalance coefficient estimates in Figure 6 in Appendix E.7.

## 5 Application: Counterfactual Trade Flows

After having shown that the theory of learning-by-importing generates a more accurate description of the data, predicts trade adjustment dynamics with greater precision, and is consistent with a theoretical link between the trade persistence parameter and the country's integration into global value chains, we turn to a stylized application of our dynamic gravity model. Due to high policy relevance of the recent trade war between the USA and China, we analyze this particular country pair to illustrate the importance of trade persistence heterogeneity in predicting the counterfactual trade adjustment dynamics in response to a multilateral trade imbalance shock.[25] Notice that in a standard panel data gravity model such an analysis is not feasible as neither heterogeneity nor multilateral imbalances are taken into account. And we argue that this exercise is important, because large and rising bilateral trade imbalances spurred much attention towards the policymakers and arguably incited the recent USA-China tariff war. Empirical evidence also suggests that the multilateral trade imbalances tend to affect bilateral trade balances, but not the other way around (see IMF (2019)). Our model predicts an intrinsic relationship between the multilateral and the bilateral trade imbalances, which we exploit in what follows.

We start by constructing the impulse response function for the size-adjusted trade flows due to the aggregate trade imbalance term. It is given by $\frac{\partial \mathbb{E}_t \ln A_{ij,t+h}}{\partial \ln \Xi_{j,t}} = (\chi_{ij}(\eta - 1))^h$, because learning-by-importing predicts a unitary trade imbalance coefficient (see equation (2.18) and set $S = 1$). Also recall that our baseline model specification predicts a unitary trade imbalance coefficient (see column (1) in Table 2). The impulse response dynamics are therefore fully inherited from the trade persistence coefficient. We therefore illustrate counterfactual trade changes due to multilateral trade imbalances, resorting to estimates of $\chi_{ij}(\eta - 1)$. Under homogeneous parameters with homogeneous fixed effects, the persistence parameter is around 0.9, whereas our preferred CCEMG estimator yields 0.35. This results in a different transition path and size of the cumulative effect of the shock. For instance, if the USA reduces its trade balance deficit by 1 percent (yielding a rise in $\Xi_{j,t}$), then given unitary elasticity in the equation (2.18) and empirical evidence in Table 2, such a change results in a 1 percent increase in size-adjusted bilateral trade flows with the USA. Over the long term, since trade is persistent, the full effect reaches around 1.5 for our preferred method, and the overall adjustment takes just a few years. In a model with zero trade persistence, instead, the short- and long-term effects coincide, and the impact is instantaneous. On the contrary, very persistent processes such as those predicted by homogeneous fixed effects take decades of adjustment after a shock (e.g., the persistence of 0.9 implies a 10 percent cumulative long-term effect on bilateral

[25] There is a vast literature on welfare implications of the trade relationship between these two large economies as well as the rest of the world (e.g., Allen et al. (2020) identify expenditure changes in a static gravity environment in the face of a trade war, Fajgelbaum et al. (2021) explore global trade adjustments due to tariff increases, emphasizing the importance of heterogeneity in response elasticities, and Adao et al. (2017) develop a framework to allow for elasticity heterogeneity and flexible functional forms to predict counterfactual trade costs if China remained weakly integrated into the global economy). Our framework is aggregate, but extends along the temporal dimension as well as emphasizes the importance of parameter heterogeneity even at this aggregation level. We leave welfare implications for future research but rather explore counterfactual trade flows.

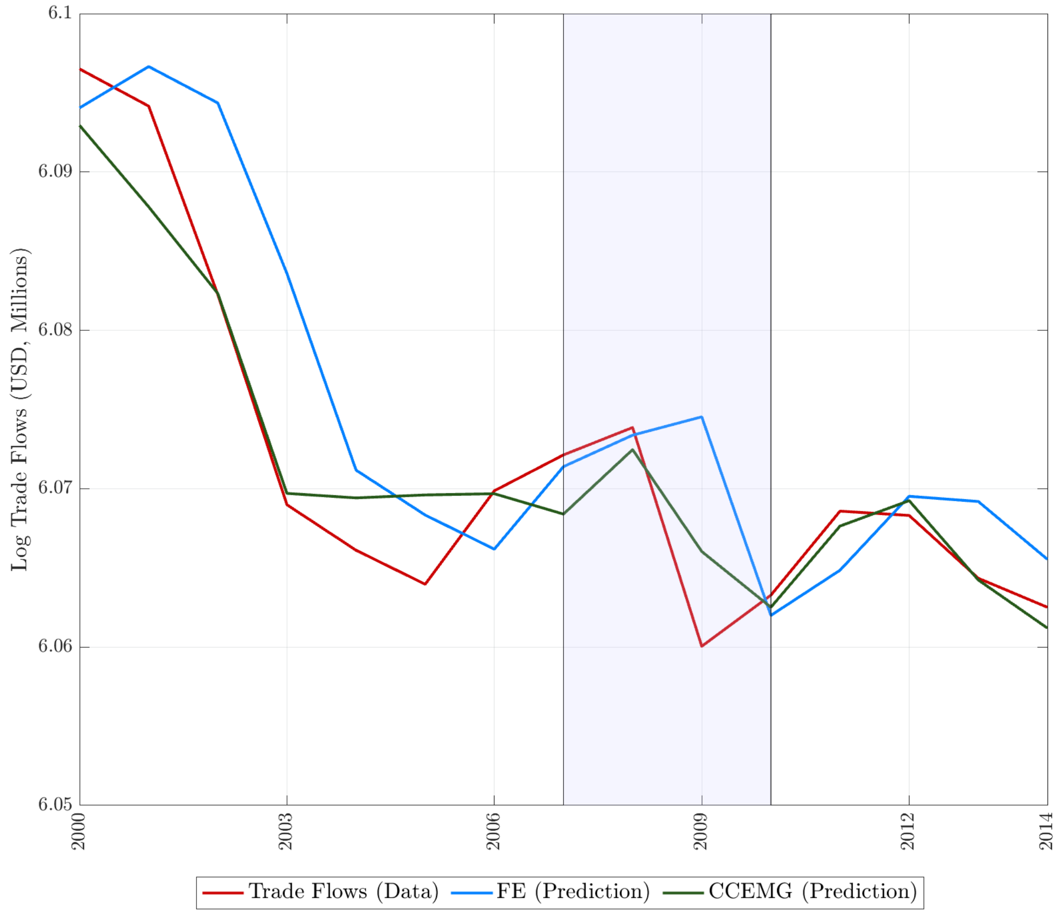



**Figure 2: Predicted Size-Adjusted Trade Flows Between USA and China (Log Scale)**

*Note:* The figure depicts observed size-adjusted and logged bilateral flows, CCEMG in-sample predictions and Fixed Effects (FE) in-sample predictions, as covered in Table 2 and explained in Section 3.2.

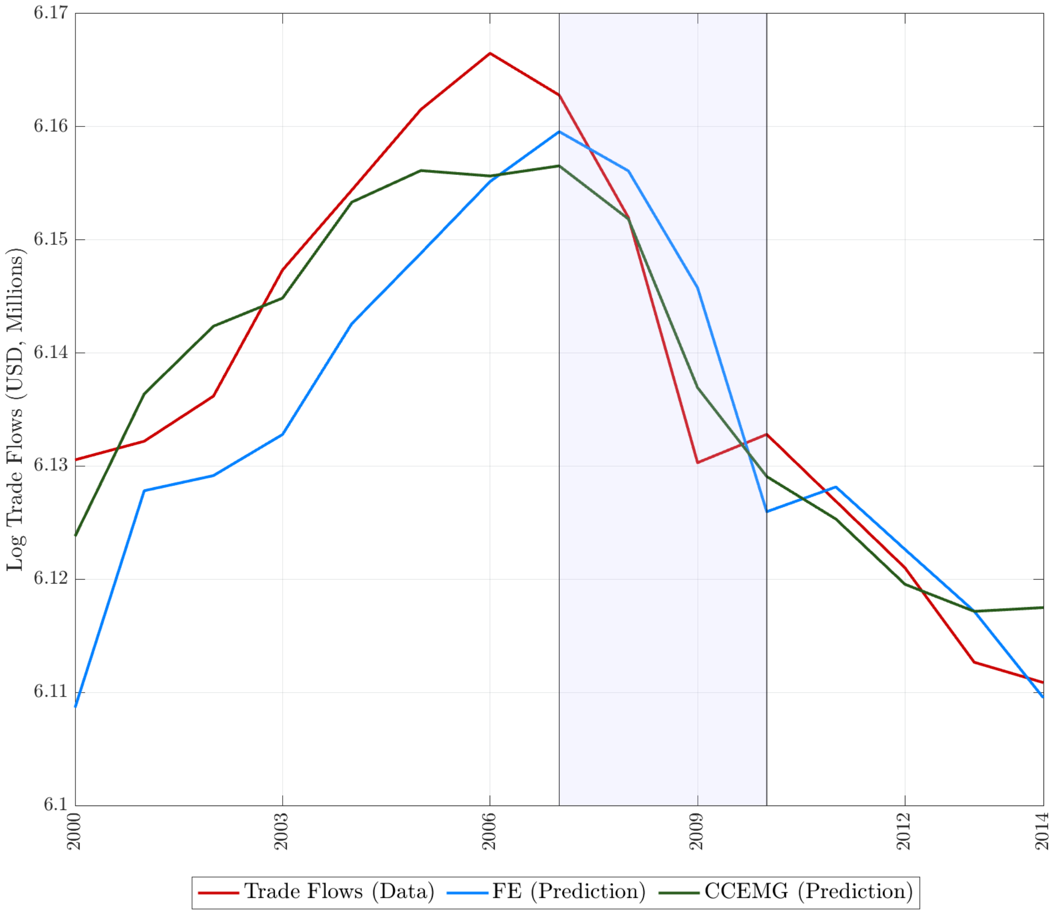



**Figure 3: Predicted Size-Adjusted Trade Flows Between China and USA (Log Scale)**

*Note:* The figure depicts observed size-adjusted and logged bilateral flows, CCEMG in-sample predictions and Fixed Effects (FE) in-sample predictions, as covered in Table 2 and explained in Section 3.2.

exports caused by a 1 percent shock in the aggregate trade balance).

Though our econometric framework relies on relatively large cross-sectional and temporal dimensions, we also dig into country-specific trade flows. Looking at the USA-China and China-USA trade flows, we obtain trade persistence coefficient estimates of 0.44 and 0.76, respectively. Clearly, the USA faces a more flexible and faster adjustment process, with less dependence on Chinese inputs in its production. In contrast, it is more difficult to rewire the Chinese trade flows from the USA economy to other countries. The long-run elasticity for the Chinese trade flows to the USA is more than 4, a substantial change due to the USA trade balance shock.[26]

The difference in adjustment exemplifies the importance of heterogeneous responses for counterfactual predictions. Figures 2-3 visualize how well the predicted and the observed bilateral trade flows for the USA-China and China-USA resemble one another. In particular, the FE approach with homogeneous fixed effects induce longer and more persistent fluctuations than observed in the data, thereby not only affecting the accuracy of the prediction, but also overestimating the counterfactual values of the short and long-run impacts of shocks and the adjustment dynamics (see the shaded area that highlights the GTC years and the FE approach displays delays in reaction for both country pairs). By contrast, our preferred model specification CCEMG, which retains parameter heterogeneity and incorporates the unobservable common factors (multilateral resistance terms) generates a more precise prediction of bilateral trade flows. Having a more accurate estimate of trade persistence and knowing its drivers help to predict long-term effects that supply shocks could have on the main macroeconomic aggregates such as potential output growth (see Le Roux (2022)). It also impacts dynamic welfare calculations which are left for future research.

## 6 Concluding Remarks

International trade flows are volatile and imbalanced. Yet, not much is known about the mechanism through which trade flows adjust in response to shocks over time. As things stand, the bulk of the modern trade literature relies on the ubiquitous gravity equation to predict the value of trade flows across countries. And it is notoriously successful at predicting both "who trades with whom" as well as "how much is traded" when trade shocks are local or country-specific. But when trade shocks are common, the observed value of trade flows adjusts by more and more rapidly than predicted by the standard gravity equations presented in the literature. While the static gravity equation remains the workhorse framework for trade policy analysis in the context of permanent, one-off, and exogenous trade shocks, it is silent about the transitional dynamics. By contrast, the neo-classical gravity equation that relies on the theory of capital accumulation predicts excessively persistent and homogeneous international trade flows that are difficult to square with the sharp trade adjustments in response to common shocks and heterogeneous ensuing recoveries.

We put forward a dynamic gravity model, which outperforms existing alternatives empirically, helping policymakers make more precise predictions and build more sound counterfactuals. Our tractable dynamic gravity equation is derived from a theory of learning-by-importing. It takes

[26] Despite the high cumulative effect, the implied value of the trade persistence coefficient, even for the bilateral trade with the largest economy, the USA, is substantially lower than inferred from the standard method (its value is 0.9, implying a long-run elasticity of around 10). What is more, such a counterfactual exercise is usually infeasible in alternative models that do not give rise to aggregate imbalances in bilateral trade flows.

time to internalise the productivity gains, such that exporter efficiency depends on lagged bilateral trade flows and the capacity to internalise those gains depends on a parameter that reflects, among other things, bilateral infrastructure. Our framework offers several advantages. First, learning-by-importing predict autocorrelated (i.e., persistent) trade flows and trade persistence that is heterogeneous across country pairs. Second, cross-country learning-by-importing asymmetry creates differences in home-bias, which explains additional variation in bilateral trade flows over time and across space beyond the standard gravity measures, such as aggregate income and geographic distance. Third, learning-by-importing creates differences in the "inward" and "outward" propensities to trade (i.e., multilateral trade resistance terms) whose contemporaneous as well as pre-determined values enter the dynamic gravity equation. Common shocks are thus heterogeneously disruptive, because multilateral trade resistance terms are strongly correlated with country-specific trade imbalances as well as foreign demand and supply. Despite these new channels, our model conveniently nests the exisitng models of the gravity equation.

We estimate the dynamic gravity equation for 39 countries over the period of 1950-2014 using several dynamic panel regression techniques. We show that in addition to the standard variables in the gravity equation, multilateral trade imbalance is an important determinant of bilateral trade flows both theoretically and empirically. We establish two causes of the trade persistence puzzle. First, the prevalent methods of estimating (dynamic) gravity equations do not appropriately account for the unobserved dynamic multilateral factors. For instance, the standard "country" fixed effects are time-invariant, while the "country-time" fixed effects are homogeneous for all country pairs. Our preferred interactive fixed effects estimator provides a flexible way to account for unobserved dynamic terms impacting each trade pair heterogeneously. Second, the inference is commonly drawn from pooled gravity equation coefficient estimates, which ignores the fact that trade flows between some country pairs are significantly less persistent than others, emphasizing structural differences across trading partners and echoing the emerging literature on heterogeneous trade elasticities (see, e.g., Adao et al. (2017); Boehm et al. (2020); Carrere et al. (2020); Fajgelbaum et al. (2021)).

Contrary to the antecedents, we exploit a relatively large temporal dimension of our panel in order to retain the cross-country parameter heterogeneity. We also account for unobservable variation in multilateral dynamic factors, which we model empirically as the cross-sectionally averaged country-specific regressors. We show that absent of the unobservable dynamic factors, the value of the pooled trade persistence coefficient is 0.91, which is comparable to the estimates in the existing literature. But this estimate is upwardly-biased as it contracts markedly in our benchmark model specification that retains the cross-country parameter heterogeneity and introduces the unobservable common factors (i.e., the cross-sectionally averaged coefficient is 0.35). If we expend (retain) the unobservable common factors, but retain (expend) parameter heterogeneity, the trade persistence coefficient nonetheless shrinks to 0.54 (0.37). This provides strong evidence in favor of a modern trade theory that predicts heterogeneous trade persistence across country pairs, such as our proposed theory of learning-by-importing.

We document pervasive heterogeneity of the trade persistence coefficients across countries. Our theory is also applied in generating a dynamic response to the aggregate trade imbalance on the bilateral trade flow between the US and China. And despite some success, the question of what drives the cross-country differences in the empirical estimates of the trade persistence coefficients remains an open discussion. The theory of learning-by-importing makes valuable progress in terms of resolving the trade persistence puzzle and offers a simple alternative framework to the neo-classical gravity equation. But in the end, we call for a more structural approach to tackling the global trade network dynamics and heterogeneity in trade elasticities. In particular, we encourage more research aimed at separating the short- and the long-run run effects in trade elasticities, which may portray substantial structural heterogeneity as is recently illustrated by Boehm et al. (2020). Another area that we forfeit to future research is dynamic non-linear panel regression models, which would appropriately account for the "zero trade problem" but simultaneously retain parameter heterogeneity and enrich the model specification with unobservable dynamic multilateral resistance terms with heterogeneous impacts on trade flows.

## References

Acharya, R. C. and W. Keller (2009, November). Technology transfer through imports. Canadian Journal of Economics 42(4), 1411–1448.

Adao, R., A. Costinot, and D. Donaldson (2017, March). Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade. American Economic Review 107(3), 633–689.

Alessandria, G., H. Choi, and K. J. Ruhl (2021). Trade adjustment dynamics and the welfare gains from trade. Journal of International Economics 131, 103458.

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

Alvarez, F. (2017). Capital Accumulation and International Trade. Journal of Monetary Economics 91(C), 1–18.

Amiti, M. and J. Konings (2007, December). Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia. American Economic Review 97(5), 1611–1638.

Anderson, J. E. (1979). A Theoretical Foundation for the Gravity Equation. American Economic Review 75(1), 106–116.

Anderson, J. E. (2011). The Gravity Model. Annual Review of Economics 3(1), 133–160.

Anderson, J. E., M. Larch, and Y. V. Yotov (2020). Transitional Growth and Trade with Frictions: A Structural Estimation Framework. The 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 E. van Wincoop (2004). Trade Costs. Journal of Economic Literature 42(3), 691–751.

Anderson, J. E. and Y. V. Yotov (2010). The Changing Incidence of Geography. American Economic Review 100(5), 2157–86.

Anderson, J. E. and Y. V. Yotov (2020). Short Run Gravity. Journal of International Economics 126, 103341.

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012). New Trade Models, Same Old Gains? The American Economic Review 102(1), 94–130.

Bai, J. (2009). Panel Data Models With Interactive Fixed Effects. Econometrica 77(4), 1229–1279.

Baldwin, R. and D. Taglioni (2006). Gravity for Dummies and Dummies for Gravity Equations. Working Paper 12516, National Bureau of Economic Research.

Boehm, C. E., A. A. Levchenko, and N. Pandalai-Nayar (2020). The Long and Short (Run) of Trade Elasticities. Working Paper 27064, National Bureau of Economic Research.

Carrere, C., M. Mrazova, and J. P. Neary (2020). Gravity Without Apology: the Science of Elasticities, Distance and Trade. The Economic Journal 130(628), 880–910.

Chaney, T. (2008, September). Distorted Gravity: The Intensive and Extensive Margins of International Trade. American Economic Review 98(4), 1707–1721.

Chudik, A. and M. H. Pesaran (2015). Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors. Journal of Econometrics 188(2), 393–420.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011, February). Weak and strong cross section dependence and estimation of large panels. Econometrics Journal 14, 45–90.

Cunat, A. and R. Zymek (2019). Bilateral Trade Imbalances. CESifo Working Paper Series 7823, CESifo.

Davis, D. R. and D. E. Weinstein (2002). The Mystery of the Excess Trade (Balances). American Economic Review 92(2), 170–174.

De Vos, I. and O. Stauskas (2021, November). Bootstrap Improved Inference for Factor-Augmented Regressions with CCE. Working Papers 2021:16, Lund University, Department of Economics.

Dekle, R., J. Eaton, and S. Kortum (2007). Unbalanced Trade. American Economic Review 97(2), 351–355.

Dekle, R., J. Eaton, and S. Kortum (2008). Global Rebalancing with Gravity: Measuring the Burden of Adjustment. IMF Staff Papers 55(3), 511–540.

Disdier, A. C. and K. Head (2008). The Puzzling Persistence of the Distance Effect on Bilateral Trade. Review of Economics and Statistics 90(1), 37–48.

Ditzen, J. (2018, September). Estimating dynamic common-correlated effects in Stata. Stata Journal 18(3), 585–617.

Ditzen, J. (2021, September). Estimating long-run effects and the exponent of cross-sectional dependence: An update to xtdcce2. Stata Journal 21(3), 687–707.

Dix-Carneiro, R., J. P. Pessoa, R. M. Reyes-Heroles, and S. Traiberman (2021, January). Globalization, Trade Imbalances and Labor Market Adjustment. NBER Working Papers 28315, National Bureau of Economic Research, Inc.

Eaton, J. and S. Kortum (2002, September). Technology, Geography, and Trade. Econometrica 70(5), 1741–1779.

Eberhardt, M. and F. Teal (2013). No Mangoes in the Tundra: Spatial Heterogeneity in Agricultural Productivity Analysis. Oxford Bulletin of Economics and Statistics 75(6), 914–939.

Eichengreen, B. and D. Irwin (1998). The Role of History in Bilateral Trade Flows. In The Regionalization of the World Economy, pp. 33–62. National Bureau of Economic Research, Inc.

Elliott, R. J. R., L. Jabbour, and L. Zhang (2016). Firm productivity and importing: Evidence from Chinese manufacturing firms. The Canadian Journal of Economics / Revue canadienne d'Economique 49(3), 1086–1124.

Ethier, W. (1982). National and International Returns to Scale in the Modern Theory of International Trade. American Economic Review 72(3), 389–405.

Fajgelbaum, P., P. K. Goldberg, P. J. Kennedy, A. Khandelwal, and D. Taglioni (2021, December). The US-China Trade War and Global Reallocations. Working Paper 29562, National Bureau of Economic Research.

Fally, T. (2015). Structural gravity and fixed effects. Journal of International Economics 97(1), 76–85.

Feenstra, R. (2016). Advanced International Trade: Theory and Evidence Second Edition (2 ed.). Princeton University Press.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The Next Generation of the Penn World Table. American Economic Review 105(10), 3150–3182.

Fieler, A. C. (2011). Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation. Econometrica 79(4), 1069–1101.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The Generalized Dynamic-Factor Model: Identification and Estimation. The Review of Economics and Statistics 82(4), 540–554.

Gopinath, G., E. Boz, C. Casas, F. J. Diez, P.-O. Gourinchas, and M. Plagborg-Moller (2020, March). Dominant Currency Paradigm. American Economic Review 110(3), 677–719.

Grossman, G. M. and E. Helpman (1995). Technology and trade. In G. M. Grossman and K. Rogof (Eds.), Handbook of International Economics, Volume 3 of Handbook of International Economics, Chapter 25, pp. 1279–1337. Elsevier.

Hacioglu-Hoke, S. and G. Kapetanios (2021). Common correlated effect cross-sectional dependence corrections for nonlinear conditional mean panel models. Journal of Applied Econometrics 36(1), 125–150.

Halpern, L., M. Koren, and A. Szeidl (2015, December). Imported Inputs and Productivity. American Economic Review 105(12), 3660–3703.

Head, K. and T. Mayer (2014). Chapter 3 - Gravity Equations: Workhorse, Toolkit, and Cookbook. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), Handbook of International Economics, Volume 4 of Handbook of International Economics, pp. 131 – 195. Elsevier.

Imbs, J. and I. Mejean (2017). Trade Elasticities. Review of International Economics 25(2), 383–402.

IMF (2015). Making Public Investment More Efficient. IMF Policy Papers, International Monetary Fund, Washington, D.C.

IMF (2019). World Economic Outlook, April 2019: Growth Slowdown, Precarious Recovery, Chapter Chapter 4: Drivers of Bilateral Trade and Spillovers from Tariffs, pp. 103–129. International Monetary Fund, Washington, D.C.

Juodis, A., H. Karabiyik, and J. Westerlund (2021). On the robustness of the pooled CCE estimator. Journal of Econometrics 220(2), 325–348. Annals Issue: Celebrating 40 Years of Panel Data Analysis: Past, Present and Future.

Kapetanios, G., C. Mastromarco, L. Serlenga, and Y. Shin (2017). Modelling in the Presence of Cross-Sectional Error Dependence, pp. 291–322. Cham: Springer International Publishing.

Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011). Panels with Non-Stationary Multifactor Error Structures. Journal of Econometrics 160(2), 326–348.

Keller, W. (2004, September). International Technology Diffusion. Journal of Economic Literature 42(3), 752–782.

Kose, M. A., N. Sugawara, and M. E. Terrones (2020). Global Recessions. CEPR Discussion Papers 14397, C.E.P.R. Discussion Papers.

Le Roux, J. (2022). How persistent supply chain disruptions could affect euro area potential output. In ECB Economic Bulletin, Volume 1/2022.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. Econometrica 71(6), 1695–1725.

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

Pesaran, M. H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. Econometrica 74(4), 967–1012.

Pesaran, M. H. and R. Smith (1995). Estimating Long-Run Relationships from Dynamic Heterogeneous Panels. Journal of Econometrics 68(1), 79–113.

Piermartini, R. and Y. Yotov (2016). Estimating Trade Policy Effects with Structural Gravity. School of Economics Working Paper Series 2016-10, LeBow College of Business, Drexel University.

Ravn, M., S. Schmitt-Grohé, and M. Uribe (2006). Deep Habits. The Review of Economic Studies 73(1), 195–218.

Santos-Silva, J. and S. Tenreyro (2007). The Log of Gravity. Review of Economics and Statistics 88, 641–658.

Schmidt-Grohé, S. and M. Uribe (2017). Open Economy Macroeconomics. Princeton, New Jersey: Princeton University Press.

Stauskas, O. and J. Westerlund (2022, October). Tests of Equal Forecasting Accuracy for Nested

Models with Estimated CCE Factors. Journal of Business & Economic Statistics 40(4), 1745–1758.

Weidner, M. and T. Zylkin (2019). Bias and Consistency in Three-Way Gravity Models. Papers 1909.01327, arXiv.org.

Westerlund, J. (2018). CCE in Panels with General Unknown Factors. The Econometrics Journal 21(3), 264–276.

Westerlund, J., Y. Petrova, and M. Norkute (2019). Cce in fixed-t panels. Journal of Applied Econometrics 34(5), 746–761.

Westerlund, J. and F. Wilhelmsson (2009). Estimating the Gravity Model Without Gravity Using Panel Data. Applied Economics 43(6), 641–649.

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

Zhang, H. (2017). Static and dynamic gains from costly importing of intermediate inputs: Evidence from Colombia. European Economic Review 91, 118–145.

Zwinkels, R. C. and S. Beugelsdijk (2010). Gravity Equations: Workhorse or Trojan Horse in Explaining Trade and FDI Patterns Across Time and Space? International Business Review 19(1), 102 – 115.

Online Appendix
(Not For Publication)

# A Theoretical Model

## A.1 Distributor

**Cost Minimisation.** Production technology of a distributor:

$$x_{ij,t} = \left[ \int_0^1 \left( m_{ij,t}(\omega) x_{ij,t-1}^{\chi_{ij}} \right)^{1-1/\eta} d\omega \right]^{1/(1-1/\eta)}, \tag{A.1}$$

The distributor operates in a perfectly competitive market structure and minimizes production costs by choosing the amount of wholesale varieties to import subject to the above CES technology:

$$\begin{aligned} \min_{\{m_{ij,t}(\omega)\}} \quad & \tilde{P}_{ij,t} x_{ij,t} - \int_0^1 P_{ij,t}(\omega) m_{ij,t}(\omega) d\omega \\ \text{s.t.} \quad x_{ij,t} = & \left[ \int_0^1 \left( m_{ij,t}(\omega) x_{ij,t-1}^{\chi_{ij}} \right)^{1-1/\eta} d\omega \right]^{1/(1-1/\eta)}. \end{aligned}$$

The first order condition is given by:

$$\begin{aligned} & \tilde{P}_{ij,t} x_{ij,t}^{1/\eta} (m_{ij,t}(\omega) x_{ij,t-1}^{\chi_{ij}})^{-1/\eta} x_{ij,t-1}^{\chi_{ij}} - P_{ij,t}(\omega) = 0, \\ & \Rightarrow m_{ij,t}(\omega) = \left[ \frac{P_{ij,t}(\omega)}{\tilde{P}_{ij,t}} \right]^{-\eta} x_{ij,t} x_{ij,t-1}^{\chi_{ij}(\eta-1)}, \end{aligned}$$

Distributors break-even when the total revenue is equal to the total costs:

$$\tilde{P}_{ij,t} x_{ij,t} = \int_0^1 P_{ij,t}(\omega) m_{ij,t}(\omega) d\omega \tag{A.2}$$

Substituting the above demand for wholesale varieties into this 'zero-profit' condition gives rise to the the break-even price index of the distributors:

$$\tilde{P}_{ij,t} = \left[ \int_0^1 (P_{ij,t}(\omega) x_{ij,t-1}^{-\chi_{ij}})^{1-\eta} d\omega \right]^{1/(1-\eta)}. \tag{A.3}$$

**Symmetric Equilibrium.** In the symmetric equilibrium, the break-even price index boils down to $\tilde{P}_{ij,t} = P_{ij,t} x_{ij,t-1}^{-\chi_{ij}}$, such that $P_{ij,t}/\tilde{P}_{ij,t} = x_{ij,t-1}^{\chi_{ij}}$. Consequently, demand for wholesale varieties in the symmetric equilibrium is given by

$$m_{ij,t} = x_{ij,t} x_{ij,t-1}^{\chi_{ij}(\eta-1)} \underbrace{\left[ \frac{P_{ij,t}}{\tilde{P}_{ij,t}} \right]^{-\eta}}_{= x_{ij,t-1}^{-\chi_{ij}\eta}} d\omega = x_{ij,t} x_{ij,t-1}^{-\chi_{ij}}, \tag{A.4}$$

such that the real output of the distributor, which represents real trade flows, is persistent and given by $x_{ij,t} = m_{ij,t} x_{ij,t-1}^{\chi_{ij}}$.

## A.2 Consumer

### A.2.1 Intra-Temporal Optimisation

The representative consumer minimizes consumption expenditure on composite goods sourced by the distributor of each country subject to CES preferences:

$$\begin{aligned} \min_{\{x_{ij,t}\}} \; & P_{j,t}c_{j,t} - \sum_{i=1}^{N} \tilde{P}_{ij,t}x_{ij,t} \\ \text{s.t.} \;\; & c_{j,t} = \left[\sum_{i=1}^{N} x_{ij,t}^{1-1/\eta}\right]^{1/(1-1/\eta)}. \end{aligned}$$

The first-order condition with respect to the demand for a composite good $x_{ij,t}$ from any source country $i = 1, ..., N$ is given by

$$\begin{aligned} & P_{j,t}c_{j,t}^{1/\eta}x_{ij,t}^{-1/\eta} - \tilde{P}_{ij,t} = 0, \\ \Rightarrow \;\; & x_{ij,t} = c_{j,t}\left(\frac{\tilde{P}_{ij,t}}{P_{j,t}}\right)^{-\eta}. \end{aligned}$$

The consumer price index, measuring the aggregate cost of living, is then derived by substituting the above demand schedule back into the CES preferences:

$$P_{j,t} = \left[\sum_{i=1}^{N} \tilde{P}_{ij,t}^{1-\eta}\right]^{1/(1-\eta)}. \tag{A.5}$$

### A.2.2 Inter-Temporal Optimisation

Consumers maximize utility subject to an indefinite sequence of budget constraints:

$$\max_{\{c_{j,t}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \log\left(c_{j,t}\right) \quad \text{s.t.} \quad c_{j,t} + \mathbb{E}_t[\zeta_{j,t,t+1}b_{j,t+1}] = b_{j,t} + w_{j,t}h_j.$$

Using the standard no-Ponzi scheme condition, we re-write this as a Current Value Lagrangian:

$$\max_{\{c_{j,t}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \left[\beta^t \log\left(c_{j,t}\right) + \Lambda_{j,0}\zeta_{j,0,t}\left(w_{j,t}h_j + \varpi_{j,t} - c_{j,t} - b_{j,0}\right]\right).$$

First order condition:

$$\frac{\beta^t}{c_t} - \Lambda_{j,0}\zeta_{j,0,t} = 0 \quad \Leftrightarrow \quad \frac{\beta^t}{c_t} = \Lambda_{j,0}\zeta_{j,0,t}. \tag{A.6}$$

Analogous first order condition holds for any time period $t = \{0, 1, 2, ...\}$. Therefore,

$$\frac{\beta^{t+1}}{\beta^t}\frac{c_{t+1}}{c_t} = \frac{\Lambda_{j,0}\zeta_{j,0,t+1}}{\Lambda_{j,0}\zeta_{j,0,t}}, \qquad \Leftrightarrow \qquad 1 = \beta\frac{c_{t+1}}{\zeta_{j,t,t+1}c_t}, \qquad \Leftrightarrow \qquad 1 = \beta\mathbb{E}_t\left[\frac{c_{t+1}}{\zeta_{j,t,t+1}c_t}\right], \tag{A.7}$$

since $\zeta_{j,t,t+1} = \zeta_{j,0,t+1}/\zeta_{j,0,t}$. With perfect capital mobility, $\Lambda_{j,0}\zeta_{j,0,t} = \Lambda_{i,0}\zeta_{i,0,t}$ is the no-arbitrage condition, where $\Lambda_{j,0}$ and $\Lambda_{i,0}$ are the 'initial conditions'.

### A.2.3 Transversality Condition

Consider iterating the household budget constraint forwards in the symmetric equilibrium:

$$
\begin{aligned}
b_{j,t} =& \zeta_{j,t,t+1}\mathbb{E}_t[b_{j,t+1} + \underbrace{c_{j,t} - \varpi_{j,t} - w_{j,t}h_j}_{nx_{j,t}}], \\
=& \zeta_{j,t,t+1}\mathbb{E}_t\left[\zeta_{j,t+1,t+2}(b_{j,t+2} - nx_{j,t+1}) - nx_{j,t}\right], \\
=& \zeta_{j,t,t+1}\zeta_{j,t+1,t+2}b_{j,t+2} - \zeta_{j,t,t+1}(nx_{j,t} + \zeta_{j,t+1,t+2}nx_{j,t+1}), \\
\dots& \\
=& \zeta_{j,t,t+S}b_{j,t+S} - \sum_{s=0}^{S}\zeta_{j,t,t+s+1}nx_{j,t+s}.
\end{aligned}
\tag{A.8}
$$

Next, note that the stochastic discount factor $\zeta_{j,t,t+S} \in (0,1)$ for all $s = 1,2,...,S$ as long as the real rate of interest is strictly non-negative. Assuming that foreign economies would only be willing to lend to the domestic economy at a positive rate of interest, it follows that

$$
\lim_{S\to\infty}\zeta_{j,t,t+S} = \zeta_{j,t,t+1} \times \zeta_{j,t+1,t+2} \times \zeta_{j,t+2,t+3} \times \cdots \times \zeta_{j,S-1,S} = 0. \tag{A.9}
$$

As a result, the stock of debt is clearly non-explosive. To fully convince yourself, consider evaluating the iterated form of the budget constraint along the balanced growth path:

$$
b_j = \beta^S b_j - nx_j\sum_{s=0}^{S}\beta^{1+s} \quad\Rightarrow\quad \lim_{S\to\infty} b_j = -nx_j\sum_{s=0}^{\infty}\beta^{1+s}, \tag{A.10}
$$

$$
= -nx_j\left(\frac{\beta}{1-\beta}\right) > -\infty.\Big|_{\beta\in(0,1)} \tag{A.11}
$$

## A.3 Wholesalers

Wholesalers are perfectly competitive and choose how much labor to hire in order to minimize production costs subject to linear production technology:

$$
\begin{aligned}
\min_{\{h_{ii,t}(\omega)\}}\ & P_{ii,t}(\omega)m_{ii,t} - W_{i,t}h_{ii,t}(\omega) \\
\text{s.t. }\ & m_{ii,t}(\omega) = z_{i,t}h_{ii,t}(\omega).
\end{aligned}
$$

The first order condition is given by

$$
\begin{aligned}
P_{ii,t}(\omega)z_{i,t} - W_{i,t} &= 0, \\
\Rightarrow P_{ii,t}(\omega) &= \underbrace{\frac{W_{i,t}}{z_{i,t}}}_{\text{Marginal Cost}}.
\end{aligned}
\tag{A.12}
$$

Assuming that international arbitrageurs are subject to the iceberg trade costs, but otherwise efficient, implies that

$$
P_{ij,t}(\omega) = d_{ij}P_{ii,t}(\omega) = \frac{d_{ij}W_{i,t}}{z_{i,t}}. \tag{A.13}
$$

# B Proofs

## B.1 Proposition 1

**Proposition.** *Let* $\theta_{i,t} = Y_{i,t}/Y_t$*, where* $Y_t = \sum_{j=1}^{N} Y_{j,t}$*. Then the share of the source country aggregate income relative to the world income is a function of its export prices and the outward multilateral resistance:*

$$\theta_{i,t} = (\Phi_{i,t} P_{ii,t})^{1-\eta},$$

*where*

$$\Phi_{i,t} = \left[ \sum_{j=1}^{N} \theta_{j,t} \Xi_{j,t} \left( \frac{d_{ij} x_{ij,t-1}^{-\chi_{ij}}}{P_{j,t}} \right)^{1-\eta} \right]^{1/(1-\eta)}$$

*Proof.* Consider the optimal demand for imports, the aggregate consumption identity, and the break-even price index, respectively:

$$X_{ij,t} = \tilde{P}_{ij,t} x_{ij,t} = C_{j,t} \left[ \frac{\tilde{P}_{ij,t}}{P_{j,t}} \right]^{1-\eta}, \tag{B.1}$$

$$C_{j,t} = Y_{j,t} \Xi_{j,t}, \tag{B.2}$$

$$\tilde{P}_{ij,t} = P_{ij,t} x_{ij,t-1}^{-\chi_{ij}}. \tag{B.3}$$

Now substitute (B.3) and (B.2) into (B.1) to obtain

$$X_{ij,t} = Y_{j,t} \Xi_{j,t} \left[ \frac{P_{ij,t} x_{ij,t-1}^{-\chi_{ij}}}{P_{j,t}} \right]^{1-\eta}. \tag{B.4}$$

Next, substitute (B.4) into the aggregate income identity $Y_{i,t} = \sum_{j=1}^{N} M_{ij,t} \equiv \sum_{j=1}^{N} X_{ij,t}$ and replace import prices $P_{ij,t}$ with scaled export prices $d_{ij} P_{ii,t}$ (see (A.13)) to obtain

$$Y_{i,t} = P_{ii,t}^{1-\eta} \sum_{j=1}^{N} Y_{j,t} \Xi_{j,t} \left[ \frac{d_{ij} x_{ij,t-1}^{-\chi_{ij}}}{P_{j,t}} \right]^{1-\eta}, \tag{B.5}$$

Finally, let $\theta_{i,t} = Y_{i,t}/Y_t$, where $Y_t = \sum_{j=1}^{N} Y_{j,t}$. Then (B.5) can be written as

$$P_{ii,t}^{1-\eta} = \frac{Y_{i,t}}{\sum_{j=1}^{N} Y_{j,t} \Xi_{j,t} \left[ \frac{d_{ij} x_{ij,t-1}^{-\chi_{ij}}}{P_{j,t}} \right]^{1-\eta}} = \frac{\theta_{i,t}}{\sum_{j=1}^{N} \theta_{j,t} \Xi_{j,t} \left[ \frac{d_{ij} x_{ij,t-1}^{-\chi_{ij}}}{P_{j,t}} \right]^{1-\eta}} = \theta_{i,t} \Phi_{i,t}^{\eta-1}, \tag{B.6}$$

where

$$\Phi_{i,t} = \left[ \sum_{j=1}^{N} \theta_{j,t} \Xi_{j,t} \left( \frac{d_{ij} x_{ij,t-1}^{-\chi_{ij}}}{P_{j,t}} \right)^{1-\eta} \right]^{1/(1-\eta)} \tag{B.7}$$

is the ‘multilateral resistance’ of destination $i$ to trade flows from all source countries $j \in n \setminus i$. □

## B.2 Lemma 1

**Lemma.** *The gravity equation is dynamic when there is learning-by-importing, such that $\chi_{ij} > 0$ for all $i \in n \backslash j$. And when learning-by-importing is asymmetric across countries, such that $\chi_{ij} \neq \chi_{ji}$ for all $i \in n \setminus j$, and/or the iceberg costs are asymmetric, such that $d_{ij} \neq d_{ji} > 1$ for all $i \in n \setminus j$, the gravity equation features the multilateral trade imbalances:*

$$A_{ij,t} = \Xi_{j,t} \left[ \frac{d_{ij}}{\Phi_{i,t} P_{j,t}} \right]^{1-\eta} \prod_{s=1}^{S} \left( \frac{d_{ij} Y_{i,t-s}^{\eta/(1-\eta)}}{\Phi_{i,t-s} A_{ij,t-s} Y_{j,t-s} Y_{t-s}^{\eta/(1-\eta)}} \right)^{\chi_{ij}^{s}(1-\eta)},$$

*where $A_{ij,t} := (X_{ij,t} Y_t)/(Y_{i,t} Y_{j,t})$ measures the 'size-adjusted' bilateral trade flows.*

*Proof.* Substitute (A.13) and (B.6) into (B.4) to obtain

$$A_{ij,t} = \Xi_{j,t} \left[ \frac{d_{ij}}{\Phi_{i,t} P_{j,t}} \right]^{1-\eta} \left( \frac{\tilde{P}_{ij,t-1}}{X_{ij,t-1}} \right)^{\chi_{ij}(1-\eta)}. \tag{B.8}$$

Next, notice that the break-even price index (B.3) is recursive, such that using (B.6) we obtain

$$\begin{aligned}
\tilde{P}_{ij,t} &= P_{ij,t} x_{ij,t-1}^{-\chi_{ij}}, \\
&= \frac{d_{ij}}{\Phi_{i,t}} \theta_{i,t}^{1/(1-\eta)} X_{ij,t-1}^{-\chi_{ij}} \tilde{P}_{ij,t-1}^{\chi_{ij}}, \\
&= \frac{d_{ij}}{\Phi_{i,t}} \theta_{i,t}^{1/(1-\eta)} X_{ij,t-1}^{-\chi_{ij}} \left( \frac{d_{ij}}{\Phi_{i,t-1}} \theta_{i,t-1}^{1/(1-\eta)} X_{ij,t-2}^{-\chi_{ij}} \tilde{P}_{ij,t-2}^{\chi_{ij}} \right)^{\chi_{ij}}, \\
&= \frac{d_{ij}^{1+\chi_{ij}}}{\Phi_{i,t} \Phi_{i,t-1}^{\chi_{ij}}} \theta_{i,t}^{1/(1-\eta)} \theta_{i,t-1}^{\chi_{ij}\eta/(1-\eta)} \underbrace{A_{ij,t-1}^{-\chi_{ij}} Y_{j,t-1}^{-\chi_{ij}}}_{X_{ij,t-1}^{-\chi_{ij}} \theta_{i,t-1}^{\chi_{ij}}} X_{ij,t-2}^{-\chi_{ij}^2} \tilde{P}_{ij,t-2}^{\chi_{ij}^2}, \\
&\vdots \\
&= \theta_{i,t} \prod_{s=0}^{t-1} \left( \frac{d_{ij}}{\Phi_{i,t-s}} \theta_{i,t-s}^{\eta/(1-\eta)} A_{ij,t-s-1}^{-\chi_{ij}} Y_{j,t-s-1}^{-\chi_{ij}} \right)^{\chi_{ij}^{s}} \tilde{P}_{ij,0}^{\chi_{ij}^{t-1}},
\end{aligned} \tag{B.9}$$

Suppose $\lim_{s\to\infty} \tilde{P}_{ij,0}^{\chi_{ij}^{s}} = 0$, which requires that $0 < \chi_{ij} < 1$. Then notice that

$$\frac{\tilde{P}_{ij,t}}{X_{ij,t}} = \underbrace{A_{ij,t}^{-1} Y_{j,t}^{-1}}_{\frac{\theta_{i,t}}{X_{ij,t}}} \prod_{s=0}^{t-1} \left( \frac{d_{ij}}{\Phi_{i,t-s}} \theta_{i,t-s}^{\eta/(1-\eta)} A_{ij,t-s-1}^{-\chi_{ij}} Y_{j,t-s-1}^{-\chi_{ij}} \right)^{\chi_{ij}^{s}} \equiv \prod_{s=0}^{t-1} \left( \frac{d_{ij} \theta_{i,t-s}^{\eta/(1-\eta)}}{\Phi_{i,t-s} A_{ij,t-s} Y_{j,t-s}} \right)^{\chi_{ij}^{s}}. \tag{B.10}$$

Therefore, substituting (B.10) back into (B.8) gives the dynamic gravity equation:

$$A_{ij,t} = \Xi_{j,t} \left[ \frac{d_{ij}}{\Phi_{i,t} P_{j,t}} \right]^{1-\eta} \prod_{s=1}^{S} \left( \frac{d_{ij} Y_{i,t-s}^{\eta/(1-\eta)}}{\Phi_{i,t-s} A_{ij,t-s} Y_{j,t-s} Y_{t-s}^{\eta/(1-\eta)}} \right)^{\chi_{ij}^{s}(1-\eta)}, \tag{B.11}$$

where we make use of $\theta_{i,t} = Y_{i,t}/Y_t$ and assume that $S = t - 1$. □

## C Empirical Representation

The theoretical learning-by-importing gravity model is given by

$$\begin{aligned}
\ln A_{ij,t} =&(\eta-1)\sum_{s=1}^{S}\chi_{ij}^{s}\ln A_{ij,t-s}+\ln\Xi_{j,t}-(\eta-1)\sum_{s=0}^{S}\chi_{ij}^{s}\ln d_{ij}\\
&+(\eta-1)\sum_{s=0}^{S}\chi_{ij}^{s}\ln\Phi_{i,t-s}+(\eta-1)\ln P_{j,t}-\eta\sum_{s=1}^{S}\chi_{ij}^{s}\ln Y_{t-s}\\
&+\eta\sum_{s=1}^{S}\chi_{ij}^{s}\ln Y_{i,t-1}+(\eta-1)\sum_{s=1}^{S}\chi_{ij}^{s}\ln Y_{j,t-1}.
\end{aligned}$$

If $S=1$, then

$$\begin{aligned}
\ln A_{ij,t} =&(\eta-1)\chi_{ij}\ln A_{ij,t-1}+\ln\Xi_{j,t}-(\eta-1)\left(1+\chi_{ij}\right)\ln d_{ij}\\
&+(\eta-1)\ln\Phi_{i,t}+(\eta-1)\chi_{ij}\ln\Phi_{i,t-1}+(\eta-1)\ln P_{j,t}-\eta\chi_{ij}\ln Y_{t-1}\\
&+\eta\chi_{ij}\ln Y_{i,t-1}+(\eta-1)\chi_{ij}\ln Y_{j,t-1}.
\end{aligned}$$

If $S\to\infty$, then

$$\begin{aligned}
\ln A_{ij,t} =&(\eta-1)\sum_{s=1}^{\infty}\chi_{ij}^{s}\ln A_{ij,t-s}+\ln\Xi_{j,t}-(\eta-1)\sum_{s=0}^{\infty}\chi_{ij}^{s}\ln d_{ij}\\
&+(\eta-1)\sum_{s=0}^{\infty}\chi_{ij}^{s}\ln\Phi_{i,t-s}+(\eta-1)\ln P_{j,t}-\eta\sum_{s=1}^{\infty}\chi_{ij}^{s}\ln Y_{t-s}\\
&+\eta\sum_{s=1}^{\infty}\chi_{ij}^{s}\ln Y_{i,t-1}+(\eta-1)\sum_{s=1}^{\infty}\chi_{ij}^{s}\ln Y_{j,t-1}.
\end{aligned}$$

Notice that

$$\textstyle\sum_{s=0}^{\infty}\chi_{ij}^{s}=1+\chi_{ij}+\chi_{ij}^{2}+\ldots=\frac{1}{1-\chi_{ij}},\ \chi_{ij}\in(0,1).$$

Hence,

$$\begin{aligned}
\ln A_{ij,t} =&(\eta-1)\sum_{s=1}^{\infty}\chi_{ij}^{s}\ln A_{ij,t-s}+\ln\Xi_{j,t}\\
&-\frac{\eta-1}{1-\chi_{ij}}\ln d_{ij}+(\eta-1)\ln\Phi_{i,t}+(\eta-1)\ln P_{j,t}+(\eta-1)\sum_{s=1}^{\infty}\chi_{ij}^{s}\ln\Phi_{i,t-s}\\
&-\eta\sum_{s=1}^{\infty}\chi_{ij}^{s}\ln Y_{t-s}+\eta\sum_{s=1}^{\infty}\chi_{ij}^{s}\ln Y_{i,t-1}+(\eta-1)\sum_{s=1}^{\infty}\chi_{ij}^{s}\ln Y_{j,t-1}.
\end{aligned}\tag{C.1}$$

Using the lag operator $L$, such that $L^{s}x_{t}=x_{t-s}$,$s=0,1,\ldots$, and $\chi_{ij}\left(L\right)=\chi_{ij}L+\chi_{ij}^{2}L^{2}+\ldots$, we obtain

$$\begin{aligned}
\ln A_{ij,t} =&(\eta-1)\chi_{ij}\left(L\right)\ln A_{ij,t}+\ln\Xi_{j,t}-\frac{\eta-1}{1-\chi_{ij}}\ln d_{ij}\\
&+(\eta-1)\ln\Phi_{i,t}+(\eta-1)\ln P_{j,t}+(\eta-1)\chi_{ij}\left(L\right)\ln\Phi_{i,t}\\
&-\eta\chi_{ij}\left(L\right)\ln Y_{t}+\eta\chi_{ij}\left(L\right)\ln Y_{i,t}+(\eta-1)\chi_{ij}\left(L\right)\ln Y_{j,t}.
\end{aligned}$$

It it is clear that the distance elasticity is amplified by the learning-by-importing externality. The higher is $\chi_{ij}$, the more prohibitive is the bilateral trade friction on trade flows. Since our key focus in this paper is on the short-run or business cycle adjustments, we will focus on the parsimonious regressions with the first lags instead of truncating each lag polynomial at a potentially different lag for a different trade pair or country and variable. The studies that focus on the long-run can fruitfully make use of the richness of the proposed model, which we leave for future investigations.

# D Data Facts

We zoom into the experience of different countries that were hidden in the global variables in Figure 4. To ease the reading of many country groups, we focus on the well-documented GTC episode, yet our theory-driven empirical model is not limited to any one common shock and covers more than six decades of data. Figure 5 visualizes the dynamics of export value from several major country groups before, during, and after the GTC. Specifically, it depicts the export value indices over the period of 2000-2014 for the global economy, the US, the EU, and other selected groups of countries, such as Brazil, Russia, India and China (abbreviated as BRICs), the group of seven (G7), and a cohort of all emerging and developing countries in our sample. Three empirical observations stand out the most. First, independently of the country group, the value of international trade declines in a sharp and synchronized fashion in response to the 2008-09 shock (see the shaded area of Figure 5). Second, the speed of recovery from the common shock is remarkably heterogeneous compared to the pre-shock trend growth. In particular, the BRICs recovered most rapidly followed by emerging and developing countries, leaving the EU and G7 well behind. In other words, even after removing pre-trends, the after-shock dynamics are far from homogeneous, thereby necessitating further explanations about heterogeneous responses to the same global shocks. Last, as already documented for the past six decades in the main text, the value of international trade is substantially more volatile than aggregate income, which has declined by only around 1% globally during this time (not displayed). This indicates relatively low persistence in the value of international trade, particularly in response to common shocks, thereby characterizing the essence of what we call the "trade persistence" puzzle.

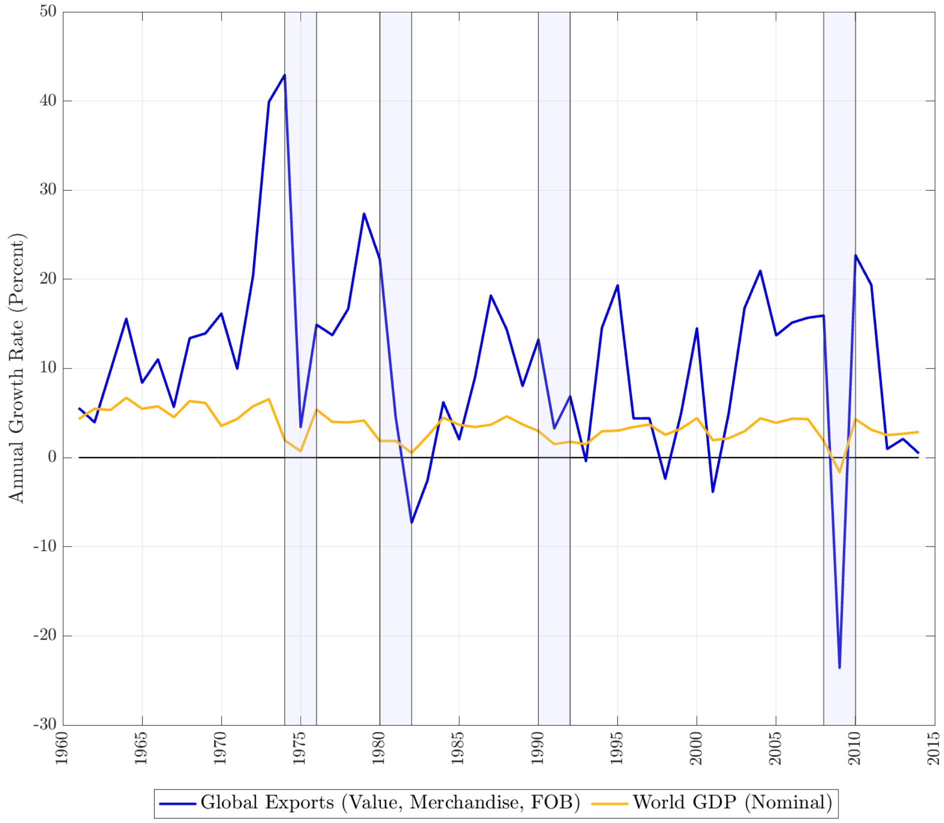



**Figure 4: Dynamics of Global Trade and GDP Growth Rates**

*Notes:* The figure depicts the annual growth rates (in percent) of the global merchandise (FOB) exports, extracted from the World Trade Organization, and the world GDP, taken from the World Bank. The shaded areas depict "global recessions", as identified by Kose et al. (2020).

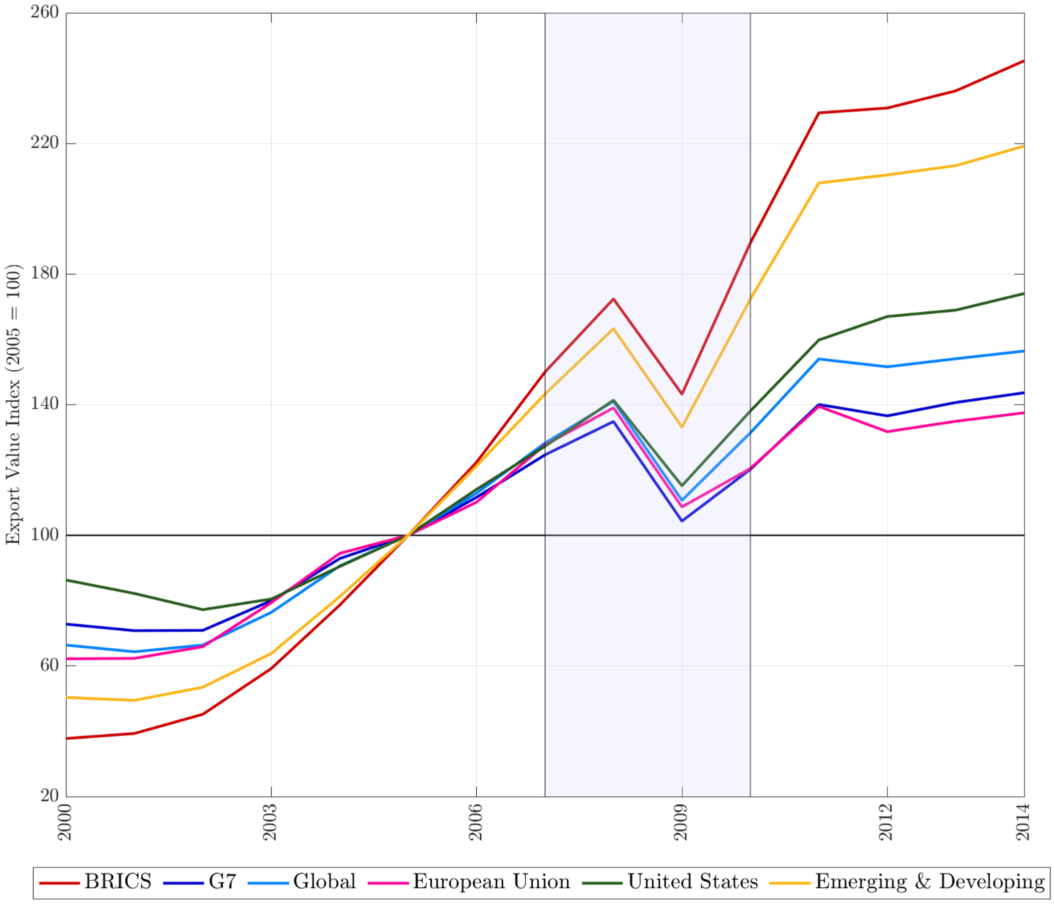



**Figure 5: Trade Flows in Major Country Groups**

*Notes:* The figure depicts the export value indices over the period of 2000-2014 for the global economy, the US, the EU, and other selected groups of countries, such as Brazil, Russia, India and China (abbreviated as BRICs), the group of seven (G7), and a cohort of all emerging and developing countries in our sample. The reference year is 2005 when the index value is equal to 100. The shaded area is a time period known as the "Great Trade Collapse" (GTC).

# E Empirical Models and Results

## E.1 Panel Data Estimators of Gravity Models

In this section, we describe and motivate possible alternative techniques to model trade gravity using panel data. We pay special attention to unobservable bilateral and multilateral trade resistance, starting with the most widely used in the literature to end with our preferred one, selected to best match the proposed dynamic gravity equation.

There exist several alternative techniques of modeling the unobservable bilateral and multilateral trade resistance empirically. In the conventional static gravity equation (2.19), bilateral trade resistance ($d_{ij}$) is time-invariant, while multilateral trade resistance ($P_{j,t}$ and $\Phi_{i,t}$) is static. Starting with Feenstra (2016), a large stream of the trade literature adopted a panel regression model with: (i) unobservable time-invariant heterogeneity (i.e., country fixed effects); and (ii) an unobservable homogeneous trend (i.e., time fixed effects). Both country and time fixed effects are expected to simultaneously capture the unobservable inward and outward trade resistance for each country pair and for each time period. We call the conventional strategy as the "Fixed Effects" (FE) approach. The antithesis of the conventional FE approach is to ignore all unobservable bilateral and multilateral trade resistance altogether, imposing independence across all country pairs. Specifically, following Pesaran and Smith (1995), given a relatively large time dimension $T$ in our sample, we can estimate $(N-1)N$ number of country-specific time series, one for each unique country pair, and average all of the coefficient estimates across all of the country pairs. We call this restrictive strategy the "Mean Group" (MG) approach.

Unlike the FE estimator, which provides pooled coefficient estimates that are homogeneous for all country pairs, the key advantage of the MG estimator is that it reflects the observed cross-sectional heterogeneity of the panel by generating coefficient estimates specific to each country pair. In the context of the learning-by-importing gravity equation, this means that the MG estimator allows to retain more details, distinguishing between country pairs for which trade flows are persistent and unbalanced, and those that are not. The FE estimator, on the other hand, rather "paints with a broad brush". Heterogeneous trade persistence is a property we wish to retain in our empirical estimates, given that the learning-by-importing-augmented gravity equation (2.19) predicts a heterogeneous trade persistence coefficient across different country pairs (i.e., $\chi_{ij}(\eta-1)$). However, the main disadvantage of the MG approach is that it accounts for neither time-varying nor time-invariant *unobservable* heterogeneity. Specifically, if we take the inference drawn about the coefficient of trade persistence based on the rudimentary MG estimator at face value, then it is as if geographic distance between countries or their overall propensity to trade have no differential impact on the degree of trade persistence across any country pairs. As a consequence, we also consider, what we coin as a "Hybrid Fixed Effects" (HFE) approach, which reflects both the observed and the time-invariant unobserved heterogeneity of the panel in addition to retaining parameter heterogeneity.

There are two important reasons why, despite their popularity, none of the aforementioned approaches are chosen as the preferred technique in this paper. First, the homogeneous time fixed effects do not appropriately reflect the fact that the unobservable time-varying multilateral resistance can be strongly correlated with observable regressors in the learning-by-importing-augmented gravity equation. In practice, we have every reason to believe that it is indeed the case. Our theoretical model endogenously links the multilateral resistance to trade flows, trade imbalance, and aggregate income (see equation (2.16)). Our empirical results provide additional support for this hypothesis, which we discuss below in Section 3.3. In fact, Anderson and Yotov (2010) and Anderson (2011) argue that the unobservable inward and the outward multilateral resistance may be heterogeneous across different country pairs. And if so, then the correlation between the observable regressors and the unobservable time-varying inward and the outward multilateral resistance may also be heterogeneous, which is not addressed by the time fixed effect approach (see Kapetanios et al. (2017)). Moreover, the learning-by-importing-augmented gravity equation predicts a heterogeneous coefficient next to lagged multilateral resistance (i.e., factor loading) that is specific to each country pair (see equation (2.18)). This means that even if the time fixed effects are common for all country pairs, their effect on bilateral trade flows is nonetheless specific to each country pair. Further, the learning-by-importing-augmented gravity equation is dynamic, not static as is traditionally the case. And while this may seem rather innocuous, Pesaran and Smith (1995) show that neglected parameter heterogeneity associated with the FE approach generates biased and even inconsistent coefficient estimates when the panel regression model is indeed dynamic. This observation is particularly alarming, since the existing trade literature tends to ignore parameter heterogeneity in spite of the three-dimensional data structure, which comprises of the source country, the destination country, and time.

Since dynamic multilateral resistance terms, entering with the trade pair specific intensity parameter $\chi_{ij}$, is crucial for consistent estimation and construction of valid counterfactuals, we take the "Common Correlated Effects Mean Group" (CCEMG) estimator as our baseline choice.[27] Following Pesaran (2006), Kapetanios et al. (2011), and Chudik and Pesaran (2015), CCEMG replaces unobservable factors, which are dynamic multilateral resistance terms in our appliation, with a set of cross-sectional averages of all regressors entering the learning-by-importing-augmented gravity equation. The CCEMG estimator accounts for the fact that the unobservable time-varying multilateral resistance is dynamic, not static (i.e., both $\Phi_{i,t}$ and $\Phi_{i,t-1}$ are controlled for). It does so by explicitly incorporating cross-sectional averages of the contemporaneous as well as lagged values of all variables as additional regressors.

In addition to a standard CCEMG approach, we also consider the "Restricted Common Correlated Effects Mean Group" (CCEMGR), in which the vector of unobservable dynamic multilateral resistance terms is based solely on the cross-sectional averages of the contemporaneous and lagged trade flows. We explore this option since, in theory, if $N$ is sufficiently large, the cross-sectional average of the trade imbalance variable $\mathrm{TB}_{j,t}$ tends to unity, because the net trade flows of the world economy as a whole are always balanced. Similarly, the cross-sectional averages of aggregate income are strongly related to the world aggregate income, which enters the dynamic gravity equation by default (see equation (2.18)).

Lastly, in order to gauge the relative importance of the "trade persistence puzzle" drivers, we also employ the Pooled Common Correlated Effects (CCEP) approach, which ignores the intrinsic para-

[27] The main alternative estimator to CCEMG is the "Augmented Mean Group" (AMG) as in Eberhardt and Teal (2013). However,the latter consists in 2 steps with the first one not taking heterogeneity into consideration, and incorporates only a single unobserved factor. More details are available in Section 3.2 and Appendix E.2.

meter heterogeneity, but incorporates the unobservable dynamic multilateral resistance terms.[28]

**Alternative fixed effects and Poisson.** For completeness and robustness, we consider other variations of the FE approach commonly adopted in the literature (e.g., Piermartini and Yotov (2016); Anderson and Yotov (2020)). While the aforementioned FE approach incorporates both country and time fixed effects, it assumes that all countries share a homogeneous time trend component and it does not fully account for the time-invariant heterogeneity specific to each country pair. For this reason, the FE2 approach replaces the country and time fixed effects with the so-called "time-varying" fixed effects, which allows for a heterogeneous time trend component specific to each country. The FE3 approach applies the standard time fixed effects as does the conventional FE approach, but it also controls for the time-invariant heterogeneity specific to each country pair. Controlling for the latter should fully incorporate the contribution of the country fixed effects. And finally, the FE4 approach controls for both the heterogeneous time trend, specific to source and destination economies, as well as time-invariant heterogeneity specific to each country pair, which replaces again the country-specific time-invariant heterogeneity (see Section 3.2 for technical details). As demonstrated in the main text, these alternative methods can be nested within the same empirical model. Finally, we also report a popular estimator in the context of gravity models, namely the Poisson Pseudo-Maximum-Likelihood (PPML), see Santos-Silva and Tenreyro (2007); Westerlund and Wilhelmsson (2009). It is a nonlinear alternative, unlike all other linear estimators, and cannot be nested into the same framework. The existing PPML framework is not developed for heterogeneous parameters, dynamic unobserved factors and cannot account for zero trade flows entering as explanatory variables. Due to dynamic structure, endogenous trade imbalances, and zero lagged trade flows, the benefits of the PPML applicable in the static framework, as spelled out in Fally (2015), cease to apply. As a result, we stick to dynamic (log)linear models. We discuss dynamic extension of the PPML and related issues in more detail in Section E.3.

## E.2 Other Estimation Methodologies

In addition to the above covered estimators, we also recognize two further MG-based techniques that are able to not only reconcile parameter heterogeneity, but also proxy the unobservable time-varying multilateral trade resistance specific to each country pair.[29] The first technique is known as the "Augmented Mean Group" (AMG) estimator. Following Eberhardt and Teal (2013), AMG involves estimating the standard FE regression model with individual and time fixed effects, extracting the pooled time fixed effect coefficients for each time period, and then using their time series as an additional regressor (i.e., unobservable common factor) in an otherwise standard MG regression model. Consequently, the AMG coefficient estimates are heterogeneous for each country pair, analogous to the regular MG approach. But unlike the regular MG estimator, the AMG coefficient

[28]The CCEP estimator is also biased, unlike CCEMG, unless the cross-sectional dimension dominates time periods (as is the case with the trade pairs), see De Vos and Stauskas (2021). Luckily, the CCEP estimator remains consistent under various assumptions about the unobserved dynamic factor (multilateral resistance term) structure and properties, albeit asymptotic normality may require additional conditions; refer to Juodis et al. (2021). Our setting is more complicated due to the lagged trade flow, requiring relatively large number of trade pairs and time periods to deal with heterogeneity and persistence.

[29]There is a large literature on dynamic factor models with homogeneous parameters in a panel setting, e.g., Forni et al. (2000); Bai (2009), among others.

estimates also reflect the fact that the unobservable common factor exerts a heterogeneous influence on bilateral trade flows for each country pair. The second technique is referred to as the "Common Correlated Effects Mean Group" (CCEMG) estimator. Following Pesaran (2006), Kapetanios et al. (2011), and Chudik and Pesaran (2015), CCEMG replaces the pre-estimated homogeneous time fixed effects with a proxy for a vector of unobservable common factors, which then enter the panel regression model as additional regressors. Specifically, CCEMG measures the unobservable common factors as a cross-sectional average of all regressors entering the learning-by-importing-augmented gravity equation.

Despite the flexibility of the AMG estimator, there are two reasons why our preferred approach of estimating the learning-by-importing-augmented gravity equation is the CCEMG estimator. First, if the pre-estimated pooled regression coefficients in the AMG approach are inconsistent due to the fact that our panel regression model is inherently dynamic with heterogeneous coefficients, then the inference drawn from the subsequent country pair-specific regressions is misleading because it inherits the inconsistencies from the pre-estimation stage. Second, the learning-by-importing-augmented gravity equation in equation (2.18) incorporates four types of unobservable trade resistance (i.e., $d_{ij}$, $P_{j,t}$, $\Phi_{i,t}$, and $\Phi_{i,t-1}$). Unlike the aforementioned techniques, CCEMG accounts for the fact that the unobservable time-varying multilateral resistance is dynamic, not static (i.e., both $\Phi_{i,t}$ and $\Phi_{i,t-1}$ are controlled for). It does so by explicitly incorporating proxies for the contemporaneous and lagged unobservable time-varying multilateral trade resistance. Those proxies are the cross-sectional averages of the contemporaneous as well as lagged bilateral trade flows, which enter the learning-by-importing-augmented gravity equation as additional regressors through the vector of unobservable common factors.

### E.3 Threats to Results Validity of the Baseline Model

All approaches described in the main text draw inference about the regression coefficients from a log-linear specification of the learning-by-importing-augmented gravity equation. But we admit that all log-linear applications of the bilateral trade flow data entail one simple caveat, which is commonly referred to as the "zero trade problem" due to Santos-Silva and Tenreyro (2007). Specifically, given that our dataset comprises of $N = 39$ and $T = 65$, around 10% of total observations $TN(N-1)$ in our sample contain zero entries. This finding documents the fact that the bilateral trade flows between a subset of country pairs during a subset of consecutive time periods were either unrecorded or non-existent. And if so, then the cross-sectional heteroscedasticity caused by the zero entries leads to at least somewhat biased and inconsistent coefficient estimates. A common approach to address the zero trade problem is to use the Poisson Pseudo-Maximum-Likelihood (PPML) approach, which estimates the regression model in a multiplicative form (e.g., Santos-Silva and Tenreyro (2007); Westerlund and Wilhelmsson (2009)). While we incorporate the results from the PPML specification as yet another tentative approach, we recognize several reasons why the results from the CCEMG approach are generally preferred to those of the PPML approach in the context of our empirical application.

First of all, the CCEMG estimator in principle allows the error structure to exhibit unknown heteroscedasticity over time, so long as it is subject to a finite order of integration (see Westerlund (2018)). Second, given the large $N$ and large $T$ nature of our panel, we argue that the observed cross-sectional heteroscedasticity is dominated by the time-varying component captured by the multi-factor error structure. Third, and most importantly, the existing PPML applications are confined exclusively to static gravity equations, such as Weidner and Zylkin (2019). A formal extension of the static PPML framework into a dynamic counterpart with a three-dimensional data structure goes beyond the scope of this paper, since it involves non-trivial practical hurdles. Specifically, the zero trade problem in the (lagged) dependent variable, introduction of a multi-factor error structure, as well as retention of parameter heterogeneity.[30] In other words, even PPML estimator would be biased due to disregarded lagged zero trade flows. Last but not least, it is not nested into the equation system (3.1)-(3.3) and, as a result, cannot shed light on the empirical support for our proposed theory compared to simpler alternatives (e.g., symmetric, homogeneous dynamic gravity model or static gravity).

[30] Though our dynamic extension preserves a multiplicative form of gravity equation, its independent variables, among others, include lagged trade value, which can be zero. On top of that, parameter heterogeneity as predicted by our theory and recently proved empirically relevant in, among others, Carrere et al. (2020), who use quantile regression, needs to be taken care of along with the time-varying unobserved multilateral resistance terms. Thus, we foresee extensions into nonlinear, including Poisson, settings along the lines of, e.g., Hacioglu-Hoke and Kapetanios (2021), to be particularly fruitful, paying attention to the incidental parameter problem, as emphasized by Weidner and Zylkin (2019).

## E.4 Other Estimators

**Table 4: Learning-By-Importing-Augmented Gravity Equation (Other Common Methodologies)**

| VARIABLES | AMG (1) $\text{FLOW}_{ij,t}$ | FE2 (2) $\text{FLOW}_{ij,t}$ | FE3 (3) $\text{FLOW}_{ij,t}$ | FE4 (4) $\text{FLOW}_{ij,t}$ | PPML (5) $\text{FLOW}_{ij,t}$ | PPML2 (6) $\text{FLOW}_{ij,t}$ | PPML3 (7) $\text{FLOW}_{ij,t}$ | PPML4 (8) $\text{FLOW}_{ij,t}$ |
|---|---|---|---|---|---|---|---|---|
| $\text{FLOW}_{ij,t-1}$ | 0.433*** | 0.908*** | 0.743*** | 0.682*** | 0.956*** | 0.965*** | 0.770*** | 0.743*** |
| | (0.00720) | (0.00322) | (0.00632) | (0.00732) | (0.00327) | (0.00264) | (0.00939) | (0.00971) |
| $\text{TB}_{j,t}$ | 0.839*** | | 0.418*** | | 0.166*** | | 0.320*** | |
| | (0.0980) | | (0.0334) | | (0.0404) | | (0.0390) | |
| $\text{GDP}_{i,t-1}$ | -0.197*** | | -0.0198** | | -0.00798 | | -0.0348*** | |
| | (0.0298) | | (0.00963) | | (0.0123) | | (0.0114) | |
| $\text{GDP}_{j,t-1}$ | -0.0234 | | -0.0529*** | | -0.00602 | | -0.0759*** | |
| | (0.0325) | | (0.00935) | | (0.0158) | | (0.0153) | |
| $\text{GDP}_{t-1}$ | 0.536 | | | | | | | |
| | (0.114) | | | | | | | |
| Time fixed effects | N | N | Y | N | Y | N | Y | N |
| Country fixed effects | - | N | - | - | Y | N | - | - |
| Time-varying country fixed effects | - | Y | N | Y | N | Y | N | Y |
| Pair fixed effects | Y | N | Y | Y | N | N | Y | Y |
| Unobservable Common Factors | Y | N | N | N | N | N | N | N |
| Constant | 8.443*** | 1.110*** | 3.987*** | 3.900*** | 0.764*** | 0.497*** | 4.536*** | 3.790*** |
| | (1.631) | (0.0406) | (0.194) | (0.0906) | (0.218) | (0.0392) | (0.289) | (0.144) |
| Observations | 70,579 | 70,604 | 70,602 | 70,602 | 71,365 | 71,312 | 71,364 | 71,311 |
| Number of pairs | 1,473 | 1,480 | 1,480 | 1,480 | 1,487 | 1,485 | 1,487 | 1,486 |
| R-squared | | 0.92 | 0.91 | 0.93 | 0.96 | 0.97 | 0.96 | 0.97 |

*Note*: Robust standard errors in parentheses; $\text{FLOW}_{ij,t}$ measured in levels and only non-zero lagged trade flows are retained in all PPML specifications.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

### E.5 Econometric Challenges of the Learning-by-Importing-Augmented Gravity Model

There are two major issues related to the empirical model specification in equation (2.18). First, the theory of learning-by-importing predicts a homogeneous trade imbalance coefficient equal to unity, whereas the other coefficients in the learning-by-importing-augmented gravity equation are heterogeneous. Second, unlike all other regressors, the multilateral trade imbalance $\Xi_{j,t}$ enters the learning-by-importing-augmented gravity equation contemporaneously, which opens up the possibility of model misspecification due to simultaneity. We propose two ways to address these issues. First, we expend $\Xi_{j,t}$ from the list of regressors and check the extent of the simultaneity bias. Second, we re-specify the learning-by-importing-augmented gravity equation by defining another variant of the size-adjusted trade flows $\tilde{A}_{ij,t} = A_{ij,t}/\Xi_{j,t} = X_{ij,t}Y_t/(Y_{i,t}C_{j,t})$, which effectively replaces $\Xi_{j,t}$ with $\Xi_{j,t-1}$ and retains full consistency with the theory of learning-by-importing. Such transformation delivers the following model, expressed in terms of pre-determined observables:

$$
\begin{aligned}
\ln \tilde{A}_{ij,t} = & \underbrace{\chi_{ij}(\eta-1)\ln \tilde{A}_{ij,t-1}}_{\text{size-adjusted bilateral trade flow persistence}} \\
& + \underbrace{\chi_{ij}(\eta-1)\ln(\Xi_{j,t-1})}_{\text{lagged destination multilateral trade imbalance}} \\
& - \underbrace{(1+\chi_{ij})(\eta-1)\ln d_{ij} + (\eta-1)\ln P_{j,t} + (\eta-1)\ln \Phi_{i,t} + \chi_{ij}(\eta-1)\ln \Phi_{i,t-1}}_{\text{bilateral and multilateral trade resistance}} \\
& + \underbrace{\chi_{ij}\eta \ln Y_{t-1} - \chi_{ij}\eta \ln Y_{i,t-1} + \chi_{ij}(\eta-1)\ln Y_{j,t-1}}_{\text{aggregate income}}. \qquad \text{(E.1)}
\end{aligned}
$$

Ultimately, Sections 3.3 and 4.2 show that neither of these issues are empirically important in terms of main conclusions and results.

As predicted by the re-specified gravity equation in (E.1), all parameters inherit bilaterally varying learning-by-importing. One of the major differences, compared to the baseline model in (2.18), is that the pre-determined trade imbalance becomes insignificant due to large variability across different trade pairs. We find that for any given destination country, the bilateral trade imbalance coefficients are remarkably heterogeneous, which are likely to depend on the structural differences between source and destination economies. This implies that bilateral trade reforms may exhibit consequences for international trade flows and the corresponding trade imbalance of countries not directly targeted by the reforms. In fact, using a static gravity equation with homogeneous coefficients, Cunat and Zymek (2019) find that bilateral imbalances depend on aggregate imbalances only if they are explained jointly with the multilateral resistance terms and the structural differences, such as production and spending patterns or trade wedges, which points to the heterogeneous influence of the trade-network-wide factors analyzed in this paper.

Last, and just as discussed in the main text, there are two main possibilities for model misspecification. One relates to the fact that both the bilateral trade flows (i.e., the dependent variable) and the multilateral trade balance (i.e., a regressor) are determined contemporaneously and simultaneously. The other goes back to the fact that the theory of learning-by-importing predicts a homogeneous trade imbalance coefficient equal to unity for each country pair. Tables 5 and 6 present two robustness checks in which we investigate the extent to which our results presented in Table 2 are affected by these issues of model misspecification. Our results show that the trade persistence coefficient remains virtually unchanged when we either exclude the multilateral trade imbalance from the vector of regressors (see "XFLOW" in Table 5) or when we re-specify the size-adjusted trade flows in terms of consumption expenditure in the destination country (see Table 6), both of which mitigate the aforementioned model misspecification issues. Our robustness checks demonstrate that the trade persistence coefficient following the CCEMG approach remains more than twice lower than the FE approach. We therefore conclude that in the context of the learning-by-importing-augmented gravity equation, these model misspecification issues are not empirically important. Our baseline results are further reinforced by the fact that there exists a long-standing trade literature that incorporates contemporaneous multilateral trade imbalance as a weakly exogenous regressor in static gravity equations (e.g., Davis and Weinstein (2002); Dekle et al. (2007, 2008)).

**Table 5: Coefficient Estimates (Learning-by-Importing-Augmented Gravity Equation Without Trade Imbalance)**

| | CCEMG | FE | MG | CCEP | HFE | AMG | CCEMGR |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| VARIABLES | $\text{FLOW}_{ij,t}$ | $\text{FLOW}_{ij,t}$ | $\text{FLOW}_{ij,t}$ | $\text{FLOW}_{ij,t}$ | $\text{FLOW}_{ij,t}$ | $\text{FLOW}_{ij,t}$ | $\text{FLOW}_{ij,t}$ |
| $\text{FLOW}_{ij,t-1}$ | 0.435*** | 0.908*** | 0.578*** | 0.443*** | 0.527*** | 0.475*** | 0.496*** |
| | (0.00789) | (0.00448) | (0.00692) | (0.0160) | (0.00684) | (0.00704) | (0.00736) |
| $\text{TB}_{j,t}$ | | | | | | | |
| $\text{GDP}_{i,t-1}$ | -0.269*** | -0.00182 | -0.182*** | -0.269*** | -0.231*** | -0.194*** | -0.208*** |
| | (0.0612) | (0.00754) | (0.0386) | (0.0325) | (0.0262) | (0.0277) | (0.0399) |
| $\text{GDP}_{j,t-1}$ | -0.102 | -0.0322*** | -0.141*** | -0.142*** | -0.0705*** | 0.0405 | -0.0150 |
| | (0.0725) | (0.00712) | (0.0187) | (0.0241) | (0.0265) | (0.0294) | (0.0434) |
| $\text{GDP}_{t-1}$ | 0.354** | -0.0270 | 0.502*** | | 0.338*** | -0.0513 | 0.487*** |
| | (0.181) | (0.0258) | (0.0257) | | (0.0884) | (0.102) | (0.154) |
| Time Fixed Effects | N | Y | N | N | N | N | N |
| Country/Country-Pair Fixed Effects | Y | Y | N | N | Y | Y | Y |
| Unobservable Common Factors | Y | N | N | Y | N | Y | Y |
| Constant | -3.434 | 1.860*** | | | | 8.950*** | -3.309* |
| | (2.669) | (0.352) | | | | (1.459) | (1.964) |
| Observations | 70,591 | 70,604 | 70,596 | 70,560 | 70,579 | 70,579 | 70,591 |
| Number of pairs | 1,475 | 1,480 | 1,476 | 1,471 | 1,473 | 1,473 | 1,475 |
| Adj. R-squared | 0.77 | 0.90 | 0.74 | 0.79 | 0.73 | | 0.76 |

*Note*: Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

**Table 6: Coefficient Estimates (Re-Specified Learning-by-Importing-Augmented Gravity Equation)**

| | CCEMG | FE | MG | CCEP | HFE | AMG | CCEMGR |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| VARIABLES | $XFLOW_{ij,t}$ | $XFLOW_{ij,t}$ | $XFLOW_{ij,t}$ | $XFLOW_{ij,t}$ | $XFLOW_{ij,t}$ | $XFLOW_{ij,t}$ | $XFLOW_{ij,t}$ |
| $XFLOW_{ij,t-1}$ | 0.360*** | 0.909*** | 0.556*** | 0.378*** | 0.503*** | 0.455*** | 0.481*** |
| | (0.00874) | (0.00445) | (0.00721) | (0.0160) | (0.00774) | (0.00760) | (0.00767) |
| $TB_{j,t-1}$ | -1.078 | -0.160*** | -0.0967 | -0.206*** | -0.110 | -0.247 | -0.269 |
| | (0.975) | (0.0323) | (0.132) | (0.0723) | (0.127) | (0.125) | (0.128) |
| $GDP_{i,t-1}$ | -0.521** | -0.00193 | -0.299*** | -0.288*** | -0.247*** | -0.224*** | -0.244*** |
| | (0.235) | (0.00759) | (0.0907) | (0.0316) | (0.0335) | (0.0354) | (0.0448) |
| $GDP_{j,t-1}$ | -0.191 | -0.0416*** | -0.179*** | -0.126*** | -0.0849** | -0.0692* | 0.0193 |
| | (0.211) | (0.00718) | (0.0231) | (0.0271) | (0.0367) | (0.0387) | (0.0515) |
| $GDP_{t-1}$ | 0.177 | -0.00346 | 0.628*** | | 0.366*** | -0.0820 | 0.441*** |
| | (0.399) | (0.0261) | (0.0562) | | (0.0924) | (0.104) | (0.152) |
| Time Fixed Effects | N | Y | N | N | N | N | N |
| Country/Country-Pair Fixed Effects | Y | Y | N | N | Y | Y | Y |
| Unobservable Common Factors | Y | N | N | Y | N | Y | Y |
| Constant | -5.206** | 1.307*** | | | | 6.969*** | -4.287** |
| | (2.326) | (0.353) | | | | (1.347) | (1.602) |
| Observations | 70,579 | 70,604 | 70,596 | 70,560 | 70,579 | 70,579 | 70,591 |
| Number of pairs | 1,475 | 1,480 | 1,476 | 1,471 | 1,473 | 1,473 | 1,475 |
| Adj. R-squared | 0.77 | 0.99 | 0.74 | 0.84 | 0.73 | | 0.76 |

*Notes:* Robust standard errors in parentheses; $XFLOW_{ij,t}$ measures bilateral trade flows adjusted for the size of world income, the size of income in the origin country, and the size of consumption expenditure, not income, in the destination country as discussed in Sections 2.5 and 3.3. The estimated model refers to the equation (E.1).

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

### E.6 Prediction "Horse Race"

**Table 7: Root Mean Square Error (Full Sample, Extensive List of Methods)**

| | Full Sample | "Bad Times" | | | | "Normal Times" | | | |
|---|---|---|---|---|---|---|---|---|---|
| Method | | $w=0$ | $w=1$ | $w=2$ | $w=3$ | $w=0$ | $w=1$ | $w=2$ | $w=3$ |
| CCEMG | 0.38 | **0.41** | **0.41** | **0.40** | **0.41** | **0.38** | **0.37** | **0.37** | **0.35** |
| MG | 0.45 | 0.52 | 0.51 | 0.49 | 0.49 | 0.45 | 0.44 | 0.43 | 0.41 |
| CCEP | 0.47 | 0.47 | 0.47 | 0.46 | 0.47 | 0.43 | 0.42 | 0.42 | 0.39 |
| HFE | 0.44 | 0.51 | 0.49 | 0.48 | 0.48 | 0.44 | 0.43 | 0.42 | 0.40 |
| AMG | 0.44 | 0.51 | 0.49 | 0.48 | 0.48 | 0.44 | 0.43 | 0.41 | 0.39 |
| CCEMGR | 0.42 | 0.45 | 0.46 | 0.45 | 0.45 | 0.42 | 0.41 | 0.40 | 0.38 |
| FE | 0.55 | 0.65 | 0.63 | 0.60 | 0.59 | 0.54 | 0.52 | 0.51 | 0.49 |
| FE2 | 0.52 | 0.61 | 0.58 | 0.56 | 0.55 | 0.50 | 0.49 | 0.47 | 0.46 |
| FE3 | 0.53 | 0.61 | 0.59 | 0.57 | 0.56 | 0.51 | 0.50 | 0.49 | 0.47 |
| FE4 | 0.49 | 0.56 | 0.54 | 0.52 | 0.51 | 0.47 | 0.46 | 0.45 | 0.43 |
| PPML | 0.54 | 0.67 | 0.64 | 0.61 | 0.61 | 0.55 | 0.53 | 0.52 | 0.50 |
| PPML2 | 0.44 | 0.65 | 0.61 | 0.60 | 0.59 | 0.53 | 0.52 | 0.50 | 0.48 |
| PPML3 | 0.44 | 0.65 | 0.62 | 0.59 | 0.59 | 0.55 | 0.53 | 0.53 | 0.52 |
| PPML4 | **0.34** | 0.62 | 0.59 | 0.57 | 0.57 | 0.53 | 0.52 | 0.52 | 0.51 |

*Note*: The Root Mean Square Errors (RMSE) are calculated using different methods of estimating the learning-by-importing gravity equation. The in-sample RMSEs are presented for the full sample, the "normal times", and the "bad times" in order to compare different model performance inside and outside of time periods characterized by common trade shocks. Consistent with Kose et al. (2020), the "bad times" represent the global recession years, namely 1975, 1982, 1991, and 2009, while the "normal times" are all of the remaining years in our sample that spans 1950-2014. The term $w = \{0, 1, 2, 3\}$ further indicates the length of the windows surrounding the recession years (i.e., number of years before and after global crises). The values in bold indicate the smallest RMSE.

**Table 8: Root Mean Square Error (Sub-Samples, Extensive List of Methods)**

| | Full Sample | "Bad Times" | | | | "Normal Times" | | | |
|---|---|---|---|---|---|---|---|---|---|
| Method | | $w=0$ | $w=1$ | $w=2$ | $w=3$ | $w=0$ | $w=1$ | $w=2$ | $w=3$ |
| CCEMG | 0.38 | - | **0.09** | **0.26** | **0.32** | 0.37 | 0.34 | **0.30** | **0.25** |
| MG | 0.45 | - | 0.37 | 0.42 | 0.45 | 0.44 | 0.42 | 0.40 | 0.36 |
| CCEP | 0.47 | - | - | 0.44 | 0.47 | 0.46 | 0.44 | 0.42 | 0.38 |
| HFE | 0.44 | - | 0.31 | 0.39 | 0.43 | 0.43 | 0.41 | 0.39 | 0.33 |
| AMG | 0.44 | - | 0.32 | 0.39 | 0.42 | 0.43 | 0.41 | 0.39 | 0.33 |
| CCEMGR | 0.42 | - | 0.23 | 0.34 | 0.38 | 0.41 | 0.38 | 0.35 | 0.30 |
| FE | 0.55 | 0.65 | 0.62 | 0.60 | 0.59 | 0.54 | 0.52 | 0.51 | 0.49 |
| FE2 | 0.52 | 0.62 | 0.60 | 0.58 | 0.57 | 0.52 | 0.50 | 0.49 | 0.47 |
| FE3 | 0.53 | 0.60 | 0.60 | 0.57 | 0.57 | 0.52 | 0.50 | 0.49 | 0.47 |
| FE4 | 0.49 | 0.57 | 0.56 | 0.54 | 0.53 | 0.49 | 0.47 | 0.46 | 0.44 |
| PPML | 0.54 | 0.53 | 0.70 | 0.64 | 0.60 | 0.54 | 0.49 | 0.49 | 0.49 |
| PPML2 | 0.44 | 0.50 | 0.60 | 0.54 | 0.51 | 0.44 | 0.40 | 0.40 | 0.39 |
| PPML3 | 0.44 | 0.37 | 0.44 | 0.44 | 0.43 | 0.43 | 0.40 | 0.39 | 0.47 |
| PPML4 | **0.34** | **0.35** | 0.35 | 0.35 | 0.34 | **0.34** | **0.32** | 0.31 | 0.44 |

*Note*: The Root Mean Square Errors (RMSE) are calculated using different methods of estimating the learning-by-importing gravity equation. The in-sample RMSEs are presented for the full sample, the "normal times", and the "bad times" in order to compare different model performance inside and outside of time periods characterized by global crises. Consistent with Kose et al. (2020), the "bad times" represent the global recession years, namely 1975, 1982, 1991, and 2009, while the "normal times" are all of the remaining years in our sample that spans 1950-2014. The term $w = \{0, 1, 2, 3\}$ further indicates the length of the windows surrounding the recession years (i.e., number of years before and after global crises). The values in bold indicate the smallest RMSE.

### E.7 Parameter Heterogeneity

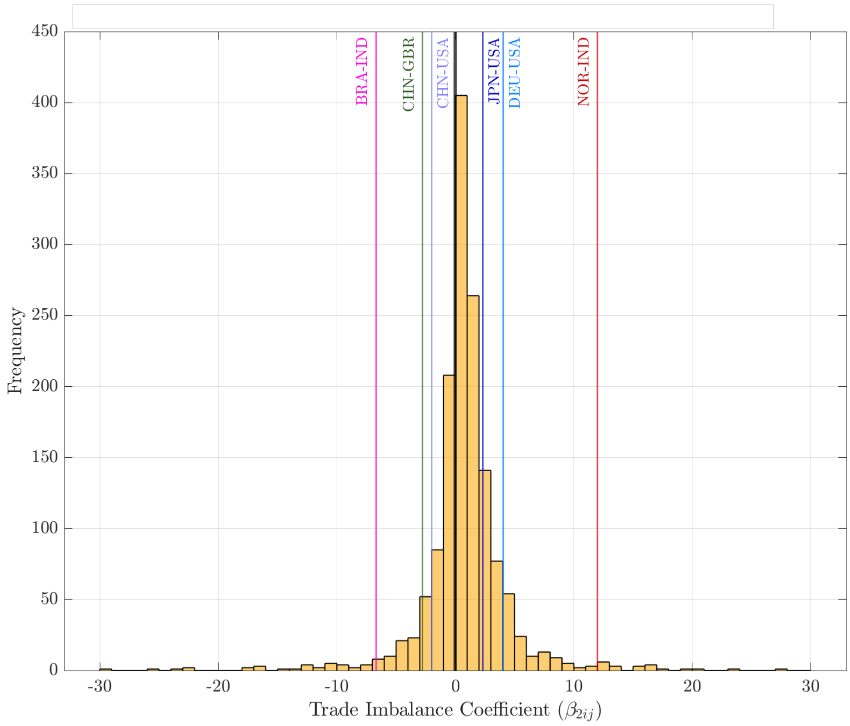



**Figure 6: Distribution of Trade Imbalance Coefficient Estimates**

*Note:* The figure presents CCEMG estimates of the trade imbalance coefficient ($\beta_{2ij}$) in the dynamic gravity model presented in equations (3.1)-(3.3). The trade imbalance coefficient estimates characterize 39 countries (i.e., up to 1482 country pairs) over the period of 1950-2014. Some country pair estimates are highlighted with according abbreviations.