What Explains Trade Persistence? A Theory of Habits in the Supply Chains*

Mariarosaria Comunale†
Bank of Lithuania, Vilnius University, and Australian National University

Justas Dainauskas‡
London School of Economics and Political Science

Povilas Lastauskas§
Bank of Lithuania and Vilnius University

Abstract

International trade flows are volatile, imbalanced, and fragmented across off-shored supply chains. Taking these empirical facts into account, we develop a theory of habits in the supply chains, 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 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 supply habit asymmetries 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; Habits; 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, Juan Passadore, and other 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. Comunale and Lastauskas gratefully acknowledge support from the European Social Fund (project No 09.3.3-LMT-K-712-01-123) under a grant agreement with the Research Council of Lithuania (LMTLT). Declarations of interest: none.

†Address at: `mariarosaria.comunale@gmail.com`, Applied Macroeconomic Research Unit (TMTS), Bank of Lithuania, Totorių 4, LT-01121, Vilnius, Lithuania; Faculty of Economics and Business Administration, Vilnius University, Saulėtekio al. 9, 2nd Building, LT-10222 Vilnius, Lithuania; Center for Applied Macroeconomic Analysis (CAMA), Australian National University, Canberra, Australia.

‡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@trinity.cantab.net`, CEFER, Bank of Lithuania, Totorių 4, LT-01121, Vilnius, Lithuania; Faculty of Economics and Business Administration, Vilnius University, Saulėtekio al. 9, 2nd Building, LT-10222 Vilnius, Lithuania.

1 Introduction

It takes time and resources to implement trade liberalization policies. That is why the patterns of "who trades with whom" are regionally-biased and slow to adjust (Eichengreen and Irwin (1998)). But at the same time, the value of "how much is traded" among those who partner up is surprisingly volatile, especially when countries are hit by common shocks. One such example is the "Great Trade Collapse" (GTC) of 2008-09 during which the world GDP shrank by 1%, while the value of global trade flows slumped by some 10% in a remarkably synchronized fashion across the world (Alessandria et al. (2010)). As empirical evidence shows, trade adjusting more abruptly than GDP stands out as a more general stylized fact. As things stand, the bulk of the modern trade literature predicts the value of international trade flows using the ubiquitous gravity equation. 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, trade flows adjust more rapidly and recover heterogeneously unlike predicted by the existing models of the homogeneous parameter dynamic gravity equation. We call this discrepancy the "trade persistence" puzzle. Failure to account for the drivers of trade dynamics leads to poor forecasts, invalid counterfactuals, and poor policy.

We contribute to the literature by putting forward a dynamic gravity model, outperforming existing alternatives empirically, thereby helping policymakers make better predictions and construct more reliable counterfactuals. Our empirical model is rooted in a tractable dynamic gravity equation, which we derive from a theory of habits in the supply chains. We acknowledge that alternative modeling assumptions can give rise to observationally similar dynamic gravity models, yet the literature misses heterogeneity, richer dynamics of multilateral resistance terms, and the role for aggregate trade imbalances, as suggested by our mechanism. Supply habits offer a simple reduced-form framework of the global trade network, where the production of final goods requires intermediate imports dispersed across space. At the aggregate level, countries develop habits of importing certain final and intermediate goods from other countries that are known for popular brands or their reputation, which may be related to shared values, history, institutions, or colonial ties (i.e., the broad definition of "distance"). At the firm level, supply habits, modeled as exogenous internal returns, capture inter-temporal contractual obligations on the globalized production belt line, which inhibit immediate assembling, disbanding, or swapping of the off-shore suppliers in response to shocks.

Compared to alternatives, our theory offers several advantages. First, supply habits predict heterogeneously autocorrelated (i.e., persistent) trade flows for each country *pair*. Second, cross-country habit asymmetry creates differences in consumption home-bias, which generates endogenous trade imbalances that explain additional variation in bilateral trade flows beyond the standard gravity measures, such as aggregate income and geographic distance. Third, supply habits enhance the geographic distance component of international trade costs, because it applies not only to goods that are "made here, sold there", but also to intermediate inputs that are "bought, sold, and bought again". Fourth, supply habits create differences in the "inward" and "outward" propensities to trade (i.e., multilateral trade resistance terms) whose contemporaneous and pre-determined values enter our version of the dynamic gravity equation. Common shocks are thus heterogeneously disruptive, because multilateral trade resistance is strongly correlated with country-specific trade imbalances as well as off-shore demand and supply. As a result, our model differentiates between different trade pairs, helping to produce a more nuanced picture of trade adjustment in the world of major disruptions, trade wars, and rewiring globalization.

Our estimates of the supply-habit-augmented dynamic gravity equation reveal a trade persistence coefficient of 0.35, which is 2.5 times lower than predicted by existing dynamic gravity models. We argue that this discrepancy occurs for two different reasons. First, the prevalent methods of estimating (dynamic) gravity equations do not appropriately account for the common shocks. For instance, the standard "country" fixed effects are time-invariant, while the "time" fixed effects are homogeneous for all country pairs. This implies that shocks originating from third countries are not fully reflected in either the source or the destination economies. Second, the inference is commonly drawn from the pooled gravity equation coefficient estimates, which ignores the fact that trade flows between some country pairs are significantly less persistent than others.

We show that the trade persistence puzzle in the existing literature is caused by (i) the pooled coefficient estimators; and (ii) the assumed homogeneity of the "time" fixed effects. Our main contribution on the empirical side is the theory-consistent estimation of a dynamic gravity equation with heterogeneous parameters across country pairs by exploiting a relatively large temporal dimension of the panel. We account for the heterogeneous responses to common shocks specific to each country pair by introducing a proxy to the multilateral trade resistance terms, which we call "unobservable common factors" that we define as the vector of cross-sectionally averaged regressand and regressors. We show that absent of the unobservable common 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 habits in the supply chains. Consistent with the theory, we map trade persistence to the global value chains and find empirical support for the proposed channel.

Despite considerable research efforts, not much is still known about the mechanism through which the value of trade flows adjust in response to either local, regional or worldwide trade shocks over time. The standard gravity equation due to Anderson (1979), Anderson and van Wincoop (2003), and Feenstra (2016) remains the workhorse framework for trade policy analysis in the context of permanent, unilateral, and exogenous trade shocks (also see Allen et al. (2020)). But it is static and silent about the transitional dynamics. Several others extend the gravity equation into a dynamic setting using the neo-classical theory of capital accumulation (e.g., Yotov and Olivero (2012); Alvarez (2017); and Anderson et al. (2020)). Their theory suggests that the trade persistence coefficient corresponds to the annual share of undepreciated capital stock. Yet the empirical estimates of the capital depreciation rate reveal that it is relatively homogeneous across countries and equals around 10% (see IMF (2015)). The neo-classical theory therefore predicts exceedingly high and homogeneous trade persistence, which is consistent with our pooled estimates absent of common factors, but inconsistent with the sharp decline and heterogeneous recovery of trade flows in face of common shocks observed over past six decades. Alternative models with richer micro-foundations cannot explain observed dynamics either. Eaton and Kortum (2002) derive observationally equivalent to Anderson and van Wincoop (2003) gravity model, driven by the supply-side considerations (heterogeneous productivity across countries, following a Fréchet distribution), generalized to the multi-sector economy by Costinot et al. (2012). Similarly, a heterogeneous firm model due to Melitz (2003), coupled with the parameterized productivity distribution (commonly used Pareto as in Chaney (2008)), also delivers multiplicative gravity model, though necessitating care when interpreting trade elasticity. Lastly, as demonstrated by Arkolakis et al. (2012), these equivalent representations generalize to a wider class of trade models with different micro-foundations, leaving us wonder about the neglected forces that are behind the empirically relevant dynamics of trade flows.

Our theory of habits in the supply chains delivers an intermediate degree of trade persistence relative to the static and the neo-classical gravity equations, such that consistent with the data, it predicts sharp and heterogeneous trade flow adjustments in response to local and global shocks. It also nests the static gravity equation à la Anderson and van Wincoop (2003) if we assume that all bilateral supply habits are infinitesimally weak. And if we assume that habits are strong and homogeneous for all country pairs, they cause less volatile and more persistent bilateral trade flows that are analogous to Anderson et al. (2020). Our theory of habits in the supply chains therefore delivers substantially richer dynamics of trade flows by admitting a more flexible domain for the trade persistence coefficient that is not tied to the empirical restrictions implied by the estimates of the capital depreciation rate. Admittedly, capital accumulation plays a role in the persistence of virtually all macroeconomic fundamentals. Supply habits are also not rooted in the first principles as strongly as the process of capital accumulation. However, they are a widely-established tool of characterizing dynamic properties of fundamentals in the macro-finance literature (e.g., Abel (1990); Campbell and Cochrane (1999); Ravn et al. (2006, 2007); and Herbst and Schorfheide (2016)). Some recent literature is also focused on micro-founding endogenous spatial firm networks (e.g., Chaney (2014); Arkolakis et al. (2021); Panigrahi (2021)), but it goes beyond the scope of this paper, as the theory of habits in the supply chains offers simple *prima facie* insights. However, we show that our theory predicts a direct mapping between the strength of supply habits and global value chain indicators that are in turn strongly related to the bilateral trade persistence heterogeneity. To preserve tractability and focus on the aggregate predictions, we abstract from a firm-level literature, which also addresses dynamics.[1]

This paper is also related to several other strands of international macroeconomics and the modern trade literature. First, the analysis of the bilateral trade persistence goes back to the seminal contribution of Eichengreen and Irwin (1998). We build on their work by developing an economic theory to support the dynamic nature of the gravity equation. Second, a number of studies

[1] The leading dynamic trade models with sunk costs à la Krugman (1980) and Melitz (2003), or market penetration costs à la Arkolakis (2010) remain widely used to conduct counterfactual analysis, to explore transitional dynamics, and to study the relationship between trade and growth. For the model in the former category, refer to Sampson (2015), whereas for the latter see Arkolakis (2016); recently, Morales et al. (2019) modeled market access costs by considering similarity to previous destination markets, producing what is coined as "extended gravity". Refer to Alessandria et al. (2021) for the recent review on firm dynamics and trade. Since we are primarily motivated by the empirical gravity literature using aggregate data and abstract from these structural, more computationally challenging models, we put forward a tractable model, which accounts for the empirical stylized facts on trade dynamics and can be estimated at the trade flows pair level.

explore the persistence of trade costs during the period of hyper-globalization (e.g., Anderson and van Wincoop (2004); Disdier and Head (2008); Zwinkels and Beugelsdijk (2010); Head and Mayer (2014)). The persistence of trade costs is directly related to the dynamic structure of the multilateral trade resistance terms in our supply-habit-augmented gravity equation and aligns with the stylized facts established by Baldwin and Taglioni (2006). Third, a number of prominent studies do not recognize the importance of unobservable and dynamic common factors in gravity equations (e.g., Egger (2000); Micco et al. (2003); and Helpman et al. (2008)). But Serlenga and Shin (2007) were the first to explore the role of contemporaneous unobservable common shocks in the context of a static gravity equation. Motivated by the theory of habits in the supply chains, we extend their framework by incorporating both the contemporaneous as well as pre-determined unobservable common factors.[2] Fourth, a number of studies examine the welfare consequences of mitigating exogenously pre-existing trade imbalances (e.g., Davis and Weinstein (2002); Dekle et al. (2007, 2008)). Other studies emphasize the importance of trade imbalances to understand transitional in response to common shocks (i.e., IMF (2019a); Beirne et al. (2020); Dix-Carneiro et al. (2021)). We also show that trade imbalances are an important determinant of *bilateral* trade flows, but unlike the existing gravity literature, the theory of habits in the supply chains generates trade imbalances endogenously. Fifth, our theory complements the recent literature on the time-varying and/or heterogeneous trade elasticity driven by preferences in the demand-side of the economy (e.g., Fieler (2011), Novy (2013), Carrere et al. (2020), and Boehm et al. (2020)). Our approach does not separate the short- and the long-run run effects in trade elasticities. But contrary to the existing literature, the theory of habits in the supply chains generates a cross-sectionally heterogeneous trade elasticity that originates from the supply-side structural differences across countries. We relegate the discussion about the welfare consequences of our theory for future research.

The rest of this paper is organized as follows. In Section 2, we describe the key stylized empirical facts about global trade flows that are not well accounted for by currently popular dynamic gravity models. More precisely, a relatively low persistence in trade and the substantial heterogeneity in reactions to shocks motivate a general equilibrium theory of habits in the supply chains, covered in the first part of Section 3. The second part of Section 3 derives the supply-habit-augmented dynamic gravity equation. Section 4 describes the data and the methodology of all panel regression techniques applied in this paper. We also discuss different choices related to the empirical modeling of the unobservable common factors. We then present the coefficient estimates of all dynamic gravity equation specifications and compare them with those in the existing literature. Section 5 presents prediction performance of our model in comparison of the nested alternatives, also the extent of the cross-country parameter heterogeneity and breaks down its source and relationship to the proposed theory of habits in the supply chains. An application of the model in constructing a stylized impulse response function and counterfactual trade flows is featured in Section 6. Finally, Section 7 summarizes and concludes, whereas Online Appendix collects all technical details and other supporting material.

[2] One of the most prominent common shocks that received substantial coverage in the trade literature is the GTC after the global financial crisis (see Alessandria et al. (2010); Bems et al. (2010); Altomonte et al. (2012); Antonakakis (2012); Levchenko et al. (2010); Eaton et al. (2016); Novy and Taylor (2020), among others). Our theory and empirical model are more general in that we focus on the role of intermediates trade through production externality over the period of more than six decades rather than a single episode.

2 Data Facts

We focus on the interaction between the value of trade flows and the global business cycle. We start by depicting dynamics of output and trade growth rates over more than five decades in Figure 1. According to the classification of Kose et al. (2020), there were four global recessions over the past seven decades, not counting the most recent pandemic-induced global economic disruption. All global recessions are preceded by a substantial adjustment in trade value, well exceeding changes in output. In other words, Figure 1 visualizes not only a higher volatility of trade as compared to output but also substantially larger reactions to global recessions. This stylized fact raises a question about other mechanisms than the ones that give rise to output persistence (e.g., investment dynamics) to explain trade dynamics. Notably, the stylized fact of sharp and substantially larger decline in trade volume compared to output is true for all global shock episodes, extending well beyond the global financial crisis (and the ensuing GTC). Therefore, these empirical patterns need to be incorporated into a tractable quantitative model that features plausible dynamics, explains data and can form a basis for counterfactual predictions and policy questions.

Before embarking on such a task, we zoom into the experience of different countries that were hidden in the global variables in Figure 1. 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 2 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 2). 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. Last, as already documented for the past six decades, 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. The remaining parts of this paper assimilate these empirical stylized facts into the dynamic extension of the gravity equation.

3 Theoretical Model

We lay down key ingredients of the theoretical model.[3] The world economy evolves over discrete time $t = 0, 1, 2, ...$ and comprises of a finite number of countries indexed by $i, j \in n = \{1, 2, ..., N\}$. Each country is populated by two types of interacting agents: (i) consumers; and (ii) producers. The producers in each country operate in two different sectors: (i) wholesale; and (ii) distribution. The

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

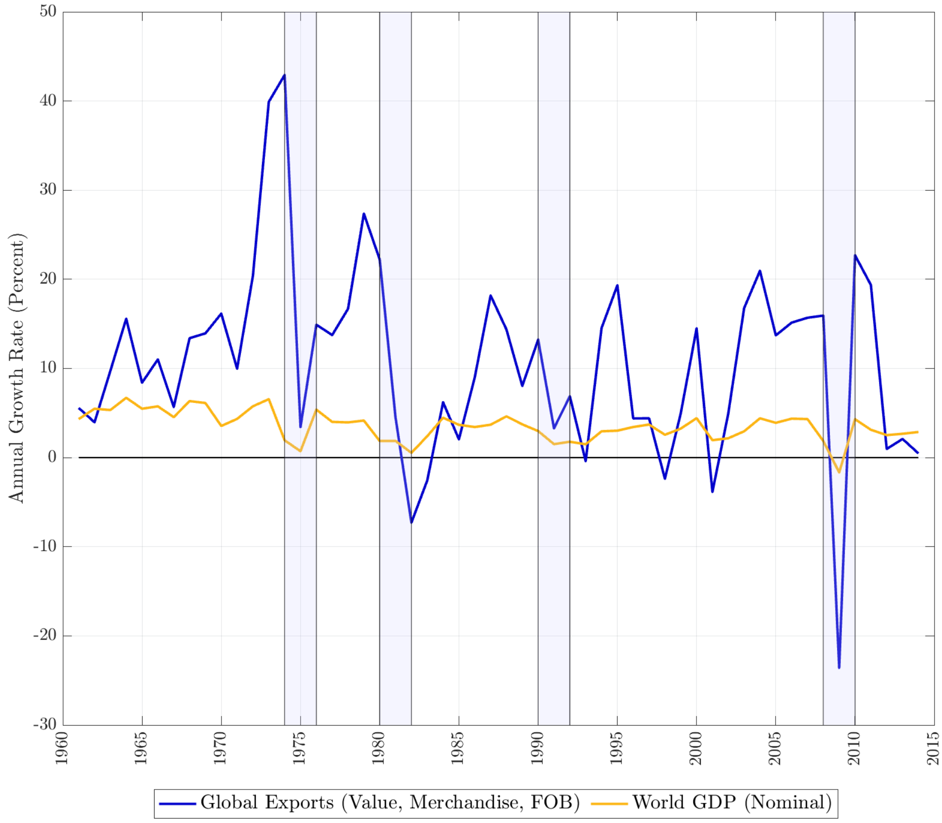

Figure 1: Dynamics of Global Trade and GDP Growth Rates

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).

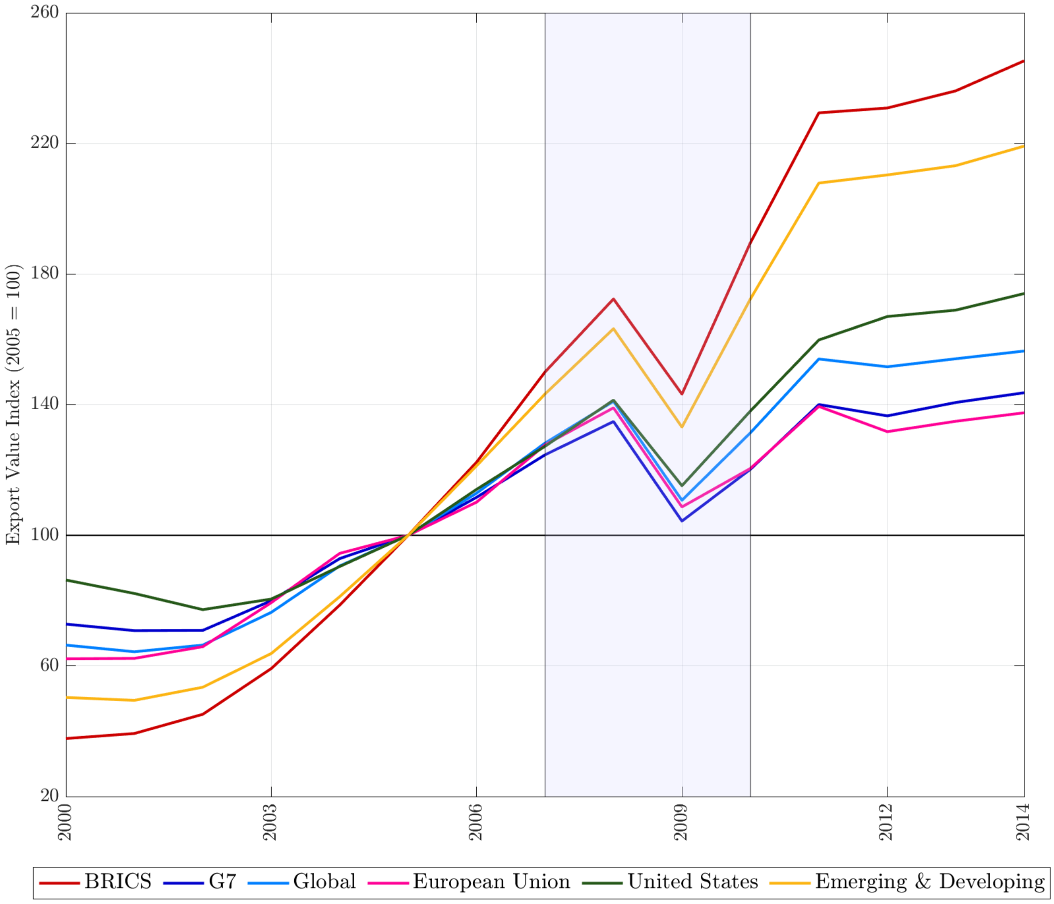

Figure 2: Trade Flows in Major Country Groups

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).

wholesale sector is populated by a unit mass of firms indexed by $\omega \in [0,1]$. All economies are open to trade wholesale varieties with one another, but bilateral trade flows are dampened by Samuelson's 'iceberg costs'.[4] The distributor merges the imported and domestically-produced wholesale varieties into a composite good. The consumers can only purchase the composite good and supply an inelastic fraction of their time endowment as labor to the wholesale firms. There is no entry or exit, but the production technology of the distributor is subject to a "learning-by-importing" type of technology in which pre-determined exports are positively related to current exports. We call this "habits" in the supply chains in order to distinguish the trade persistence implied by "learning-by-importing" from the potential productivity gains. In equilibrium, multilateral trade imbalances arise when different country pairs are subject to asymmetrical home-bias, which arise when there cross-country differences in habits.

3.1 Supply Side

Wholesale varieties are imperfectly substitutable and produced using linear labor-intensive technology: $m_{ij,t}(\omega) = z_{i,t}h_{ij,t}(\omega)$, where $i,j \in n$. Aggregate labor productivity in country i denoted as $z_{i,t}$ is exogenously given and the only source of uncertainty in this model. The hours of labor spent by workers domiciled in source country i to produce varieties that are sold in destination j are denoted as $h_{ij,t}(\omega)$. Delivering one unit of the wholesale variety from the source country i to the destination j costs $d_{ij} - 1 > 0$ proportion of the unit costs of production (i.e., iceberg cost).[5]

3.1.1 Technology

The distributor aggregates the wholesale varieties into an infinitely-divisible composite good according to the following constant elasticity of substitution (CES) production technology:

$$x_{ij,t} = \left[\int_0^1 \left(m_{ij,t}(\omega) x_{ij,t-1}^{\chi_{ij}} \right)^{1-1/\eta} d\omega \right]^{1/(1-1/\eta)}, \tag{3.1}$$

where $\eta > 1$ is the elasticity of substitution, $\chi_{ij} > 0$ denotes the habit intensity, and $x_{ij,t-1}$ is the stock of supply or production habit.[6] In this framework, the production of the final goods $x_{ij,t}$ requires intermediate imports $m_{ij,t}(\omega)$ dispersed across space. When the distributor in destination j develops a habit of sourcing intermediate imports from country i, the demand for

[4] Iceberg costs is a catch-all time-invariant bilateral trade resistance term, which subsumes both tariff and non-tariff barriers to trade, including the exogenously determined geographic distance.

[5] The standard triangular equation holds at all times, namely, $d_{ij} \leq d_{i\iota}d_{\iota j}$ for all $i, j, \iota \in n$, such that direct shipment of merchandise is the least expensive route.

[6] The existing literature provides several ways of modeling habits, sometimes referred to as "catching up with the Joneses". In macro-finance, the stock of habit enters the lifetime utility of the consumer as a function of past consumption, which introduces a richer autocorrelation structure and improves the model-implied fit of the observed data (Abel (1990); Campbell and Cochrane (1999); Herbst and Schorfheide (2016)). In closed and open economy macroeconomics, the stock of habit enters the CES preferences as a function of past consumption of individual varieties (i.e., "deep habits"), which generates counter-cyclical mark-up adjustments (Ravn et al. (2006, 2007)). In this paper, the stock of habit instead enters CES production technology, which is dual to CES preferences, but the stock of habit is aggregate and independent of individual varieties of intermediate imports. An interesting extension that is not considered in this paper is to incorporate "deep habits" when firms are subject to idiosyncratic productivity shocks à la Melitz (2003). However, the aggregate stock of habit is sufficient to generate autocorrelated bilateral trade flows specific to each country pair that is of key interest in this paper.

intermediate imports and therefore the production of final goods becomes persistent, such that $\partial \ln x_{ij,t} / \partial \ln x_{ij,t-1} = \chi_{ij}$. Trade flow adjustments are therefore gradually decaying, permanent, or explosive in response to shocks when $\chi_{ij} \in (0,1)$, $\chi_{ij} = 1$, or $\chi_{ij} > 1$, respectively. And the determinants of trade flows are static, as per usual, when production habits are infinitesimally weak, such that $\chi_{ij} \to 0$.[7]

Technology or supply habits put forward in this paper require a further elaboration. First, a multiplicative form of past experience of sourcing from a particular pair delivers a familiar structural gravity model, extended to a dynamic setting, and nesting a static alternative. Second, though preference habits are well documented for final (consumption) goods, technology adaptation to a particular supplier of inputs and learning from the past experience has been less explored in the gravity literature despite existing empirical and theoretical evidence. For instance, Grossman and Helpman (1995) cover early literature on learning-by-doing as an important source of technical change where trade plays a key role. The idea is that accumulated knowledge to manufacture the product makes firms more productive. In other words, the initial knowledge makes history relevant for the trade patterns. The importance of intermediate goods to produce productivity gains and economies due to increasing specialization can be traced back in Ethier (1982). A more recent revival of the learning-by-doing technology and trade is sometimes referred to as the "learning-by-importing" hypothesis or embodied technology in imports, and it has received substantial empirical support in Acharya and Keller (2009); Amiti and Konings (2007); Elliott et al. (2016); Halpern et al. (2015); Zhang (2017), among many others. In a review article, Keller (2004) concludes that "importing is associated with technology spillovers."

The production function (3.1) provides a concise and tractable way to capture learning from past trade between countries i and j. In effect, the term $x_{ij,t-1}^{\chi_{ij}}$ can be seen as a technology parameter that affects the relative demand for imported imports $m_{ij,t}$, parameterized as the (indexed by χ_{ij}) past production level, embodying accumulated knowledge to combine inputs. The larger the past aggregate production level, the better is a firm at using inputs. We will remain agnostic as to which forces, contractual, institutional, customization or learning and technological, play a more important role in giving rise to the supply habits augmented production technology (3.1). We will later show that all these explanations can be mapped into a measure of participation in global value chains (or foreign value-added in domestic output), thereby enabling us to empirically evaluate parameter χ_{ij} (that is why our preferred label as "habits in the supply chains"). We also leave endogenous mechanisms that govern production habits for future research.

3.1.2 Wholesale Varieties

Let $P_{ij,t}(\omega)$ denote the price of variety ω that is produced in economy i and sold in destination j at time t. The distributor chooses the amount of wholesale varieties to purchase $m_{ij,t}(\omega)$ by minimizing the total expenditure on intermediate imports $\tilde{P}_{ij,t} x_{ij,t} - \int_0^1 P_{ij,t}(\omega) m_{ij,t}(\omega) d\omega$ subject to the augmented CES preferences in equation (3.1). The first-order condition with respect to

[7] We consider the multiplicative habit specification, 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 habits, a more appropriate analytical benchmark are not the CES, but the translog preferences (see Novy (2013)).

$m_{ij,t}(\omega)$ gives rise to the following optimal demand schedule for wholesale varieties:

$$m_{ij,t}(\omega) = x_{ij,t} x_{ij,t-1}^{\chi_{ij}(\eta-1)} \left[\frac{P_{ij,t}(\omega)}{\tilde{P}_{ij,t}} \right]^{-\eta}. \tag{3.2}$$

The demand for intermediate imports is increasing in the contemporaneous stock of the composite good $x_{ij,t}$ and decreasing in the relative price of that variety $P_{ij,t}(\omega)/\tilde{P}_{ij,t}$, but since $\chi_{ij} > 0$, it is increasing in the stock purchased in the previous period $x_{ij,t-1}$. The price level $\tilde{P}_{ij,t}$ is such that the distributor in destination j breaks-even.

The wholesalers are monopolistically-competitive and adopt the strategy of "pricing-to-habits" (Ravn et al. (2007)).[8] Specifically, they recognize the demand schedule of the distributors in each destination when setting the optimal price for their variety, which is derived by maximizing the nominal profits $P_{ij,t}(\omega)m_{ij,t}(\omega) - d_{ij}MC_{i,t}m_{ij,t}(\omega)$, subject to the demand schedule in equation (3.2). The first-order condition gives rise to the standard expression for a fixed price-cost margin:

$$P_{ij,t}(\omega) = \left(\frac{\eta}{\eta - 1} \right) d_{ij} MC_{i,t}, \tag{3.3}$$

where $MC_{i,t}$ are the unit costs of producing the wholesale variety and $\eta/(\eta - 1) > 1$ is a constant price mark-up. Given the absence of idiosyncratic shocks to labor productivity in the wholesale production technology, the unit costs for all wholesaler firms in equilibrium are homogeneous and equivalent to the nominal hourly wage rate in the source country $W_{i,t}$, normalized by the aggregate labor productivity $z_{i,t}$, such that $MC_{i,t} = W_{i,t}/z_{i,t}$.

3.2 Demand Side

Each destination $j \in n$ is populated by a representative consumer characterized by CES preferences over consumption of composite goods originating from each source country:

$$c_{j,t} = \left[\sum_{i=1}^{N} x_{ij,t}^{1-1/\eta} \right]^{1/(1-1/\eta)}. \tag{3.4}$$

The representative consumer chooses the amount of composite goods to purchase from any trade partner by minimizing the total expenditure $P_{j,t}c_{j,t} - \sum_{i=1}^{N} \tilde{P}_{ij,t}x_{ij,t}$, subject to the CES preferences, which gives rise to the optimal demand schedule for composite goods:

$$x_{ij,t} = c_{j,t} \left(\frac{\tilde{P}_{ij,t}}{P_{j,t}} \right)^{-\eta}, \tag{3.5}$$

where $P_{j,t}$ denotes the aggregate consumer price index. The demand for composite goods is increasing in the aggregate consumption of the destination country $c_{j,t}$, but decreasing in the relative price of composite goods originating from each source country $\tilde{P}_{ij,t}/P_{j,t}$.

Suppose the representative consumer derives utility from the consumption of an infinitely-

[8] When the stock of habit $x_{ij,t-1}$ is independent of individual varieties, as is our proposed setting, "pricing-to-habits" implies local currency pricing (LCP). But the choice of invoicing currency does not play an important role in gravity models due to the low frequency of bilateral trade data. We therefore abstract from the choice of invoicing currency in modelling price setting.

divisible basket of composite goods $c_{j,t}$, while the aggregate hours of labor are supplied inelastically. The lifetime utility of the representative consumer is therefore given by

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \ln\left(c_{j,t}\right), \tag{3.6}$$

where $\mathbb{E}_0$ is the rational expectations operator and $\beta \in (0,1)$ stands for the time preference parameter.[9] The representative consumer is subject to an indefinite sequence of budget constraints:

$$c_{j,t} + \mathbb{E}_t[\zeta_{j,t,t+1} b_{j,t+1}] \leq b_{j,t} + w_{j,t} h_j + \varpi_{j,t}, \tag{3.7}$$

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, h_j are the inelastically supplied hours of labor relative to the total endowment of time, and $\varpi_{j,t}$ is the real aggregate profit dividend. The profit dividend forms part of the representative household income, since they own the wholesale firms that are domiciled in their country.

When the national stock of bonds is in zero net supply (i.e., $b_{j,t+1} = b_{j,t} = 0$), the aggregate consumption in each source country equals to the wage bill plus the profit dividends (i.e., the "hand-to-mouth" budget constraint implies $c_{j,t} = w_{j,t} h_j + \varpi_{j,t}$). However, if the international financial market structure is complete and aggregate stock of bonds is allowed to be positive (negative) in equilibrium, such that the country is able to lend (borrow), then the aggregate consumption would increase (decrease) in the contemporaneous stock of bonds $b_{j,t}$, but decrease (increase) in the stock of bonds held until maturity in the next period $b_{j,t+1}$ (see Obstfeld and Rogoff (1996)). As a consequence, international borrowing and lending generates a widely-observed consumption smoothing pattern and ultimately allows for short-run and long-run multilateral trade imbalances to arise. Whether or not the multilateral trade imbalance is saddle-path stable depends on the time preference parameter β. If β is bounded between zero and unity, as is usually the case, then it can be shown that the transversality condition holds, since it implies that the long-run real rate of interest is strictly non-negative. And if so, then the long-run multilateral trade imbalance is always finite and, in the long-run, it is equivalent to the present discounted value of the trade imbalance, which is constant and bounded in the long-run steady state (see Appendix A.5). Any country $j \in n$ can therefore sustain a current account deficit perpetually without violating the transversality condition, but only if it is a sufficiently large net creditor to begin with.[10]

The representative consumer chooses aggregate consumption $c_{j,t}$ and aggregate stock of bonds $b_{j,t}$ by maximizing their lifetime utility (3.6), subject to an indefinite sequence of budget constraints (3.7), taking the aggregate profit dividend $\varpi_{j,t}$, the real price of one-period bonds $\zeta_{j,t,t+1}$, and the real wage bill $w_{j,t} h_j$ as given. The first-order conditions give rise to the standard Euler equation

[9] For simplicity, the international financial markets are complete. We use compact notation that suppresses the state of nature, which is a conventional approach in the 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.

[10] In the context of complete international financial markets, Dix-Carneiro et al. (2021) also consider endogenous trade imbalances as an important vehicle for driving trade adjustment dynamics.

and the perfect consumption risk sharing relationship, respectively:

$$1 = \mathbb{E}_t \left[\frac{c_{j,t}}{\zeta_{j,t,t+1} c_{j,t+1}} \right], \tag{3.8}$$

$$\zeta_{j,0,t} = \zeta_{i,0,t}, \tag{3.9}$$

The Euler equation (3.8) implies that countries smooth consumption expenditure over time relative to the contemporaneous stream of income by saving or borrowing against the expected future income. The complete financial market structure and additively separable preferences give rise to the standard consumption risk-sharing relationship à la Backus and Smith (1993) as shown in equation (3.9).

3.3 General Equilibrium

General equilibrium is a set of dynamic processes characterizing a unique set of state and control variables given by $\{b_{i,t+1}, \zeta_{i,t,t+1}, x_{ij,t}, m_{ij,t}, p_{ij,t}, \tilde{p}_{ij,t}, w_{i,t}, \varpi_{i,t}, z_{i,t}, c_{i,t}, \pi_{ij,t}, \xi_{i,t}, \Xi_{i,t}\}_{t=0}^{\infty}$ for all $i, j \in n$ that are consistent with the utility-maximizing behavior of the representative household and the profit-maximizing behavior of the representative firm. The dynamic processes are conditional on pre-determined variables $\{b_{i,t}, , z_{i,t-1}, x_{ij,t-1}\}_{t=0}^{\infty}$, labor productivity shocks $\{\epsilon_{i,t}\}_{t=0}^{\infty}$, and fixed parameters $d_{ij}, \chi_{ij}, \rho_i, \sigma_i, \mu_i, h_i, \eta, \beta$. Due to the standard price level indeterminacy issues, the general equilibrium is defined in terms of the relative prices $p_{ij,t} = P_{ij,t}/P_{j,t}$ and $\tilde{p}_{ij,t} = \tilde{P}_{ij,t}/P_{j,t}$.

The relative prices are independent of ω in general equilibrium, because all wholesale firms are subject to country-specific uncertainty associated with aggregate labor productivity over time, but they do not face any idiosyncratic risk. As a consequence, the general equilibrium is symmetric, such that wholesale prices are homogeneous across firms for any given source country and they are set as a constant mark-up over the unit costs (see equation (3.3)). It implies not only $P_{ij,t}(\omega) = P_{ij,t}$ and $m_{ij,t}(\omega) = m_{ij,t}$ for all $\omega \in [0, 1]$, but also that the supplied hours to produce varieties for all destination markets are constant. The equilibrium supply of intermediate goods for each wholesale variety is thus perfectly price inelastic in each source country. But as long as $\chi_{ij} > 0$, the break-even price index of the distributor $\tilde{P}_{ij,t}$ does not correspond to the aggregate wholesale price index $P_{ij,t}$, because it is influenced by the stock of habit:

$$\tilde{P}_{ij,t} = x_{ij,t-1}^{-\chi_{ij}} P_{ij,t}. \tag{3.10}$$

Similarly, the duality problem gives rise to the aggregate consumer price index as a function of the break-even price indices from each source country:

$$P_{j,t} = \left[\sum_{i=1}^{N} \tilde{P}_{ij,t}^{1-\eta} \right]^{1/(1-\eta)}. \tag{3.11}$$

Observe that $\lim_{\chi_{ij} \to 0} \tilde{P}_{ij,t} = P_{ij,t}$, whereas more generally $P_{ij,t}/\tilde{P}_{ij,t}$ is not equal to unity, such that the aggregate stock of intermediate goods in equilibrium is increasing (decreasing) in the current (past) stock of the composite good:

$$m_{ij,t} = x_{ij,t}x_{ij,t-1}^{-\chi_{ij}}. \tag{3.12}$$

The multilateral trade flows can therefore be expressed as a geometrically decaying function of the demand for intermediate goods in the past:

$$x_{ij,t} = m_{ij,t}x_{ij,t-1}^{\chi_{ij}} = \prod_{s=0}^{\infty} m_{ij,t-s}^{\chi_{ij}^{s}}, \tag{3.13}$$

thereby introducing the central mechanism through which dynamic gravity effects are sustained even in the symmetric general equilibrium.

Let $X_{ij,t} = P_{ij,t}x_{ij,t}$, $C_{j,t} = P_{j,t}c_{j,t}$, and $Y_{j,t} = \bar{P}_{j,t}y_{j,t}$ denote the aggregate nominal value of trade flows, consumption, and output, respectively. Notice that absent of government expenditure and capital formation, the deflator pertaining to the gross domestic product, namely $\bar{P}_{j,t}$, is identical to the consumer price index derived from the duality problem only if trade is balanced at all times. The aggregate income of each country therefore amounts to its global sales of composite goods to each destination in each time period:

$$Y_{j,t} = \sum_{i=1}^{N} X_{ji,t}. \tag{3.14}$$

By contrast, the aggregate consumption in each economy is equal to its global expenditure on composite goods at any given time period:

$$C_{j,t} = \sum_{i=1}^{N} X_{ij,t}. \tag{3.15}$$

In this model, the difference between aggregate income and aggregate consumption defines the multilateral trade balance:

$$NX_{j,t} = Y_{j,t} - C_{j,t} = \sum_{i=1}^{N} X_{ji,t} - \sum_{i=1}^{N} X_{ij,t}, \tag{3.16}$$

such that aggregate consumption is proportional to aggregate income:

$$C_{j,t} = Y_{j,t}\Xi_{j,t}, \tag{3.17}$$

where $\Xi_{j,t} = 1/(1 + \sum_{i=1}^{N} \pi_{ji,t} - \sum_{i=1}^{N} \pi_{ij,t}) = 1/(1 + \xi_{j,t})$ captures multilateral trade imbalance relative to the total consumption expenditure and $\pi_{ij,t} = X_{ij,t}/C_{j,t}$ is defined as the import penetration ratio. The relative trade imbalance term is strictly non-negative $\Xi_{j,t} > 0$ at each time period, since $\xi_{j,t} = NX_{j,t}/C_{j,t} \in (-1, \infty)$ is the ratio of net exports to aggregate consumption expenditure.

In the related gravity literature, $\Xi_{j,t}$ is traditionally assumed to be equal to unity or simply exogenously given. But instead of implicitly ruling out the possibility of multilateral trade imbalances or assuming that they prevail *ad hoc*, our model provides a theoretical justification for their existence in both the short-run and the long-run. The model remains consistent with the accounting identity in which the import penetration ratios associated with each source country sum up to unity in the destination country, such that $\sum_{i=1}^{N} \pi_{ij,t} = 1$. Hence, an increase in the consumption of foreign goods in relative terms (i.e., a rise in $\pi_{ij,t}$ for any $i \in n$) implies a decline in the consumption of domestic goods in relative terms (i.e., a fall in $\pi_{jj,t}$). Consistent with Obstfeld and Rogoff (1996), international trade in this model is one of the forces through which local or country-specific labor productivity shocks are deflected in the short-run by saving or borrowing against the permanent income. While a negative labor productivity shock at home leads to a decline in aggregate consumption in the special case of autarky, with compelete markets, the loss in productivity is substituted with foreign production by running a transitory trade deficit.

When the influence of labor productivity shocks fades away in the long-run, and the economy reverts back to the steady state, the term $\Xi_{j,t}$ would only be equal to unity if there were no structural heterogeneities across countries (i.e., symmetric steady state). And since consumers supply labor to the wholesale firms inelastically, there can only be two dimensions along which structural heterogeneities distort the import penetration ratios across countries: (i) tariff and non-tariff barriers to trade subsumed within the iceberg costs d_{ij}; and (ii) technological import dependence encapsulated by the stock of habits in the production technology $x_{ij,t-1}^{\chi_{ij}}$, the differences of which are driven by the parameter characterizing the habits of the distributor χ_{ij}. If economies differ in either of these two dimensions, then trade is not balanced in the long-run and the term $\Xi_{j,t}$ is endogenously and indefinitely shifted away from unity.

Formally, when all exporters adopt "pricing-to-habits" strategies, import and export prices are proportional, such that $P_{ij,t} = (\eta/(\eta-1))d_{ij}MC_{i,t} = d_{ij}P_{ii,t}$. In turn, the break-even price index is $\tilde{P}_{ij,t} = d_{ij}P_{ii,t}x_{ij,t-1}^{\chi_{ij}}$, such that the import penetration ratio is given by $\pi_{ij,t} = (\tilde{P}_{ij,t}/P_{j,t})^{1-\eta} = (d_{ij}P_{ii,t}x_{ij,t-1}^{\chi_{ij}}/P_{j,t})^{1-\eta}$. Although the numerator $P_{j,t}$ itself is also influenced by both d_{ij} and $x_{ij,t-1}^{\chi_{ij}}$, the consumer price index is a function of trade costs and habits across all trade partner countries, while the frictions pertaining to the break-even price index in the denominator $\tilde{P}_{ij,t}$ are bilateral. Consequently, if either habits or trade costs are asymmetric across countries, such that $\chi_{ij} \neq \chi_{ji}$ and/or $d_{ij} \neq d_{ji}$, then $\Xi_{j,t} \neq 1$ in the long-run. And as discussed above, long-run trade imbalances are sustained by a corresponding imbalance in the financial account, which in this model corresponds to a permanent and sustainable inflow or outflow of bonds.

3.4 Supply-Habit-Augmented Gravity Equation

We have thus far established a well-defined and internally-consistent general equilibrium in a model of the world economy characterized by supply habits in the wholesale distribution network. But the general equilibrium itself does not provide an identifiable link between the empirical estimates of international trade flow persistence and the intensity of habits in our model. In the absence of habits in the supply chains, Anderson and van Wincoop (2003) establish the ubiquitous "gravity equation" approach of taking static trade models with homothetic preference aggregators to the data.[11] The static gravity equation links the bilateral trade flows to aggregate income in the source

[11]In principle, equation (3.13) alone could naively be used to determine the level of bilateral trade flow persistence and to identify parameter χ_{ij} for any country pair. But in a general equilibrium environment, the naive approach does not account for the fact that the stock of intermediate trade flows $m_{ij,t}$ and the stock of composite goods $x_{ij,t}$ are determined simultaneously, which causes the estimates of trade persistence to be biased. We circumvent the problem of simultaneity by replacing the intermediate trade flows entering equation (3.13) with a function of the multilateral trade resistance, which extends the static Anderson and van Wincoop (2003) approach into a dynamic framework.

and destination countries as well as the unobservable bilateral and multilateral trade resistance. This section of the paper shows that habits extend the static gravity equation into a dynamic counterpart in which bilateral trade flows depend on: (i) the lagged bilateral trade flows capturing their heterogeneous persistence across different country pairs; (ii) multilateral trade imbalance in the destination economy; (iii) trade resistance in the form of bilateral geographic distance as well as contemporaneous and pre-determined overall propensity to trade in the source and destination countries; and (iv) the aggregate income in the source and destination countries relative to the world economy as a whole.

Proposition 1. *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:*

$$\theta_{i,t} = (\Phi_{i,t} P_{ii,t})^{1-\eta}, \tag{3.18}$$

where

$$\Phi_{i,t} = \left[\sum_{j=1}^{N} \theta_{j,t} \Xi_{j,t} \left(\frac{d_{ij} x_{ij,t-1}^{-\chi_{ij}}}{P_{j,t}} \right)^{1-\eta} \right]^{1/(1-\eta)} \tag{3.19}$$

Proof. See Appendix C.1. □

Proposition 2. *If firms from source country $i \in n$ set prices $P_{ij,t}$ for each $j \in n$, and the production technology of final exported goods exhibits constant returns to scale, then import prices are proportional to the outward multilateral resistance of each source country:*

$$P_{ij,t} = \frac{d_{ij} \theta_{i,t}^{1/(1-\eta)}}{\Phi_{i,t}}. \tag{3.20}$$

Proof. See Appendix C.2. □

Lemma 1. *The gravity equation is dynamic when habits are non-zero, such that $\chi_{ij} > 0$ for all $i \in n \setminus j$. And when habits are asymmetric across countries, such that $\chi_{ij} \neq \chi_{ji}$ for all $i \in n \setminus j$, and/or the inward and outward the bilateral iceberg costs are non-identical, such that $d_{ij} \neq d_{ji} > 1$ for all $i \in n \setminus j$, the gravity equation is subject to the multilateral trade imbalance:*

$$A_{ij,t} = X_{ij,t} \times \frac{Y_t}{Y_{i,t} Y_{j,t}} = \Xi_{j,t} \left[\frac{d_{ij}^{1+\chi_{ij}}}{\Phi_{i,t} \Phi_{i,t-1}^{\chi_{ij}} P_{j,t}} \right]^{1-\eta} \left[\frac{\theta_{i,t-1}^{-\eta}}{A_{ij,t-1}^{1-\eta} Y_{j,t-1}^{1-\eta}} \right]^{\chi_{ij}}. \tag{3.21}$$

Proof. See Appendix C.3. □

Taking the natural logs on both sides of (3.21) and imposing the identity, $\theta_{i,t} = Y_{i,t}/Y_t$, gives the log-linear regression model specification for the dynamic supply-habit-augmented gravity equation:

$$\ln A_{ij,t} = \underbrace{\chi_{ij}(\eta-1)\ln A_{ij,t-1}}_{\text{size-adjusted bilateral trade flow persistence}} + \underbrace{\ln(\Xi_{j,t})}_{\text{destination multilateral trade imbalance}} - \underbrace{(1+\chi_{ij})(\eta-1)\ln d_{ij} + (\eta-1)\ln P_{j,t} + (\eta-1)\ln \Phi_{i,t} + \chi_{ij}(\eta-1)\ln \Phi_{i,t-1}}_{\text{bilateral and multilateral trade resistance}} + \underbrace{\chi_{ij}\eta \ln Y_{t-1} - \chi_{ij}\eta\ln Y_{i,t-1} + \chi_{ij}(\eta-1)\ln Y_{j,t-1}}_{\text{aggregate income}}. \tag{3.22}$$

According to the supply-habit-augmented gravity equation, the persistence of the size-adjusted bilateral trade flows $A_{ij,t}$ is increasing in the intensity of habits specific to each country pair $\chi_{ij} > 0$. Moreover, the size-adjusted bilateral trade flows $A_{ij,t}$ are increasing in the multilateral trade imbalance of the destination economy $\Xi_{j,t}$, decreasing in the iceberg costs d_{ij}, and decreasing in the lagged output of the home economy $Y_{i,t-1}$, but increasing in the multilateral trade resistance (consumer price index) of the destination country $P_{j,t}$, increasing in the current and lagged multilateral trade resistance of the home country $\Phi_{i,t}$ and $\Phi_{i,t-1}$, increasing in the lagged global output Y_{t-1}, and increasing in the lagged output of the destination economy $Y_{j,t-1}$. Importantly, the time-varying outward multilateral resistance term features a pair-specific coefficient, emphasizing the heterogeneous impact of aggregate trade frictions on each trade pair.

Corollary 1. *When habits are non-zero, but iceberg costs and habits are symmetrical 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 global trade flows are balanced, such that:*

$$\begin{aligned}\lim_{\chi_{ij}\to\chi\,\forall\, i\in n\setminus j} \ln A_{ij,t} = {} & \chi(\eta-1)\ln A_{ij,t-1} \\ & - (1+\chi)(\eta-1)\ln d_{ij} + (\eta-1)\ln P_{j,t} + (\eta-1)\ln\Phi_{i,t} + \chi(\eta-1)\ln\Phi_{i,t-1} \\ & + \chi\eta\ln Y_{t-1} - \chi\eta\ln Y_{i,t-1} + \chi(\eta-1)\ln Y_{j,t-1},\end{aligned} \tag{3.23}$$

since $\lim_{\chi_{ij}\to\chi\,\forall\, i\in n\setminus j}\Xi_{j,t} = 1$ *under the assumption that* $d_{ij} = d_{ji}$.

Corollary 2. *When habits are infinitesimally weak, 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):*

$$\lim_{\chi_{ij}\to 0\,\forall\, i\in n\setminus j} \ln A_{ij,t} = (1-\eta)\left[\ln d_{ij} - \ln\Phi_{i,t} - \ln P_{j,t}\right], \tag{3.24}$$

since $\lim_{\chi_{ij}\to 0\,\forall\, i\in n\setminus j}\Xi_{j,t} = 1$ *assuming that iceberg costs are symmetrical, such that $d_{ij} = d_{ji}$, which implies that* $\lim_{\chi_{ij}\to 0\,\forall\, i\in n\setminus j}\Phi_{i,t} = P_{i,t}$.

Observe that the multilateral trade resistance terms $\Phi_{i,t}$ are theoretically equivalent to the consumer price index $P_{i,t}$ in the symmetric static gravity equation, but not in the supply-habit-augmented gravity equation. This occurs even if the trade costs are symmetric both ways, because trade imbalance introduces a wedge between aggregate consumption and income, which feeds into the outward multilateral resistance term $\Phi_{i,t}$ (i.e., outward from the perspective of j), such that the aggregate export revenue is not equal to the aggregate import expenditure at any point in time. However, the inward and outward multilateral resistance terms $P_{j,t}$ and $\Phi_{i,t}$, respectively, are nonetheless dual to one another (see Proposition 1). This means that the supply-habit-augmented gravity equation is consistent with the static and the neo-classical gravity equations in that $P_{j,t}$ and $\Phi_{i,t}$ are defined up to a single normalization. As illustrated in Corollaries 1 and 2, our dynamic gravity model nests popular models in the literature and can therefore be easily tested if outperforms simpler alternatives.

In fact, the supply-habit-augmented gravity equation presents several key differences compared to the alternatives in the existing literature. First, parameter heterogeneity originates from the supply-side (i.e., technology-based), not the demand-side (i.e., preference-based), structural differences across country pairs. Consequently, bilateral trade flows in the supply-habit-augmented gravity equation adjust heterogeneously across country pairs in response to common shocks even in the absence of assumptions about special functional forms of preferences (e.g., non-homotheticity). This makes habits a remarkably tractable modelling choice (see Corollaries 1 and 2). Second, habits amplify the coefficient next to the bilateral iceberg costs d_{ij}, commonly known as the "trade elasticity", given by $|(1+\chi_{ij})(1-\eta)|$. While we do not discuss or quantify the resulting welfare gains from trade in this paper, it nonetheless hints at the fact that the elasticity of substitution η is not necessarily a catch-all parameter as argued by Arkolakis et al. (2012) for instance.

Next, habits generate a dynamic structure for the outward multilateral trade resistance term $\Phi_{i,t}$, which is not only time-varying, but its pre-determined value $\Phi_{i,t-1}$ also enters the supply-habit-augmented gravity equation with the trade pair specific intensity. By contrast, the static gravity equation contains only $\Phi_{i,t}$ (see Proposition 2). This leads to a fundamentally different transmission of country-specific (local) and worldwide (common) trade shocks, because $\Phi_{i,t}$ and $\Phi_{i,t-1}$ are strongly correlated with off-shore demand and off-shore supply (i.e., common factors). Importantly, homogeneous parameter gravity models with time fixed effects would fail to capture heterogeneous time trends. Last but not least, habit asymmetry creates differences in consumption home-bias across countries, which unlike the seminal contribution of Davis and Weinstein (2002), generates trade imbalances endogenously. It also explains additional variation in bilateral trade flows over time and across space beyond the standard gravity measures, such as aggregate income and geographic distance.

4 Empirical Analysis

In this section of the paper, we describe the data and the methodology used to estimate the supply-habit-augmented gravity equation, summarized in (3.22). The theory of habits in the supply chains 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. Finally, we cross-validate the theory of habits in

Table 1: Data Description

Variable	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
$\ln(Y_t)$	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.

the supply chains by quantifying the magnitude of habits using the global value chain indicators and linking them to the estimates of the trade persistence coefficients using a Poisson regression. We refer the reader to Appendix D.1 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.

4.1 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 habits in the supply chains. The value of size-adjusted bilateral trade flows ($\text{FLOW}_{ij,t}$) represents the dependent variable, explicitly defined in equation (3.21). Consistent with equation (3.17), 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 (GDP_t) is measured by the nominal gross domestic product in each location.

4.2 Methodology

The empirical adaptation of the supply-habit-augmented gravity equation (3.21) is a large N and large T panel regression model. Our approach extends the interactive fixed effects representation of Bai (2009) into a three-dimensional data structure. Specifically, we captures 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 (3.22) is mapped into the following estimating model:

$$\text{y}_{ij,t} = \beta_0 + \mathbf{x}'_{ij,t}\boldsymbol{\beta}_{ij} + u_{ij,t}, \tag{4.1}$$

$$u_{ij,t} = \boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t + \varepsilon_{ij,t}, \tag{4.2}$$

$$\mathbf{x}_{ij,t} = \boldsymbol{\gamma}'_{ij}\boldsymbol{\phi}_t + \boldsymbol{\nu}_{ij,t}, \tag{4.3}$$

where $\mathrm{y}_{ij,t} := \mathrm{FLOW}_{ij,t}$ are the trade flows, $\boldsymbol{\beta}_{ij} = [\beta_{1ij}, \beta_{2ij}, ..., \beta_{5ij}]'$ is a 5×1 vector of coefficients, $\mathbf{x}_{ij,t} = [\mathrm{FLOW}_{ij,t-1}, \mathrm{TB}_{j,t}, \mathrm{GDP}_{i,t-1}, \mathrm{GDP}_{j,t-1}, \mathrm{GDP}_{t-1}]'$ is a 5×1 vector of all common and country-specific observable variables, 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 factor loadings, respectively. The error terms $\varepsilon_{ij,t}$ and $\boldsymbol{\nu}_{ij,t}$ are assumed to be independently distributed of each other, uncorrelated with the unobservable common factors, and uncorrelated across country pairs.

Importantly, each estimation strategy discussed in detail in Section D.1 is nested as a special case of equations (4.1), (4.2), and (4.3) by: (i) choosing an estimator of $\boldsymbol{\beta}_{ij}$; and (ii) imposing restrictions on the inner product of $\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t$. For instance, the configuration of $\boldsymbol{\phi}_t = [1, 1, \tau_t]'$ and $\boldsymbol{\lambda}_{ij} = [\alpha_i, \alpha_j, 1]'$ gives rise to the traditional fixed effects (FE) error structure of Feenstra (2016), namely $u_{ij,t} = \alpha_i + \alpha_j + \tau_t + \varepsilon_{ij,t}$, where α_i and α_j are the country fixed effects and τ_t are the time fixed effects. However, FE pools the regressions coefficients by averaging $\mathbf{x}_{ij,t}$ across all source and destination countries, 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). By contrast, the so-called mean group (MG) estimator nullifies the inner product $\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t = 0$, but preserves parameter heterogeneity 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 $\boldsymbol{\beta} = 1/\bar{N}\sum_{ij=1}^{\bar{N}}\boldsymbol{\beta}_{ij}$.

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 $\boldsymbol{\lambda}_{ij} = [\alpha_i, \alpha_j]'$. Consequently, the only difference between the MG and the HFE approaches is that the error structure is now given by $u_{ij,t} = \alpha_i + \alpha_j + \varepsilon_{ij,t}$, 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 differences arise from the specification of the error structure. Specifically, we consider three additional FE alternatives: FE2 imposes $\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t = \alpha_{i,t} + \alpha_{j,t}$, FE3 imposes $\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t = \alpha_{ij} + \tau_t$, and FE4 imposes $\boldsymbol{\lambda}'_{ij}\boldsymbol{\phi}_t = \alpha_{i,t} + \alpha_{j,t} + \alpha_{ij}$.[12] In words, 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.

Our preferred, and baseline, empirical model, called the common correlated effects mean group (CCEMG) approach, imposes $\boldsymbol{\phi}_t = [1, 1, \mathbf{z}'_t]'$ and $\boldsymbol{\lambda}_{ij} = [\alpha_i, \alpha_j, \boldsymbol{\alpha}'_{ij}]'$, 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, are proxied by the cross-sectional averages of the dependent and independent variables.[13] As recent literature shows, the estimator is robust even if the time

[12]As covered in the Appendix D.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 $\boldsymbol{\phi}_t = [1, 1, \hat{\tau}_t]'$ and $\boldsymbol{\lambda}_{ij} = [\alpha_i, \alpha_j, \alpha_{ij}]'$, where $\hat{\tau}_t$ are the pre-estimated time fixed effects from the standard FE regression model. Notice that α_{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 $u_{ij,t} = \alpha_i + \alpha_j + \alpha_{ij}\hat{\tau}_t + \epsilon_{ij,t}$ 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.

[13]Analogous to Pesaran (2006), equation (4.3) 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 (4.1) and (4.2) generally delivers biased and inconsistent coefficient estimates whenever the unobservable common factors $\boldsymbol{\phi}_t$ are correlated with the

dimension is limited and the nature of unobserved common shocks (factors) is either deterministic or stochastic (Westerlund (2018); Westerlund et al. (2019)).

It is important to stress that the interactive fixed effect approach 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. In addition to the more flexible treatment of common shocks, we also explore the importance of parameter heterogeneity in our inference using the so-called common correlated effects pooled (CCEP) approach. It adopts an identical error structure to the CCEMG approach, but applies the pooled (homogeneous) regression coefficient estimator analogous to the FE approach. And finally, we consider the construction of the unobserved dynamic factors 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, which simply removes $\bar{\mathbf{x}}_t$ from the proxied unobserved common factors $\mathbf{z}_t$ and replaces it with $[\bar{\mathrm{y}}_t, \bar{\mathrm{y}}_{t-1}]'$. This means that the error structure of the CCEMG, CCEMGR, and CCEP approaches is given by $u_{ij,t} = \alpha_i + \alpha_j + \mathbf{z}_t' \boldsymbol{\alpha}_{ij} + \varepsilon_{ij,t}$, which distinguishes between unobservable country-specific time-invariant heterogeneity captured by α_i and α_j as well as unobservable time-varying heterogeneity specific to each country pair captured by the inner product of $\mathbf{z}_t' \boldsymbol{\alpha}_{ij}$.

4.3 Coefficient Estimates

Consider the coefficient estimates of the supply-habit-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 D.1.[14] 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 factors (CCEP), a restricted version of the common factor where only trade flows data are used (CCEMGR).

regressors $\mathbf{x}_{ij,t}$. And this constraint generally binds, since the supply-habit-augmented gravity equation (3.22) 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 (3.19)). However, we explicitly control for the endogeneity between the unobservable common factors and the regressors by assuming that the regressors are generated by equation (4.3). 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)).

[14] We do not cover Poisson regressions in the main text as they cannot be nested in the system (4.1)-(4.3) due to the inherent nonlinear structure, lacks tools to control for dynamic unobserved factors and cannot easily handle zero lagged values of trade flows. However, for the sake of completeness, the additional results generated using other variations of the FE and Poisson pseudo-maximum-likelihood (PPML) approaches are relegated to Table 5 in Appendix D.4. The PPML1-4 have the same characteristics for the fixed effects as in FE1-4. In all cases, zero lagged trade flows are dropped.

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 5.2. 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 β_{1ij} in equation (4.1)). 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 habits specific to each country pair χ_{ij} from the trade persistence coefficient $\beta_{1ij} = \chi_{ij}(\eta - 1)$. Specifically, β_1 is mapped directly to $1/\bar{N}\sum_{ij=1}^{\bar{N}} \chi_{ij}(\eta - 1)$ in equation (3.22), 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 Section 2.

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 β_{2ij} in equation (4.1)). As shown in equations (3.21) and (3.22), multilateral trade imbalance enters the dynamic gravity equation with unitary elasticity, such that 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

Table 2: Coefficient Estimates (Supply-Habit-Augmented Gravity Equation)

	CCEMG	FE	MG	CCEP	HFE	CCEMGR
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	$FLOW_{ij,t}$	$FLOW_{ij,t}$	$FLOW_{ij,t}$	$FLOW_{ij,t}$	$FLOW_{ij,t}$	$FLOW_{ij,t}$
$FLOW_{ij,t-1}$	0.347***	0.907***	0.540***	0.374***	0.488***	0.460***
	(0.00825)	(0.00451)	(0.00714)	(0.0161)	(0.00711)	(0.00730)
$TB_{j,t}$	0.975***	0.219***	0.793***	0.612***	0.865***	0.770***
	(0.126)	(0.0279)	(0.0714)	(0.0801)	(0.103)	(0.0989)
$GDP_{i,t-1}$	-0.312***	-0.00174	-0.189***	-0.296***	-0.232***	-0.210***
	(0.0778)	(0.00749)	(0.0304)	(0.0338)	(0.0283)	(0.0422)
$GDP_{j,t-1}$	-0.117	-0.0239***	-0.188***	-0.195***	-0.134***	-0.0634
	(0.0954)	(0.00714)	(0.0253)	0.0271	(0.0306)	(0.0454)
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.

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

(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 5 of Appendix D 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 ($\text{GDP}_{i,t-1}$), lagged destination country aggregate income ($\text{GDP}_{j,t-1}$), and lagged world aggregate income (GDP_{t-1}). According to the supply-habit-augmented gravity equation depicted in equation (3.22), the size-adjusted bilateral trade flows ($\text{FLOW}_{ij,t}$) are *ceteris paribus* increasing in $\text{GDP}_{j,t-1}$ and GDP_{t-1}, but decreasing in $\text{GDP}_{i,t-1}$. Consistent with the theory, we do find evidence of a small, negative, and statistically significant coefficient estimate for $\text{GDP}_{i,t-1}$ across the board. We also find that the coefficient estimate for 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 $\text{GDP}_{j,t-1}$ 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 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 D.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 supply-habit-augmented gravity equation, our results are robust to the aforementioned deviations from the baseline model.

4.4 Empirical Implications

Our opening results presented in Section 4.3 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 (3.22). 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 δ, measuring the annual rate of capital depreciation, to be anywhere from 0.06 to 0.14. Conversely, IMF (2015) estimates that the value of δ 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. And in support of this view, Section 3 of this paper develops a competing theory of trade persistence that extends the relative habits framework of Ravn et al. (2006) to capture a reduced form mechanism of inertia in the globalized wholesale distribution network. The main advantages of the theory of habits in the supply chains 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, the theory of habits in the supply chains 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 habits specific to any given country pair. If we take the values of η 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 habits 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 supply-habit-augmented gravity equation indicates that the habits-induced 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 as we describe it in Section 2.

5 Cross-Validation

5.1 Prediction Performance "Horse Race"

We have thus far established that controlling for the unobservable common factors and retaining parameter heterogeneity specific to each country pair when estimating the supply-habit-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 habits in the supply chains, 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, as covered in Section 2.

In order to illustrate that our preferred estimation strategy, titled CCEMG, outperforms the leading rival empirical strategies in terms of the data fit, especially in response to global trade shocks, we conduct a so-called "horse race" for the predictive performance of different empirical estimation strategies of our empirical model presented in equations (4.1)-(4.3). Specifically, Table 3 compares the Root Mean Square Errors (RMSE) calculated using the CCEMG, MG, CCEP, and FE methodologies (see Sections D.1 and 4.2 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 D.6 Table 8. 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 8 in Appendix D). While Tables 3 and 8 calculate 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 9 in Appendix D, 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

Table 3: Root Mean Square Error

Method	Full Sample	"Bad Times" w = 0	"Bad Times" w = 1	"Bad Times" w = 2	"Bad Times" w = 3	"Normal Times" w = 0	"Normal Times" w = 1	"Normal Times" w = 2	"Normal Times" 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 supply-habit-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.

recessions).[15] 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.

5.2 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 4.3 and 4.4). We now present Figure 3, 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 Sections D.1 and 4.2, which retains the country-pair-specific parameter heterogeneity and controls for the unobservable 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.

Nearly all of the trade persistence coefficients turn out to be positive, statistically significant, and their value is scattered around the interval of -0.2 and 0.5 (see Figure 3a) 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.[16] 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

[15] PPML-types of estimators are, however, not ideal in this context, see Appendix D.1

[16] One possible explanation for the negative trade persistence coefficient we obtain in South Africa for the period

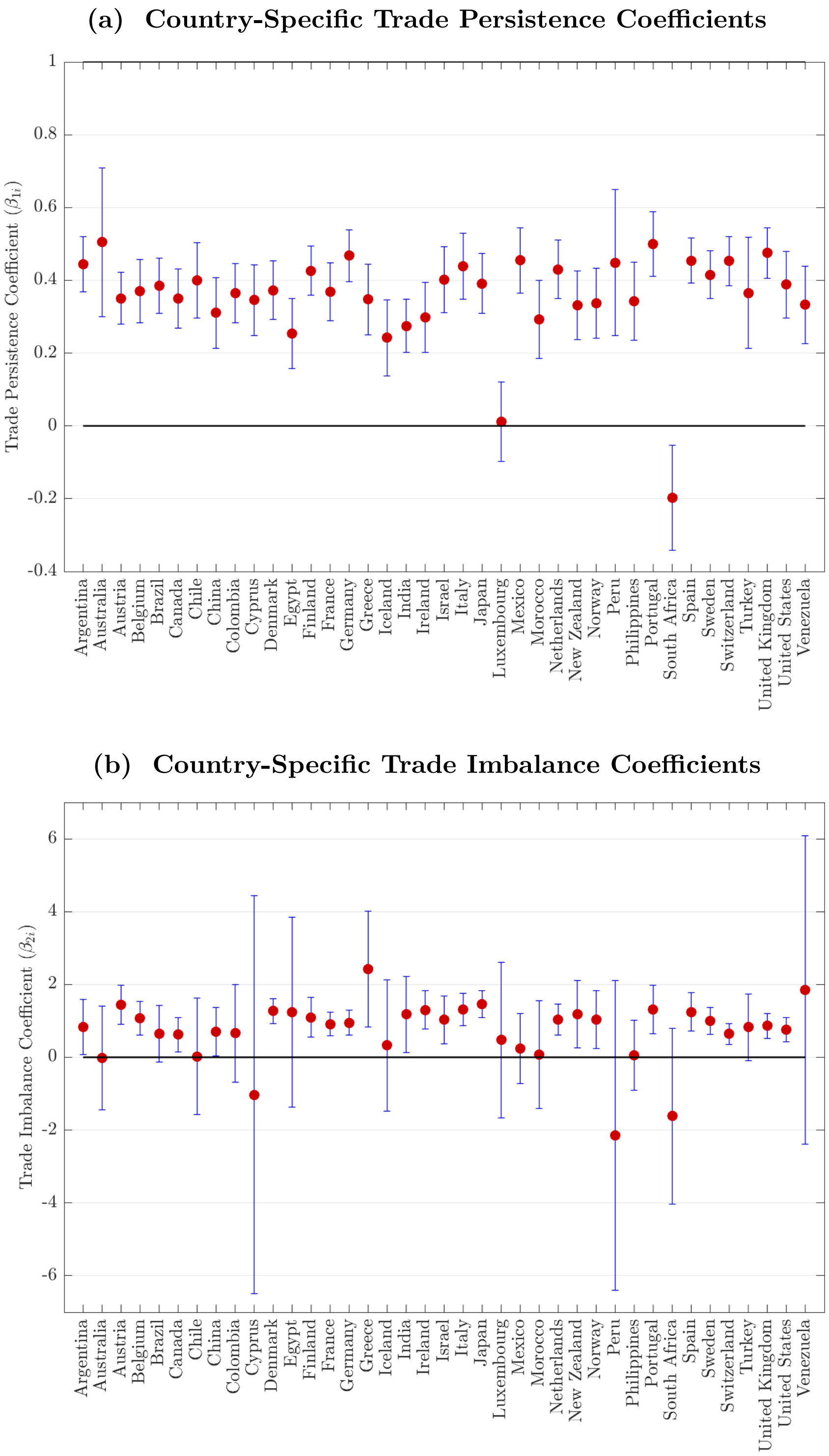

Figure 3: Cross-Country Heterogeneity of Trade Persistence and Trade Imbalance Coefficient Estimates (CCEMG)

Notes: Subplots (a) and (b) display the cross-country heterogeneity and the uncertainty surrounding the CCEMG coefficient estimates β_{1i} and β_{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.

larger (see Figure 3b); 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 habits in the supply chains.

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

5.3 Trade Persistence & Global Value Chains

The pervasive parameter heterogeneity presented in Section 5.2 raises an important question about what drives the cross-country differences in the estimates of the trade persistence coefficients. This section of the paper shows that the theoretical model presented in Section 3, in principle, links the estimated trade persistence coefficients to the participation in Global Value Chains (GVCs) specific to each country pair. We also discuss some of the obstacles we encounter when identifying the habit parameters from the estimates of the trade persistence coefficient specific to each country pair.

The theoretical model presented in Section 3 distinguishes between two different indicators of participation in GVCs that can be mapped directly to those measured by Casella et al. (2019), for instance. Specifically, the domestic value-added (DVA) and the foreign value-added (FVA) in domestic exports expressed as a share of domestic exports, such that $\text{DVA} + \text{FVA} = 1$. If the source country i is considered as the "domestic" economy and destination j is the "foreign" economy, then evaluating equation (3.12) at the steady state, it can easily be shown that

$$\text{FVA}_{ij} = \frac{X_{ij} - M_{ij}}{X_{ij}} \equiv 1 - x_{ij}^{-\chi_{ij}} \in [0,1] \Big|_{\chi_{ij}>0}, \tag{5.1}$$

where X_{ij} (M_{ij}) are the nominal trade flows of final (intermediate) goods from origin i to destination j, while x_{ij} are the analogous real trade flows of final goods in the steady state. It follows that FVA_{ij} is increasing in the habits parameter χ_{ij}. As a consequence, the unobservable and time-invariant deep habit parameter χ_{ij} could in principle be mapped directly to the observable time-averages of FVA_{ij}. Formally,

$$\chi_{ij} = -\frac{\ln(1 - \text{FVA}_{ij})}{\ln(x_{ij})}, \tag{5.2}$$

such that $\lim_{\text{FVA}_{ij} \to 0} \chi_{ij} = 0$ nests the classical "made here, sold there" case of arms length trade. Otherwise, if $\text{FVA}_{ij} > 0$, then $\chi_{ij} > 0$. Holding all else constant, the lower is the share of

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.

intermediate imports sourced from origin i in destination j (i.e., the closer FVA_{ij} is to zero), the closer is the deep habit parameter χ_{ij} to zero. And by extension, if destination j does not rely on intermediate imports from origin i, then the nominal value of trade flows from origin i to destination j are expected to be volatile rather than persistent, since the trade persistence coefficient is measured as $\beta_{1ij} = \chi_{ij}(\eta - 1)$, where $\eta > 0$ by assumption (see equation (3.22)).

While in theory the mapping between χ_{ij} and FVA_{ij} is relatively straightforward, identifying the habit parameter χ_{ij} empirically is much more difficult. Notice that the trade persistence coefficient $\beta_{1ij} = \chi_{ij}(\eta-1)$ identifies χ_{ij} and η jointly. One of the most common assumptions in modern trade theory is that the elasticity of substitution η is time-invariant and homogeneous across countries (e.g., Anderson and van Wincoop (2003)). Yet even the cross-sectionally averaged trade persistence coefficients presented in Figure 3 suggest evidence against this assumption. Specifically, if η was truly homogeneous across countries, then as long as $1 < \eta < \infty$, such that final imports from different source countries are considered to be imperfect substitutes, as is usually assumed to be the case, β_{1i} would be strictly non-negative, since $\chi_i > 0$ by definition. And yet in South Africa it is negative and statistically significant (see Figure 3) as is the case for some other bilateral trade persistence coefficients β_{1ij} (not displayed), not all of which can be treated as a sampling error. Consequently, direct mapping between χ_{ij} and FVA_{ij} is ultimately difficult to establish from the trade persistence coefficient estimates, because, contrary to the theory, our estimates suggest that parameter $\eta > 0$ is also likely to be country-specific rather than symmetric across all countries. Not least because our sample considers 39 developed and developing countries in which market structures are not only radically different at any given point in time, but also transformed at a heterogeneous pace over time.

Figure 4 presents a distribution of the data-implied destination-specific habit parameters. As expected, the habit parameters are relatively small (i.e., ranging between 0 and 0.28) and the cross-country distribution is skewed and bi-modal, suggesting two distinct clusters: one close to zero and the other equal to around 0.12. We explore whether the small habit parameter values that we obtain can successfully explain the trade persistence heterogeneity in Section 5.4. If so, then the habit framework serves not only as theoretical motivation for an empirical model of the dynamic gravity equation, but also as a theoretical tool that helps explain the trade persistence puzzle.

5.4 What Drives Trade Persistence Heterogeneity?

We validate the theory of habits in the supply chains by identifying the empirical determinants of the bilateral trade persistence heterogeneity emphasized in Section 4.3. We adopt a non-linear Poisson cross-sectional regression model. We use data spanning more than six decades to uncover persistence parameters. Specifically, we take the CCEMG estimates of β_{1ij} in equation (4.1) (i.e., the dependent variable) and construct the theory-consistent bilateral "Habit" χ_{ij} in equation (5.2) (i.e., the independent variable). The theory of habits in the supply chains predicts a trade persistence coefficient equal to $\beta_{1ij} = \chi_{ij}(\eta - 1)$ (see equation 3.21), which is a product of χ_{ij} and η. Our choice of adopting the non-linear Poisson empirical model specification, as opposed to a log-linear specification for instance, is thus guided by the fact that habits χ_{ij} are identified multiplicatively and jointly with the elasticity of substitution η. As pointed by Santos-Silva and Tenreyro (2007), ordinary least squares can be subject to severe biases when cross-sectional data are used to fit

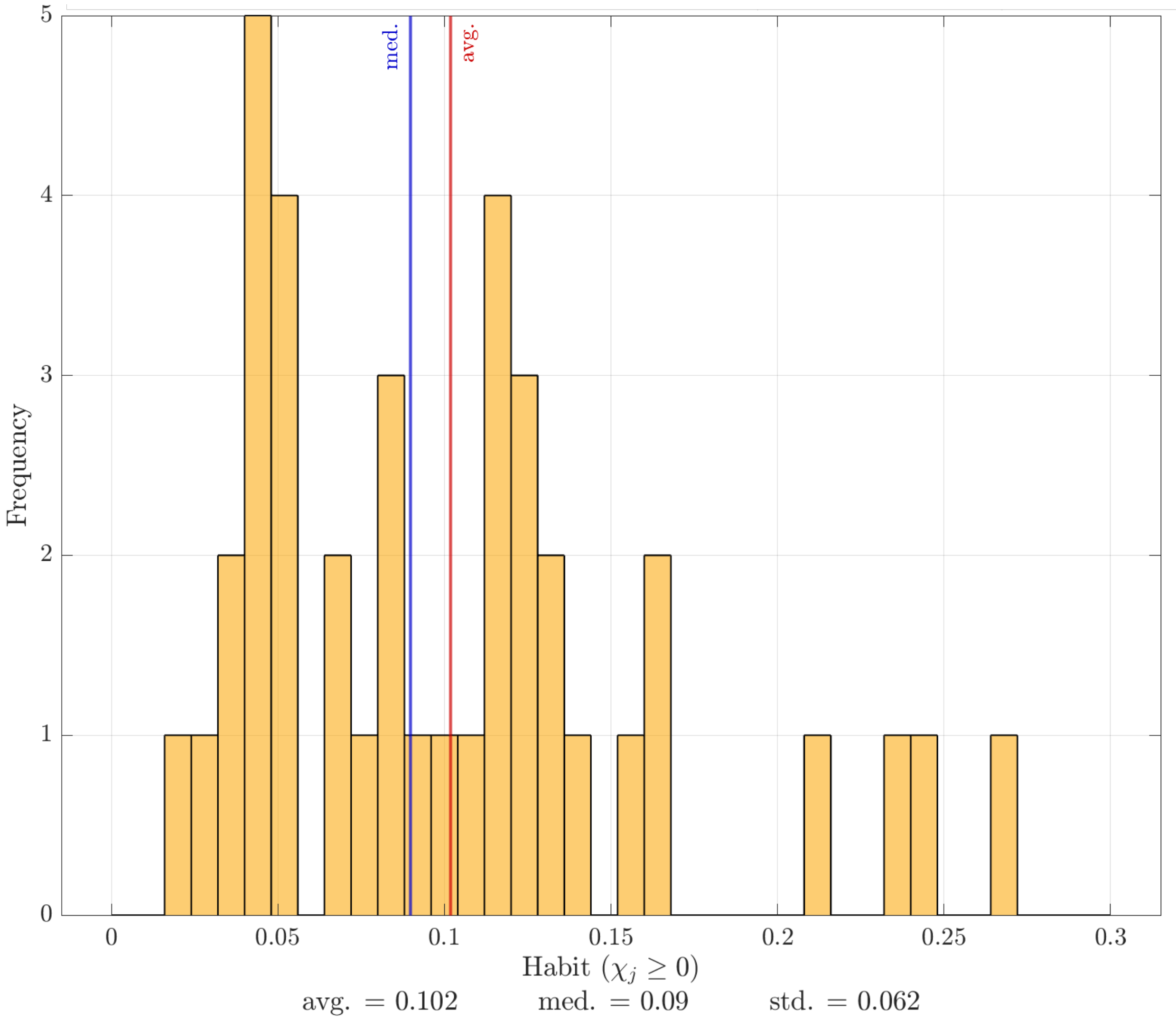

Figure 4: Distribution of Quantified Habits

Notes: The figure depicts destination-specific habits derived from the time-averaged data on the foreign value added in exports and trade flows as suggested in equation (5.2). We use data from 39 countries over the period of 1990-2014. We quantify the destination-specific habit χ_j by calculating the total foreign value-added in exports FVA_j from the UNCTAD-Eora database (see Casella et al. (2019)) and use total trade inflows at constant prices from the Penn World tables (see Feenstra et al. (2015)) in order to calculate x_j.

Table 4: The Non-Linear Cross-Sectional Poisson Regression Model of Trade Persistence and Global Value Chains

	$t_{\beta_{1ij}} > 1.96$	$t_{\beta_{1ij}} > 2.326$	$t_{\beta_{1ij}} > 2.576$	$t_{\beta_{1ij}} > 3.09$
	(1)	(2)	(3)	(4)
Variables	$\beta_{1ij} > 0$	$\beta_{1ij} > 0$	$\beta_{1ij} > 0$	$\beta_{1ij} > 0$
Log-Supply-Habit	0.0478**	0.0523**	0.0597***	0.0697***
	(0.0217)	(0.0222)	(0.0219)	(0.0247)
Log-Distance	-0.0442***	-0.0432***	-0.0456***	-0.0409***
	(0.0164)	(0.0158)	(0.0152)	(0.0145)
Common Language	0.0628*	0.0471	0.0358	0.0244
	(0.0341)	(0.0320)	(0.0311)	(0.0309)
Colony	0.229	0.204	0.158	0.116
	(0.158)	(0.133)	(0.139)	(0.141)
Constant	0.267	0.445*	0.586**	0.734**
	(0.279)	(0.268)	(0.260)	(0.285)
Number of Observations	869	781	728	607

Notes: Robust standard errors associated with the Huber/White/sandwich coefficient estimates are in parentheses. All regression models incorporate fixed effects specific to each source and destination country.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

log-linearized multiplicative models under the presence of heteroscedasticity. Unlike main model, we deal with the cross-sectional data and do not run into issues of applying PPML in a dynamic setting.

In addition, we recognize that β_{1ij} coefficients contain some randomness due to their estimation errors. In order to account for the estimation errors in our dependent variable, we condition β_{1ij} on several general-to-specific subsamples based on different thresholds of statistical significance. Finally, recognizing that a mapping in equation (4.1) includes FVA_{ij} and bilaterally varying trade of composite goods x_{ij}, we apply standard gravity resistance terms as additional controls. As a result, other independent variables in our cross-sectional analysis include: "Colony", which assumes the value of unity if the country pair has shared a common colonizer; "Common language", which assumes the value of unity if at least 9% of the population speak the same language in both countries; and "Distance", which captures the geographic distance between the capital cities in both countries. This approach also helps us address two additional empirical issues.

First, the GVC-based habits that we calculate are a simple proxy of multinational firm level activity, such as inter-temporal contractual obligations on the globalized production belt line, which inhibit immediate assembling, disbanding, or swapping of the off-shore suppliers in response to shocks. But we recognize that there are other factors beyond the GVC participation indicators that affect country-specific habit formation and bilateral trade persistence. Specifically, at the aggregate level, countries adapt their technology by developing habits of importing certain varieties of final and intermediate goods from other countries that are known for popular brands or their reputation, which may be related to shared values, history, institutions, or colonial ties (i.e., the broad definition of "distance"). Second, the elasticity of substitution η is not directly observable, but contrary to the conventional wisdom and as discussed in Section 5.3, it is likely to exhibit at least some cross-sectional variation given the long time horizon and a selection both developing and advanced economies in our estimation sample. In fact, Carrere et al. (2020) have recently demonstrated, using quantile regression instead of ordinary least squares, that the trade elasticity is variable and decreases in the volume of trade (this approach is further reinforced by the literature on the determinants of the trade elasticity, e.g., Mayer and Zignago (2011), Boehm et al. (2020)). Coupled with the need to control for exogenous drivers of bilateral trade flows x_{ij}, we proceed with a conditional cross-sectional regression that also accounts for unobservable country-specific heterogeneity (directional fixed effects).

Table 4 presents the coefficient estimates from our cross-sectional Poisson regression model. First, our results show that trade persistence is increasing in GVC-based habits and more so with greater statistical significance of the trade persistence coefficients (see the first line in Table 4). Consistent with the theory of habits in the supply chains, this result implies that the value of bilateral trade flows between country pairs where the source country is dependent on intermediate imports from the destination country (i.e., integrated) tend to be more persistent in response to negative trade shocks than when countries trade at "arm's length" (i.e., disintegrated). Second, trade persistence is decreasing in geographic distance between the source and the destination countries (see the second line in Table 4). This means that the value of trade flows between trade partners with greater geographic proximity tend to be more persistent over time than the value of trade flows between geographically remote trade partners. Third, we find some evidence that countries who share a common language may exhibit greater trade persistence as they may share similar values and institutions (see the third line in Table 4). However, once we control for the cross-sectional variation in GVC-based habits and distance, we find that the effects of colonial ties and common language on trade persistence are generally not statistically significant. For the sake of robustness, we also implement an analogous cross-sectional Poisson regression for the trade persistence coefficient estimates obtained from the re-specified supply-habit-augmented gravity equation as discussed in Sections 3.4 and 4.3 (see Table 10 in Appendix D.7). We find that our results from the cross-sectional analysis are robust to the potential issues of model misspecification in the dynamic gravity equation.

Our empirical results offer nuanced policy implications. While global value chains are a source of "strength" in the face of common shocks, long-distance trade appears to be a sign of a "weakness". This means that bilateral trade flows are more resilient to local and worldwide shocks between countries that source their intermediate goods from a regionally diversified and more proximate set of suppliers. Therefore, policymakers should take dynamic considerations into account, particularly the trade network resilience and trade volatility, when developing multilateral trade policies in the face of global business cycles.

6 Counterfactual Trade Flows

After having shown that the theory of habits in supply chains 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. We study a specific country pair of particular interest, namely the USA and China, and illustrate the importance of trade persistence heterogeneity in predicting the trade adjustment dynamics in response to a multilateral trade imbalance shock.[17] 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 (2019b)). 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 $\frac{\partial \mathbb{E}_t \ln A_{ij,t+h}}{\partial \ln \Xi_{j,t}} = \left(\chi_{ij}\left(\eta - 1\right)\right)^h$, because the theory of habits in the supply chains predicts a unitary trade imbalance coefficient (see equation (3.22)). 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 $\chi_{ij}\left(\eta - 1\right)$. 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 (3.22) 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 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.[18]

The difference in adjustment exemplifies the importance of heterogeneous responses for coun-

[17] 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.

[18] 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.

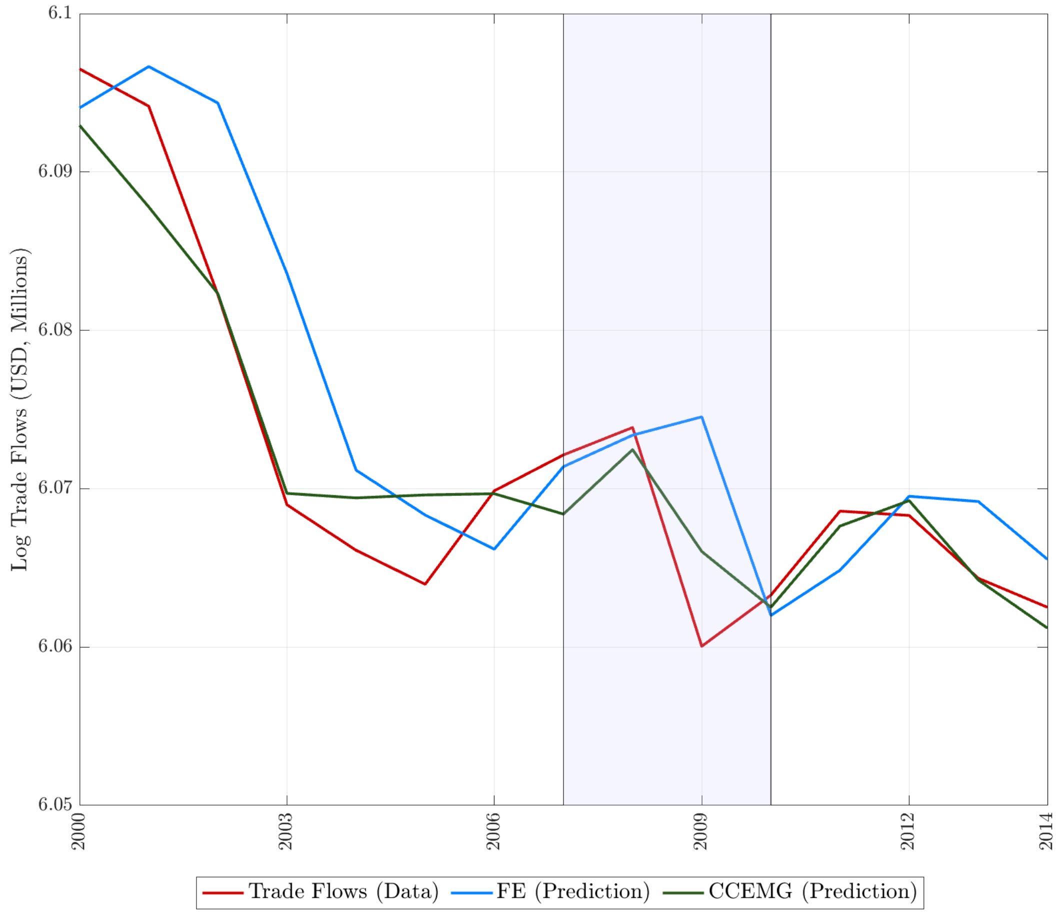

Figure 5: Predicted Size-Adjusted Trade Flows Between USA and China (Log Scale)

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 4.2.

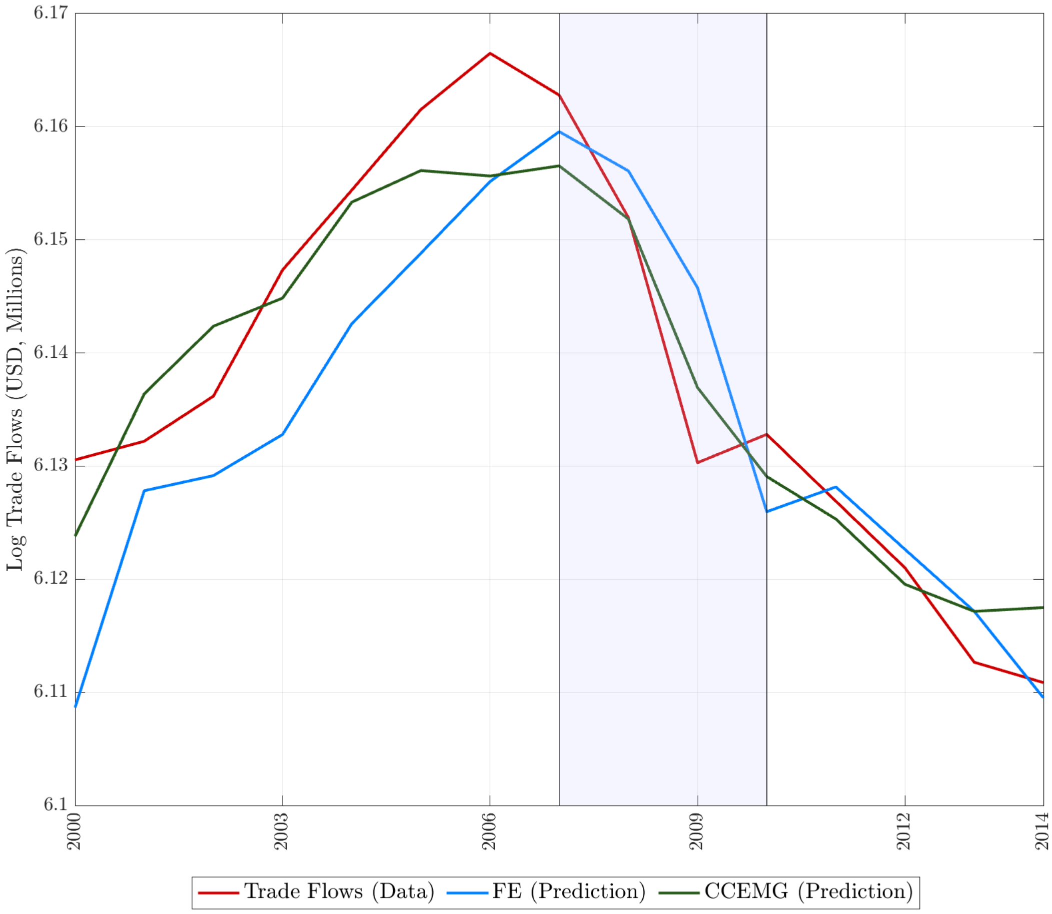

Figure 6: Predicted Size-Adjusted Trade Flows Between China and USA (Log Scale)

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 4.2.

terfactual predictions. Figures 5-6 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 the future research.

7 Concluding Remarks

International trade flows are volatile, imbalanced, and fragmented across off-shored supply chains. 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 international trade flows that are difficult to square with the sharp and synchronized trade adjustments in response to common shocks.

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 habits in the supply chains. Our framework offers several advantages. First, habits predict autocorrelated (i.e., persistent) trade flows and trade persistence that is heterogeneous across country pairs. Second, cross-country habit 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, habits enhance the geographic distance component of international trade costs, because distance applies not only to goods that are "made here, sold there", but also to intermediate inputs that are "bought, sold, and bought again". Fourth, habits create 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 off-shore 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 common trade shocks. For instance, the standard "country" fixed effects are time-invariant, while the "time" fixed effects are homogeneous for all country pairs. This implies that shocks originating from third countries are not fully reflected in either the source or the destination economies. 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, 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 heterogeneous responses to common shocks specific to each country pair by approximating the multilateral trade resistance terms as unobservable variation in common dynamic factors, which we model empirically as the cross-sectionally averaged country-specific regressors. We show that absent of the unobservable common 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 habits in the supply chains.

We document pervasive heterogeneity of the trade persistence coefficients across countries. We demonstrate that our theory predicts a direct mapping between habits and the trade persistence coefficients, which are related to the cross-sectional variation in the indicators of global value chain participation. Our results offer several policy implications. While global value chains are a source of "strength" in the face of common shocks, long-distance trade appears to be a sign of a "weakness". This means that bilateral trade flows are more resilient to local and worldwide shocks between countries that source their intermediate goods from a regionally diversified and closer set of suppliers.

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 habits in the supply chains 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). An extension to functional elasticities, varying by the GVC participation as well as more elaborate technology to account for the GVC complexity are worth pursuing further. 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 common shocks.

References

Abel, A. B. (1990). Asset Prices under Habit Formation and Catching Up with the Joneses. American Economic Review 80(2), 38–42.

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., C. Arkolakis, and K. J. Ruhl (2021). Firm Dynamics and Trade. Annual Review of Economics 13(1), 253–280.

Alessandria, G., J. P. Kaboski, and V. Midrigan (2010). The Great Trade Collapse of 2008–09: An Inventory Adjustment? IMF Economic Review 58(2), 254–294.

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

Altomonte, C., F. D. Mauro, G. Ottaviano, A. Rungi, and V. Vicard (2012). Global Value Chains During the Great Trade Collapse: A Bullwhip Effect? Working Paper Series 1412, European Central Bank.

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.

Antonakakis, N. (2012). The Great Synchronization of International Trade Collapse. Economics Letters 117(3), 608–614.

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

Arkolakis, C. (2016, 02). A Unified Theory of Firm Selection and Growth. The Quarterly Journal of Economics 131(1), 89–155.

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

Arkolakis, C., F. Huneeus, and Y. Miyauchi (2021). Spatial Production Networks. Mimeo, Yale University.

Backus, D. K. and G. W. Smith (1993). Consumption and Real Exchange Rates in Dynamic Economies with Non-Traded Goods. Journal of International Economics 35(3), 297 – 316.

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.

Beirne, J., N. Renzhi, and U. Volz (2020). Persistent Current Account Imbalances: Are they Good or Bad for Regional and Global Growth? ADBI Working Papers 1094, Asian Development Bank Institute.

Bems, R., R. C. Johnson, and K.-M. Yi (2010). Demand Spillovers and the Collapse of Trade in the Global Recession. IMF Economic Review 58(2), 295–326.

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.

Campbell, J. Y. and J. H. Cochrane (1999). By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. Journal of Political Economy 107(2), 205–251.

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.

Casella, B., R. Bolwijn, D. Moran, and K. Kanemoto (2019). Improving the Analysis of Global Value Chains: the UNCTAD-Eora Database. Transnational Corporations 26(3), 115–142.

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

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

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.

Costinot, A., D. Donaldson, and I. Komunjer (2012). What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas. Review of Economic Studies 79(2), 581–608.

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.

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.

Eaton, J., S. Kortum, B. Neiman, and J. Romalis (2016). Trade and the Global Recession. American Economic Review 106(11), 3401–38.

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.

Egger, P. (2000). A Note on the Proper Econometric Specification of the Gravity Equation. Economics Letters 66(1), 25 – 31.

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.

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.

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.

Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. The Quarterly Journal of Economics 123(2), 441–487.

Herbst, E. P. and F. Schorfheide (2016). Bayesian Estimation of DSGE Models (1 ed.). The Econometric and Tinbergen Institutes Lectures. Princeton University Press.

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 (2019a). Global Imbalances. IMF Staff Note, International Monetary Fund, Washington, D.C.

IMF (2019b). 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.

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

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

Levchenko, A. A., L. T. Lewis, and L. L. Tesar (2010). The Collapse of International Trade During the 2008-09 Crisis: In Search of the Smoking Gun. IMF Economic Review 58(2), 214–253.

Mayer, T. and S. Zignago (2011). Notes on CEPII's Distances Measures: The GeoDist Database. Working Papers 2011-25, CEPII research center.

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

Micco, A., E. Stein, and G. Ordonez (2003). The Currency Union Effect on Trade: Early Evidence from EMU. Economic Policy 18(37), 315–356.

Morales, E., G. Sheu, and A. Zahler (2019). Extended Gravity. Review of Economic Studies 86(6), 2668–2712.

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

Novy, D. and A. M. Taylor (2020). Trade and Uncertainty. The Review of Economics and Statistics 102(4), 749–765.

Obstfeld, M. and K. S. Rogoff (1996). Foundations of International Macroeconomics, Volume 1 of MIT Press Books. The MIT Press.

Panigrahi, P. P. (2021). Endogenous Spatial Production Networks: Quantitative Implications for Trade and Productivity. Working papers, University of Berkeley.

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.

Ravn, M. O., S. Schmitt-Grohe, and M. Uribe (2007). Pricing to Habits and the Law of One Price. American Economic Review 97(2), 232–238.

Sampson, T. (2015, 09). Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth. The Quarterly Journal of Economics 131(1), 315–380.

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.

Serlenga, L. and Y. Shin (2007). Gravity Models of Intra-EU Trade: Application of the CCEP-HT Estimation in Heterogeneous Panels with Unobserved Common Time-Specific Factors. Journal of Applied Econometrics 22(2), 361–381.

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

The production technology of a distributor adopts the following functional form:

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

The distributor operates in a perfectly competitive market structure, such that they minimize production costs by choosing the amount of commodities to import form each sector subject to the above augmented CES production technology

$$\begin{aligned} \min_{\{m_{ij,t}(\omega)\}} \quad & \tilde{P}_{ij,t} x_{ij,t} - \int_0^1 P_{ij,t}(\omega) m_{ij,t}(\omega) d\omega \\ \text{s.t.} \quad x_{ij,t} &= \left[\int_0^1 \left(m_{ij,t}(\omega) x_{ij,t-1}^{\chi_{ij}} \right)^{1-1/\eta} d\omega \right]^{1/(1-1/\eta)}. \end{aligned} \tag{A.2}$$

The first order condition is given by:

$$\tilde{P}_{ij,t} x_{ij,t}^{1/\eta} (m_{ij,t}(\omega) x_{ij,t-1}^{\chi_{ij}})^{-1/\eta} x_{ij,t-1}^{\chi_{ij}} - P_{ij,t}(\omega) = 0, \tag{A.3}$$

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

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

$$\tilde{P}_{ij,t} x_{ij,t} = \int_0^1 P_{ij,t}(\omega) m_{ij,t}(\omega) d\omega \tag{A.5}$$

The break-even price index of the distributors is then derived by substituting the demand for intermediate imports into the 'zero-profit' condition:

$$\tilde{P}_{ij,t} = \left[\int_0^1 (P_{ij,t}(\omega) x_{ij,t-1}^{-\chi_{ij}})^{1-\eta} d\omega \right]^{1/(1-\eta)}. \tag{A.6}$$

The aggregate demand for intermediate imports is therefore derived by integrating across all varieties of intermediate imports:

$$\begin{aligned} m_{ij,t} &= \int_0^1 m_{ij,t}(\omega) d\omega, \\ &= x_{ij,t} x_{ij,t-1}^{\chi_{ij}(\eta-1)} \int_0^1 \left[\frac{P_{ij,t}(\omega)}{\tilde{P}_{ij,t}} \right]^{-\eta} d\omega, \\ &= x_{ij,t} x_{ij,t-1}^{-\chi_{ij}}, \end{aligned} \tag{A.7}$$

such that the dynamic demand for aggregate imports is given by

$$x_{ij,t} = m_{ij,t} x_{ij,t-1}^{\chi_{ij}}. \tag{A.8}$$

A.2 Duality Problem

The representative consumer minimizes the consumption expenditure on composite goods from each source country subject to CES preferences:

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

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

$$P_{j,t} c_{j,t}^{1/\eta} x_{ij,t}^{-1/\eta} - \tilde{P}_{ij,t} = 0. \tag{A.9}$$

Rearranging the above gives the demand schedule for each composite tradable good:

$$x_{ij,t} = c_{j,t} \left(\frac{\tilde{P}_{ij,t}}{P_{j,t}} \right)^{-\eta}, \tag{A.10}$$

The consumer price index is derived by substituting the above demand schedule into the CES preferences displayed above and solving for $P_{j,t}$, which gives rise to the following expression:

$$P_{j,t} = \left[\sum_{i=1}^{N} \tilde{P}_{ij,t}^{1-\eta} \right]^{1/(1-\eta)}. \tag{A.11}$$

A.3 Consumption Smoothing

The consumer maximizes the lifetime utility subject to an indefinite sequence of budget constraints by choosing the aggregate consumption:

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

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

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

First order condition:

$$\frac{\beta^t}{c_t} - \Lambda_{j,0} \zeta_{j,0,t} = 0 \quad \Leftrightarrow \quad \frac{\beta^t}{c_t} = \Lambda_{j,0} \zeta_{j,0,t}. \tag{A.12}$$

The expected ratio of the first order conditions at period $t+1$ then givens the Euler equation (3.8). With perfect capital mobility, $\zeta_{j,0,t} = \zeta_{i,0,t}$ is the no-arbitrage condition.

A.4 Wholesalers

The optimal nominal flexible price of the intermediate good exporters $P_{ij,t}(\omega)$, who adopt the "pricing-to-habits" strategy, is one that maximizes the current monopolistically-competitive profit dividends:

$$\max_{\{P_{ij,t}(\omega)\}} \left[P_{ij,t}(\omega) - d_{ij} MC_{i,t}\right] m_{ij,t}(\omega)$$
$$\text{s.t.} \quad m_{ij,t}(\omega) = \left[\frac{P_{ij,t}(\omega)}{\tilde{P}_{ij,t}}\right]^{-\eta} x_{ij,t} x_{ij,t-1}^{\chi_{ij}(\eta-1)}.$$

The first-order conditions with respect to the nominal price $P_{ij,t}(\omega)$ is given by

$$(1-\eta) m_{ij,t}(\omega) + \eta \left(\frac{d_{ij} MC_{i,t}}{P_{ij,t}(\omega)}\right) m_{ij,t}(\omega) = 0, \tag{A.13}$$

or alternatively

$$P_{ij,t}(\omega) = \left(\frac{\eta}{\eta-1}\right) d_{ij} MC_{i,t}. \tag{A.14}$$

A.5 Transversality Condition

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

$$\begin{aligned}
b_{j,t} =& \zeta_{j,t,t+1} \mathbb{E}_t [b_{j,t+1} + \underbrace{c_{j,t} - \varpi_{j,t} - w_{j,t} h_j}_{nx_{j,t}}], \\
=& \zeta_{j,t,t+1} \mathbb{E}_t \left[\zeta_{j,t+1,t+2}(b_{j,t+2} - nx_{j,t+1}) - nx_{j,t}\right], \\
=& \zeta_{j,t,t+1}\zeta_{j,t+1,t+2} b_{j,t+2} - \zeta_{j,t,t+1}(nx_{j,t} + \zeta_{j,t+1,t+2} nx_{j,t+1}), \\
\ldots& \\
=& \zeta_{j,t,t+S} b_{j,t+S} - \sum_{s=0}^{S} \zeta_{j,t,t+s+1} nx_{j,t+s}.
\end{aligned} \tag{A.15}$$

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

$$\lim_{S\to\infty} \zeta_{j,t,t+S} = \zeta_{j,t,t+1} \times \zeta_{j,t+1,t+2} \times \zeta_{j,t+2,t+3} \times \cdots \times \zeta_{j,S-1,S} = 0. \tag{A.16}$$

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:

$$b_j = \beta^S b_j - nx_j \sum_{s=0}^{S} \beta^{1+s} \quad \Rightarrow \quad \lim_{S\to\infty} b_j = -nx_j \sum_{s=0}^{\infty} \beta^{1+s}, \tag{A.17}$$

$$= -nx_j \left(\frac{\beta}{1-\beta}\right) > -\infty. \Big|_{\beta\in(0,1)} \tag{A.18}$$

B Dynamic Gravity Equation

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

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

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

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

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, consider the aggregate income identity:

$$Y_{i,t} = \sum_{j=1}^{N} X_{ij,t}, \tag{B.5}$$

Substituting (B.4) into (B.5) gives

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

Note that the import price $P_{ij,t}$ is proportional to export price $P_{ii,t}$, where the proportionality corresponds to the iceberg costs:

$$P_{ij,t} = d_{ij}P_{ii,t}. \tag{B.7}$$

Substituting (B.7) into (B.6) gives

$$Y_{i,t} = P_{ii,t}^{1-\eta}\sum_{j=1}^{N} Y_{j,t}\Xi_{j,t}\left[\frac{d_{ij}x_{ij,t-1}^{-\chi_{ij}}}{P_{j,t}}\right]^{1-\eta}, \tag{B.8}$$

Now let $\theta_{i,t} = Y_{i,t}/Y_t$, where $Y_t = \sum_{j=1}^{N} Y_{j,t}$. Then solving (B.8) for the export price scaled by the trade elasticity $P_{ii,t}^{1-\eta}$ gives

$$\begin{aligned} P_{ii,t}^{1-\eta} &= \frac{Y_{i,t}}{\sum_{j=1}^{N} Y_{j,t}\Xi_{j,t}\left[\frac{d_{ij}x_{ij,t-1}^{-\chi_{ij}}}{P_{j,t}}\right]^{1-\eta}}, \\ &= \frac{\theta_{i,t}}{\sum_{j=1}^{N} \theta_{j,t}\Xi_{j,t}\left[\frac{d_{ij}x_{ij,t-1}^{-\chi_{ij}}}{P_{j,t}}\right]^{1-\eta}}, \\ &= \theta_{i,t}\Phi_{i,t}^{\eta-1}, \end{aligned} \tag{B.9}$$

where

$$\Phi_{i,t} = \left[\sum_{j=1}^{N} \theta_{j,t} \Xi_{j,t} \left(\frac{d_{ij} x_{ij,t-1}^{-\chi_{ij}}}{P_{j,t}} \right)^{1-\eta} \right]^{1/(1-\eta)} \tag{B.10}$$

defines the 'multilateral resistance' of destination i to trade flows from all trade partner countries $j \in n \setminus i$. Next, substitute (B.9) out of (B.4) using the proportionality condition (B.7) to obtain

$$X_{ij,t} = \Xi_{j,t} \left[\frac{d_{ij} x_{ij,t-1}^{-\chi_{ij}}}{\Phi_{i,t} P_{j,t}} \right]^{1-\eta} \frac{Y_{i,t} Y_{j,t}}{Y_t}. \tag{B.11}$$

Finally, note that

$$X_{ij,t-1} = P_{ij,t-1} x_{ij,t-1} = d_{ij} P_{ii,t-1} x_{ij,t-1},$$

such that the stock of habits can be replaced by

$$\begin{aligned} x_{ij,t-1}^{-\chi_{ij}} &= \left(\frac{X_{ij,t-1}}{d_{ij} P_{ii,t-1}} \right)^{-\chi_{ij}}, \\ &= \left(\frac{X_{ij,t-1} \Phi_{i,t-1}}{d_{ij} \theta_{i,t-1}^{1/(1-\eta)}} \right)^{-\chi_{ij}}. \end{aligned} \tag{B.12}$$

Substituting (B.12) into (B.11) therefore gives a dynamic gravity equation:

$$X_{ij,t} = \Xi_{j,t} \left[\frac{d_{ij}^{1+\chi_{ij}}}{\Phi_{i,t} \Phi_{i,t-1}^{\chi_{ij}} P_{j,t}} \right]^{1-\eta} \left[\frac{\theta_{i,t-1}}{X_{ij,t-1}^{1-\eta}} \right]^{\chi_{ij}} \frac{Y_{i,t} Y_{j,t}}{Y_t}, \tag{B.13}$$

$$\frac{X_{ij,t} Y_t}{Y_{i,t} Y_{j,t}} = A_{ij,t} = \Xi_{j,t} \left[\frac{d_{ij}^{1+\chi_{ij}}}{\Phi_{i,t} \Phi_{i,t-1}^{\chi_{ij}} P_{j,t}} \right]^{1-\eta} \left[\frac{\theta_{i,t-1}}{X_{ij,t-1}^{1-\eta}} \right]^{\chi_{ij}}, \tag{B.14}$$

$$= \Xi_{j,t} \left[\frac{d_{ij}^{1+\chi_{ij}}}{\Phi_{i,t} \Phi_{i,t-1}^{\chi_{ij}} P_{j,t}} \right]^{1-\eta} \left[\frac{\theta_{i,t-1}^{\eta}}{A_{ij,t-1}^{1-\eta} Y_{j,t-1}^{1-\eta}} \right]^{\chi_{ij}}, \tag{B.15}$$

since $A_{ij,t}^{1-\eta} = [X_{ij,t} Y_t / (Y_{i,t} Y_{j,t})]^{1-\eta} = [X_{ij,t}/(\theta_{i,t} Y_{j,t})]^{1-\eta}$, thus $X_{ij,t}^{1-\eta}/\theta_{i,t} = A_{ij,t}^{1-\eta} Y_{j,t}^{1-\eta}/\theta_{i,t}^{\eta}$. Taking natural logs on both sides of (B.13) thus gives rise to the dynamic gravity equation regression function specification:

$$\begin{aligned} \ln A_{ij,t} = & \underbrace{\chi_{ij}(\eta - 1) \ln A_{ij,t-1}}_{\text{size-adjusted bilateral trade flow persistence}} \\ & + \underbrace{\ln(\Xi_{j,t})}_{\text{multilateral trade imbalance}} \\ & - \underbrace{(1 + \chi_{ij})(\eta - 1) \ln d_{ij} + (\eta - 1) P_{j,t} + (\eta - 1) \ln \Phi_{i,t} + \chi_{ij}(\eta - 1) \ln \Phi_{i,t-1}}_{\text{bilateral and multilateral trade resistance}} \\ & + \underbrace{\chi_{ij}\eta \ln Y_{t-1} - \chi_{ij}\eta \ln Y_{i,t-1} + \chi_{ij}(\eta - 1) \ln Y_{j,t-1}}_{\text{aggregate income}}, \end{aligned} \tag{B.16}$$

since $\ln \theta_{i,t-1} = \ln Y_{i,t-1} - \ln Y_{t-1}$.

C Proofs of Propositions and Lemma

C.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 demand for composite goods $X_{ij,t} = C_{j,t}[\tilde{P}_{ij,t}/P_{j,t}]^{1-\eta}$, the break-even price index $\tilde{P}_{ij,t} = P_{ij,t} x_{ij,t-1}^{-\chi_{ij}}$, and the aggregate consumption identity $C_{j,t} = Y_{j,t}\Xi_{j,t}$. Substitute each of these schedules into the aggregate income identity of the source country $i \in n$ to obtain $Y_{i,t} = \sum_{j=1}^{N} Y_{j,t}\Xi_{j,t}(d_{ij}P_{ii,t}x_{ij,t-1}^{-\chi_{ij}}/P_{j,t})^{1-\eta}$. Solve the above for $P_{ii,t}^{1-\eta}$, which gives $P_{ii,t}^{1-\eta} = Y_{i,t}/[\sum_{j=1}^{N} Y_{j,t}\Xi_{j,t}(d_{ij}x_{ij,t-1}^{-\chi_{ij}}/P_{j,t})^{1-\eta}] = (Y_{i,t}/Y_t)/[\sum_{j=1}^{N} \theta_{j,t}\Xi_{j,t}(d_{ij}x_{ij,t-1}^{-\chi_{ij}}/P_{j,t})^{1-\eta}]$ or simply $P_{ii,t}^{1-\eta} = \theta_{i,t}\Phi_{i,t}^{\eta-1}$, where $\Phi_{i,t}$ measures the outward multilateral resistance and $\theta_{i,t}$ is the share of the source country income in the world economy. □

C.2 Proposition 2

Proposition. *If the price of imports* $P_{ij,t}$ *from each source country* $i \in n$ *are set for each destination* $j \in n$*, and the production technology of final exported goods exhibits constant returns to scale, then import prices are proportional to the outward multilateral resistance of each source country:*

$$P_{ij,t} = \frac{d_{ij}\theta_{i,t}^{1/(1-\eta)}}{\Phi_{i,t}}.$$

Proof. When the production technology of final exports exhibits constant returns to scale, the unit costs of production $MC_{i,t}$ are independent of the trade flows. And if exporters set prices for each destination, then they are proportional to the unit costs of production, namely $P_{ij,t} = (\eta/(\eta-1))d_{ij}MC_{i,t}$ for all $j \in n \setminus i$, since $d_{ii} = 1$, such that $P_{ii,t} = (\eta/(\eta-1))MC_{i,t}$. It follows that import and export prices are proportional to one another: $P_{ij,t} = d_{ij}P_{ii,t}$. And if so, then using Proposition 1 to substitute out the export price gives rise to an expression for import prices as a function of outward multilateral resistance in the source country: $P_{ij,t} = d_{ij}\theta_{i,t}^{1/(1-\eta)}/\Phi_{i,t}$. □

C.3 Lemma 1

Lemma. *The gravity equation is dynamic when habits are non-zero, such that* $\chi_{ij} > 0$ *for all* $i \in n \setminus j$*. And when habits are asymmetric across countries, such that* $\chi_{ij} \neq \chi_{ji}$ *for all* $i \in n \setminus j$*, and/or the inward and outward the bilateral iceberg costs are non-identical, such that* $d_{ij} \neq d_{ji} > 1$

for all $i \in n \setminus j$, the gravity equation is subject to the multilateral trade imbalance:

$$A_{ij,t} = X_{ij,t} \times \frac{Y_t}{Y_{i,t}Y_{j,t}} = \Xi_{j,t} \left[\frac{d_{ij}^{1+\chi_{ij}}}{\Phi_{i,t}\Phi_{i,t-1}^{\chi_{ij}} P_{j,t}} \right]^{1-\eta} \left[\frac{\theta_{i,t-1}^{-\eta}}{A_{ij,t-1}^{1-\eta} Y_{j,t-1}^{1-\eta}} \right]^{\chi_{ij}} .$$

Proof. When consumers adopt homothetic preferences with constant elasticity of substitution, the demand for imports is given by $X_{ij,t} = C_{j,t}[\tilde{P}_{ij,t}/P_{j,t}]^{1-\eta}$. The aggregate consumption is proportional to aggregate income, such that $C_{j,t} = Y_{j,t}\Xi_{j,t}$, where $\Xi_{j,t} = 1/(1+\xi_{j,t})$ and $\xi_{j,t} = NX_{j,t}/C_{j,t}$, thus $X_{ij,t} = Y_{j,t}\Xi_{j,t}[\tilde{P}_{ij,t}/P_{j,t}]^{1-\eta}$. Using Proposition 1 to substitute out the break-even price index $\tilde{P}_{ij,t} = P_{ij,t}x_{ij,t-1}^{-\chi_{ij}}$ gives $X_{ij,t} = \Xi_{j,t}\left[d_{ij}x_{ij,t-1}^{-\chi_{ij}}/(\Phi_{i,t}P_{j,t})\right]^{1-\eta} Y_{i,t}Y_{j,t}/Y_t$. Next, Proposition 2 is used to substitute out the real stock of habits $x_{ij,t-1}$ with the nominal value trade flows given by $X_{ij,t-1} = P_{ij,t-1}x_{ij,t-1}$, such that $x_{ij,t-1} = X_{ij,t-1}\Phi_{i,t}/(d_{ij}\theta_{i,t}^{1/(1-\eta)})$, which implies that $X_{ij,t} = \Xi_{j,t}\left[d_{ij}^{1+\chi_{ij}}/(\Phi_{i,t}\Phi_{i,t-1}^{\chi_{ij}}P_{j,t})\right]^{1-\eta}\left[\theta_{i,t-1}/X_{ij,t-1}^{1-\eta}\right]^{\chi_{ij}} Y_{i,t}Y_{j,t}/Y_t$. Finally, let $A_{ij,t}$ denote the size-adjusted bilateral trade flows, such that $A_{ij,t} = X_{ij,t}Y_t/(Y_{i,t}Y_{j,t})$. Then it follows that $A_{ij,t} = \Xi_{j,t}\left[d_{ij}^{1+\chi_{ij}}/(\Phi_{i,t}\Phi_{i,t-1}^{\chi_{ij}}P_{j,t})\right]^{1-\eta}\left[\theta_{i,t-1}^{-\eta}/(A_{ij,t-1}^{1-\eta}Y_{j,t-1}^{1-\eta})\right]^{\chi_{ij}}$ (see Appendix B). □

D Empirical Models and Results

D.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 (3.24), 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 habit-augmented 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 supply-habit-augmented gravity equation (3.24) 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 supply-habit-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 (3.19)). Our empirical results provide additional support for this hypothesis, which we discuss below in Section 4.3. 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 supply-habit-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 (3.22)). 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 supply-habit-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 χ_{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.[19] 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 supply-habit-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 $\text{TB}_{j,t}$ 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 (3.22)).

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 parameter heterogeneity, but incorporates the unobservable dynamic multilateral resistance terms.[20]

[19] 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 4.2 and Appendix D.2.

[20] The CCEP estimator is also biased, unlike CCEMG, unless the cross-sectional dimension dominates time peri-

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 4.2 for technical details and Appendix D for results). We demonstrate that these alternative methods can be nested within the same empirical model as outlined in the next Section. 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. As a result, we cannot test our theory against simpler alternatives. Since the validity of the PPML estimator with disregarded zero past trade flow values is not well established in the literature, we stick to dynamic (log)linear models.[21]

D.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.[22] 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 estimates also reflect the fact that the unobservable common factor exerts a heterogeneous influence

ods (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.

[21] We discuss dynamic extension of the PPML and related issues in more detail in Appendix D.3.

[22] 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.

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 supply-habit-augmented gravity equation.

Despite the flexibility of the AMG estimator, there are two reasons why our preferred approach of estimating the supply-habit-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 supply-habit-augmented gravity equation in equation (3.22) 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 supply-habit-augmented gravity equation as additional regressors through the vector of unobservable common factors.

D.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 supply-habit-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 in Appendix D, 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.[23] 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 (4.1)-(4.3) 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).

[23]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).

D.4 Other Estimators

Table 5: Supply-Habit-Augmented Gravity Equation (Other Common Methodologies)

VARIABLES	AMG (1) $FLOW_{ij,t}$	FE2 (2) $FLOW_{ij,t}$	FE3 (3) $FLOW_{ij,t}$	FE4 (4) $FLOW_{ij,t}$	PPML (5) $FLOW_{ij,t}$	PPML2 (6) $FLOW_{ij,t}$	PPML3 (7) $FLOW_{ij,t}$	PPML4 (8) $FLOW_{ij,t}$
$FLOW_{ij,t-1}$	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)
$TB_{j,t}$	0.839***		0.418***		0.166***		0.320***	
	(0.0980)		(0.0334)		(0.0404)		(0.0390)	
$GDP_{i,t-1}$	-0.197***		-0.0198**		-0.00798		-0.0348***	
	(0.0298)		(0.00963)		(0.0123)		(0.0114)	
$GDP_{j,t-1}$	-0.0234		-0.0529***		-0.00602		-0.0759***	
	(0.0325)		(0.00935)		(0.0158)		(0.0153)	
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; $FLOW_{ij,t}$ measured in levels and only non-zero lagged trade flows are retained in all PPML specifications.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

D.5 Econometric Challenges of the Supply-Habit-Augmented Gravity Model

There are two major issues related to the empirical model specification in equation (3.22). First, the theory of habits in the supply chains predicts a homogeneous trade imbalance coefficient equal to unity, whereas the other coefficients in the supply-habit-augmented gravity equation are heterogeneous. Second, unlike all other regressors, the multilateral trade imbalance $\Xi_{j,t}$ enters the supply-habit-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 supply-habit-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 habits in the supply chains. Such transformation delivers the following model, expressed in terms of pre-determined observables:

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

Ultimately, Sections 4.3, 5.2, and 5.4 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 (D.1), all parameters inherit bilaterally varying habits. One of the major differences, compared to the baseline model in (3.22), 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 habits in the supply chains predicts a homogeneous trade imbalance coefficient equal to unity for each country pair. Tables 6 and 7 in Appendix D 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 6) or when we re-specify the size-adjusted trade flows in terms of consumption expenditure in the destination country (see Table 7), 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 supply-habit-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 6: Coefficient Estimates (Supply-Habit-Augmented Gravity Equation Without Trade Imbalance)

VARIABLES	CCEMG (1) $FLOW_{ij,t}$	FE (2) $FLOW_{ij,t}$	MG (3) $FLOW_{ij,t}$	CCEP (4) $FLOW_{ij,t}$	HFE (5) $FLOW_{ij,t}$	AMG (6) $FLOW_{ij,t}$	CCEMGR (7) $FLOW_{ij,t}$
$FLOW_{ij,t-1}$	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)
$TB_{j,t}$							
$GDP_{i,t-1}$	-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)
$GDP_{j,t-1}$	-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)
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 7: Coefficient Estimates (Re-Specified Supply-Habit-Augmented Gravity Equation)

	CCEMG	FE	MG	CCEP	HFE	AMG	CCEMGR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	$\text{XFLOW}_{ij,t}$	$\text{XFLOW}_{ij,t}$	$\text{XFLOW}_{ij,t}$	$\text{XFLOW}_{ij,t}$	$\text{XFLOW}_{ij,t}$	$\text{XFLOW}_{ij,t}$	$\text{XFLOW}_{ij,t}$
$\text{XFLOW}_{ij,t-1}$	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)
$\text{TB}_{j,t-1}$	-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)
$\text{GDP}_{i,t-1}$	-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)
$\text{GDP}_{j,t-1}$	-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; $\text{XFLOW}_{ij,t}$ 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 3.4 and 4.3. The estimated model refers to the equation (D.1).

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

D.6 Prediction "Horse Race"

Table 8: 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 habit-augmented 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 9: 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 habit-augmented 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.

D.7 Parameter Heterogeneity

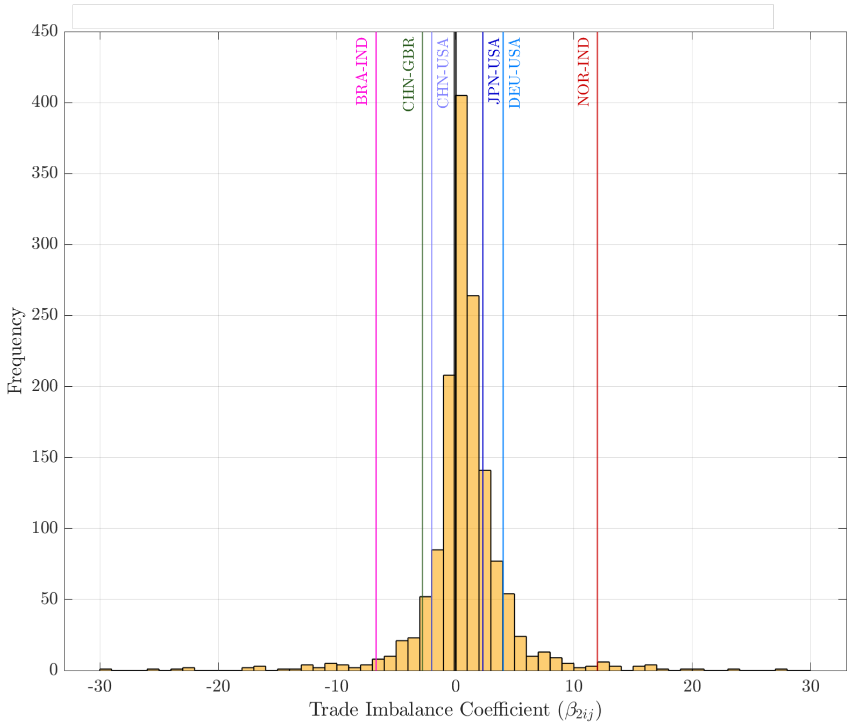

Figure 7: Distribution of Trade Imbalance Coefficient Estimates

Note: The figure presents CCEMG estimates of the trade imbalance coefficient (β_{2ij}) in the dynamic gravity model presented in equations (4.1)-(4.3). 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.

Table 10: The Non-Linear Cross-Sectional Poisson Regression Model of Trade Persistence and Global Value Chains (Re-Specified Supply-Habit-Augmented Gravity Equation)

Variables	$t_{\beta_{1ij}^{\text{restr}}} > 1.96$ (1) $\beta_{1ij}^{\text{restr}} > 0$	$t_{\beta_{1ij}^{\text{restr}}} > 2.326$ (2) $\beta_{1ij}^{\text{restr}} > 0$	$t_{\beta_{1ij}^{\text{restr}}} > 2.576$ (3) $\beta_{1ij}^{\text{restr}} > 0$	$t_{\beta_{1ij}^{\text{restr}}} > 3.09$ (4) $\beta_{1ij}^{\text{restr}} > 0$
Log-Supply-Habit	0.0127	0.0176	0.0224*	0.0252**
	(0.0136)	(0.0129)	(0.0126)	(0.0116)
Log-Distance	-0.0624***	-0.0606***	-0.0587***	-0.0457***
	(0.0138)	(0.0129)	(0.0123)	(0.0120)
Common Language	0.0416	0.0318	0.0247	0.0339
	(0.0297)	(0.0286)	(0.0268)	(0.0266)
Colony	0.243	0.202	0.120	0.0913
	(0.148)	(0.130)	(0.116)	(0.117)
Constant	0.267	0.327	0.389*	0.373*
	(0.242)	(0.227)	(0.215)	(0.198)
Number of Observations	942	870	801	678

Notes: Robust standard errors associated with the Huber/White/sandwich coefficient estimates are in parentheses. All regression models incorporate fixed effects specific to each source and destination country. *** p < 0.01, ** p < 0.05, * p < 0.1.