Global Knowledge and Trade Flows: Theory and Measurement*

Nelson Lind
Emory University

Natalia Ramondo
Boston University & NBER

February 6, 2024

Abstract

We present a model of trade, global innovation, and diffusion, inspired by Eaton and Kortum (1999). The specific structure for innovation and diffusion we propose, which leverages general results developed in our previous work (Lind and Ramondo, 2023a), allows us to measure the flow of ideas across countries and over time. By deriving tractable expressions for productivity and expenditure, we can use easily-available international trade data to estimate both innovation and diffusion rates across countries and over time. We find that, although innovation is correlated with economic growth, there are many high income countries that primarily produce using diffused ideas from foreign sources.

JEL Codes: F1, O3, O4. Key Words: innovation; diffusion; international trade; Poisson processes; Fréchet distribution; generalized extreme value.

*We thank Jonathan Eaton and Sam Kortum for their comments and fruitful discussions. This research is supported by NSF grant #1919372. All errors are our own.

E-mail: nlind@emory.edu; nramondo@bu.edu.

1 Introduction

Global waves of technological change often start from one idea in one place. Edmund Cartwright sparked the industrial revolution by inventing the power loom in the United Kingdom during 1786. Over the next hundred years, first nearby countries, like France, and later far-away places, like India and Argentina, adopted his idea. In 1959, Robert Noyce, Intel's founder, sparked the computer era by creating the first silicon microchip in the United States, which, nowadays, is widely spread across the globe.

Despite the impact of new ideas on day-to-day lives and the economic growth of creators and adopters, it is challenging to measure the accumulation and diffusion of knowledge over space and time. Eaton and Kortum (Eaton and Kortum, 1996a,b, 1997, 1999) provide one of the first attempts to measure the contribution of the creation and spread of ideas to growth using the structure of their model, together with data on where inventors from different countries seek patent protection.[1]

In this paper, we build and estimate a model of Ricardian trade, global innovation, and diffusion, inspired by Eaton and Kortum (1999) (henceforth, EK). This model allows us to use easily-available international trade data to measure knowledge flows across countries and over time.[2] We uncover the dynamics of idea flows by detecting shared knowledge and differentiating growth due to diffusion from growth due to innovation, for twenty countries and regions, over the period 1962-2019.

The model leverages a general result developed in our previous work (Lind and Ramondo, 2023a), which provides necessary and sufficient conditions for innovation and diffusion to generate the class of productivity distributions used in quantitative models of Ricardian trade. Concretely, we assume that ideas get discovered randomly over time, have a unique discovery time and location, and have a global time-invariant quality. Importantly, ideas have a random efficiency component specific to a location and that can change over time — so that when ideas diffuse, they can be more or less similar to the idea in the discovery place. The novelty relative to

[1] Some efforts on measuring knowledge relate the creation — and spread — of ideas to patent creation and citations (e.g. Keller, 2010; Akcigit, Grigsby, and Nicholas, Akcigit et al.); others efforts construct direct measures of technology adoption (e.g. Comin and Hobijn, 2004); and recent efforts use text analysis (e.g. Bloom et al., 2021).

[2] Krugman (1979) is one of the first papers to combine technology diffusion and trade in the context of a North-South model of the product cycle.

our previous work lies on the development of a specific model of the diffusion of ideas, or in other words, a model of the location-time specific component of ideas. The structure for diffusion that we propose preserves some of the key features of EK, and allows us to bring the model to the data.

Diffusion occurs in the model by assuming that once an idea gets discovered, applications of the idea become available in all countries gradually, and they randomly accumulate over time. While ideas represent general insights and broad principles of production, applications represent specific implementations of an idea to a location. That is, each application is a location-specific production technology. In this way, while ideas are global, applications can only be used in a single place.[3]

Even though our approach, based on applications, captures the spirit of the EK model, we depart from their approach in two ways, allowing us to bring the model to the trade data.

First, in the EK setup, diffusion occurs with a random lag, implying that only a subset of countries has knowledge of the idea at each moment in time. To calculate the global distribution of knowledge and obtain implications for trade flows, it is necessary to keep track of the sets of "who knows what" over time. With a large number of countries, this combinatorial problem becomes intractable. We overcome this challenge by modeling diffusion as a gradual process where all countries gain access to *some* applications immediately after an idea gets discovered. Put simply, we depart from modeling diffusion at the extensive margin (i.e. a country has or does not have an idea), and instead, we model the intensive margin through the steady accumulation of applications over time.

Second, we assume that applications are location specific. This assumption buys us tractability when modeling Ricardian trade because it introduces idiosyncratic differences in how countries can use the same idea. Without this assumption, countries have equal ability to use ideas after they diffuse, which can lead to knife-edge equilibrium conditions under head-to-head Ricardian competition. The introduction of location-specific applications of an idea "smooths" overs the knife-edge cases, an approach discussed in Eaton and Kortum (2012), and leads to expressions

[3] For example, artificial intelligence (AI) techniques are general purpose ideas that get applied throughout the world. Recent uses of AI to develop expert systems (Chat-GPT, ChatSonic, and Github Copilot), natural language processing systems (smart assistants, language translation, and predictive text), and autonomous robots (delivery systems, self-driving vehicles, and industrial automation) correspond to specific applications of those ideas.

for productivity and expenditure that are closed-form and easy to compute.

In turn, productivity in each country is the result of using, at each point in time, the most efficient application of the best idea available to them to produce each good. With this structure for knowledge, we can apply Theorem 1 in Lind and Ramondo (2023a) to derive the resulting joint distribution of productivity across space. This distribution is max-stable multivariate Fréchet, which, under Ricardian trade, leads to expenditure shares belonging to the generalized extreme value (GEV) class.[4] As a consequence, we can analyze how innovation and diffusion shape observed patterns of trade across countries and over time.

We use bilateral trade data over time, together with measures of prices, income, and standard geography data, to estimate the parameters of the model. In particular, we estimate diffusion rates, innovation rates, and trade costs, across countries and over time. We do not need to use data on patent creation or citation, multinational firms, or any other data related to R&D efforts. In fact, these measures for innovation and knowledge diffusion can be used to validate our estimates — we do so using data on multinational production and the number of researchers across countries over time.

Since, absent diffusion, the model implies constant elasticity of substitution (CES) expenditure, accounting for diffusion boils down to estimating a non-CES import demand system, as many other papers in the trade literature do.[5] But, by linking substitution expenditure patterns to a specific structure for innovation and diffusion, we are able to use changes in expenditure across countries and time to estimate primitive knowledge parameters. Intuitively, when countries share ideas, head-to-head competition becomes fiercer, resulting in more substitution in expenditure patterns. The opposite happens if ideas are not shared much across countries. In this way, observed trade patterns are informative about underlying global knowledge dynamics, and, ultimately, they help to analyze the determinants of growth. We operationalize this insight for estimation purposes by developing a specific model

[4]Under max-stability, the maximum of Fréchet is also Fréchet, and the conditional and unconditional distribution of the maximum coincides. As first presented by Eaton and Kortum (2002), an independent Fréchet distribution for productivity is max-stable and leads to the subclass of constant elasticity of substitution (CES) expenditure shares. Lind and Ramondo (2023b) show that max-stable Fréchet productivity with arbitrary correlation over space leads to GEV expenditure shares.

[5]See among others Caron et al. (2014), Lashkari and Mestieri (2016), Brooks and Pujolas (2019), Feenstra et al. (2018), Adao et al. (2017), Bas et al. (2017), and our own previous work Lind and Ramondo (2023b).

of innovation and diffusion based on EK.

Our estimates suggest that over the last sixty years countries have become more distinct in terms of knowledge. Particularly, China has surged as a source of innovated knowledge — the estimated model correctly detects the Chinese innovation wave starting in the late 1990's and early 2000's.[6] Interestingly, the trade data uncover a pattern of global knowledge where most rich (European) countries have a large share of knowledge coming from diffusion. Additionally, our estimates of innovated knowledge correlate with increases in researchers and researchers per capita in a country over time. International diffusion also correlates with more researchers over time suggesting that countries need some absorptive capacity to adopt foreign knowledge, as in Nelson and Phelps (1966). Finally, our estimates of international diffusion capture the cross-country patterns of multinational production, undoubtedly one of the main channels of technology transfers across countries.

2 Model

The global economy consists of N countries. Countries produce and trade a continuum of goods $v \in [0,1]$, which aggregate into a composite good using a constant elasticity of substitution (CES) technology with elasticity $\eta > 1$.

Time is continuous and indexed by t. Countries are indexed by n when they are the innovator and m when they are the adopter, as well as o when they are the origin location where goods are produced, and d when they are the destination market.

Each good v is produced under perfect competition with productivity $Z_o(v,t)$ using an input bundle with marginal cost $C_o(t)$. Productivity $Z_o(v,t)$ results from the adoption of ideas. We next describe how the innovation and diffusion of ideas determines productivity over space and time.

2.1 Innovation and diffusion

As in EK, for each good v, there exists an infinite, but countable, set of ideas $i = 1, 2, \ldots$. Each idea has quality Q_i, which is global. Idea i applied to the production of good v gets discovered in a unique country at a unique moment in time. We

[6] See, e.g., Chen et al. (2021) and Ma (2023) for explanations of the rise of innovation in China linked to R&D policies and human capital.

denote the innovator country where the idea is discovered by n_i and its discovery time by t_i^*. All idea-specific variables — such as n_i and t_i^* — are good specific, but to simplify the notation we suppress the index v unless necessary.

Formally, ideas get discovered according to a Poisson process. The intensity of innovation for ideas of quality q is given by $\theta q^{-\theta-1}\lambda_n(t^*)\mathrm{d}q\mathrm{d}t^*$, where $\theta > 0$, $\lambda_n(t^*)$ controls the arrival rate of ideas in n over time, and $\theta q^{-\theta-1}$ controls the arrival rate of low versus high quality ideas. The Poisson assumption implies that the expected number of ideas discovered in n up to time t with quality above q is given by $q^{-\theta}\Lambda_n(t)$, where $\Lambda_n(t) \equiv \int_{-\infty}^{t} \lambda_n(t^*)\mathrm{d}t^*$. The distribution of quality among those ideas is Pareto with lower bound q and tail parameter θ. A high θ means that ideas have very similar quality.

We model diffusion through the concept of *applications* of an idea. While ideas represent general insights and broad principles of production, applications represent specific implementations of an idea to a location. That is, each application is a location-specific production technology.

When an idea is discovered in country n at time t^*, we assume that country n gains access to all of their applications, while other countries start with none. Next, the idea begins to diffuse internationally with other countries gaining access to applications over time.

Applications of idea i arrive to different countries, at different times, and with different efficiencies. We denote the country with the j'th application of the idea by m_{ij}. This country develops application j with efficiency A_{ij} at time t_{ij}. Most applications have efficiencies close to zero.

Formally, we assume that given the quality Q_i, innovator n_i, and discovery time t_i^*, of idea i, the history of all its applications, $\{A_{ij}, m_{ij}, t_{ij}\}_{j=1,2,\ldots}$, consists of the points of a Poisson process with intensity $\Gamma(1-\theta/\sigma)^{-\sigma/\theta}\sigma a^{-\sigma-1}p'_{nm}(t-t^*)\mathrm{d}a\mathrm{d}t$. Here, $\sigma > \theta$ and $p_{nm}(t-t^*) : [0,\infty) \to [0,1]$ is an increasing differentiable function with $p_{nm}(0) = \mathbf{1}\{n = m\}$, $p'_{nm}(0) > 0$ for $n \neq m$, and $\lim_{t-t^*\to\infty} p_{nm}(t-t^*) = 1$. Mimicking the Poisson process for innovation, $\sigma a^{-\sigma-1}\mathrm{d}a$ controls the arrival rate of low versus high efficiency applications, while $p'_{nm}(t-t^*)\mathrm{d}t$ controls the arrival rate of applications in m based on ideas innovated in n.

This assumption implies that among ideas discovered by n at time t^*, the expected number of applications with efficiency of at least a in m by time $t \geq t^*$ is

$$\mathbb{E}\left[\sum_{j=1}^{\infty} \mathbf{1}\{A_{ij} > a, m_{ij} = m, t_{ij} \leq t \mid Q_i = q, n_i = n, t_i^* = t^*\}\right] = \Gamma(1-\frac{\theta}{\sigma})^{-\frac{\sigma}{\theta}} a^{-\sigma} p_{nm}(t-t^*).$$

Since $\lim_{t-t^*\to\infty} p_{nm}(t-t^*) = 1$, the term $\Gamma(1-\theta/\sigma)^{-\sigma/\theta} a^{-\sigma}$ is the expected number of applications with efficiency above a to ever be adopted in m. The probability $p_{nm}(t-t^*)$ is then the expected fraction of those applications known to m by time t. The initial condition that $p_{nm}(0) = \mathbf{1}\{n = m\}$ implies that the innovator gains all of its applications at the moment of innovation, while all other countries start with no applications. However, immediately after discovery, they begin to get some applications since $p'_{nm}(0) > 0$, for $n \neq m$.

This specification for diffusion imposes three key restrictions. First, no country has an inherent advantage at using ideas because, on average, all countries have similar ability to use every idea once they have learned all of their applications (as $t - t^* \to \infty$). Second, the expected number of applications does not depend on an idea's quality. This restriction implies that the best application is independent of quality, which is necessary for multivariate Fréchet productivity. Third, the expected number of applications in m depends on an idea's age, $t - t^*$, rather than on t and t^* separately. This restriction means that diffusion patterns are stationary over time, as in EK. Finally, the distribution of applications A_{ij} among all ideas that will ever exist in any country is Pareto with shape σ. The parameter σ controls the thickness of the tail of this Pareto distribution. Since $\sigma > \theta$, the Pareto distribution of applications is thinner than the Pareto distribution of quality. The constant $\Gamma(1-\theta/\sigma)^{-\sigma/\theta}$ scales down the total number of applications for each idea as $\sigma \to \theta$ (this condition will ensure that the productivity distribution remains finite in the limit).

If country m adopts idea i at time t, they use the best application of the idea available to them. The highest efficiency across applications is

$$A^*_{im}(t) \equiv \max_{j=1,2,\ldots} A_{ij}\mathbf{1}\{m_{ij} = m, t_{ij} \leq t\}. \tag{1}$$

This location-specific component combines multiplicatively with the idea's quality to determine the overall efficiency of the idea. The global quality, Q_i, is the common component of the idea's efficiency, while $A^*_{im}(t)$ creates idiosyncratic variation in the idea's efficiency across countries and time.

Since applications are independent of quality, the distribution of $A^*_{im}(t)$ across countries among ideas innovated by n and discovered at time t^* is also independent of quality. We denote this distribution by $M_n(a_1, \ldots, a_N; t - t^*)$ and, in Appendix A.1, we show that it is independent Fréchet,

$$M_n(a_1, \ldots, a_N; t - t^*) = \exp\left[-\Gamma(1 - \frac{\theta}{\sigma})^{-\frac{\sigma}{\theta}} \sum_{m=1}^{N} p_{nm}(t - t^*) a_m^{-\sigma}\right], \tag{2}$$

with shape σ, and scale for country m given by $\Gamma(1 - \frac{\theta}{\sigma})^{-\frac{\sigma}{\theta}} p_{nm}(t - t^*)$.[7]

This structure for knowledge allows us to depart from the setup in EK in the following ways.

First, after an idea's discovery, all countries start to gain some applications (although they may be very inefficient).

Second, we introduce idiosyncratic differences in how countries can use the same idea. There are two sources of heterogeneity: countries can differ in the short run because it takes time to access applications (i.e. application j arrives in m_{ij} at the random time t_{ij}); and they can differ in the long-run due to differences in the efficiency of applications (i.e. randomness in A_{ij}). This heterogeneity means that some high-quality ideas may still be difficult to apply in some locations both in the short run and the long run. This can be interpreted as barriers to adoption as in Parente and Prescott (1994).

Third, in the setup in EK, diffusion occurs with a random lag, so at each moment in time there is a random subset of countries with knowledge of each idea. As a consequence, one has to keep track of the distribution of ideas across all possible subsets of countries. By allowing all countries to start gaining applications immediately, we circumvent this combinatorial problem.[8]

Fourth, in EK, all countries that receive an idea have equal ability to use it —

[7] In Lind and Ramondo (2023a), the variable $A^*_{im}(t)$ was taken as a primitive with any joint distribution across m. Our setup based on applications endogenizes the location-specific efficiency of each idea.

[8] An alternative setup would be to model the extensive margin of diffusion: At the moment of discovery, only the innovator has knowledge of the idea, but over time, the set of countries with knowledge of the idea expands, as in EK, and $A^*_{im}(t)$ turns positive only for those countries. The problem with this extensive-margin setup, which we sketch in Lind and Ramondo (2023a), is that to compute the productivity distribution — and expenditure shares — one must keep track of the expected number of ideas commonly known to any subset of countries. Our formulation of diffusion in which ideas have Poisson applications avoids this high-dimensional combinatorial problem.

a feature that implies that $A^*_{im}(t)$ is degenerate at one and the idea's efficiency is equal to Q_i. This restriction leads to knife-edge equilibrium conditions with Ricardian competition, which makes quantitative analysis of trade patterns challenging. By introducing heterogeneity via applications, we "smooth" over the knife-edge cases.

This model of innovation and diffusion fits the necessary and sufficient conditions developed in Lind and Ramondo (2023a) to generate any max-stable productivity distribution. In the next section, we derive the specific Fréchet distribution arising from countries accumulating applications of ideas.

2.2 The productivity distribution

At each point in time, countries produce each good using the most efficient idea available to them. Hence, the productivity in country m at time t (for production of good v) is

$$Z_m(t) = \max_{i=1,2,\ldots} Q_i A^*_{im}(t), \tag{3}$$

where Q_i is the common component of productivity across countries, and $A^*_{im}(t)$ is the location-specific time-varying component in (1) arising from adopting the best available application of the idea.

When a country accesses a new idea, their productivity rises if the idea is more productive than the idea that they were previously using, either because the newly-available idea has high quality, or because it has a very efficient application available in that location. In this way, the dynamics of productivity are entirely driven by two forces: innovation, which can increase productivity through the introduction of new ideas in their discovery location; and diffusion, which can increase productivity by expanding access to applications of previously discovered ideas.

This structure for knowledge implies that the distribution of productivity across countries at each moment in time is max-stable multivariate Fréchet. To get this result, we apply Theorem 1 in Lind and Ramondo (2023a), which provides a constructive method to derive the productivity distribution. Appendix A.2 shows how to use the theorem with the specific structure of $A^*_{im}(t)$ arising from applications.

The Poisson assumption on the innovation process, together with the result in (2) that the efficiency of the best application available at time t in country m is independent Fréchet, yields the following expression for the joint distribution of productivity across countries,

$$\mathbb{P}\left[Z_1(t) \leq z_1, \ldots, Z_N(t) \leq z_N\right] = \exp\left[-\sum_{n=1}^{N} \int_{-\infty}^{t} \left(\sum_{m=1}^{N} p_{nm}(t-t^*) z_m^{-\frac{\theta}{1-\rho}}\right)^{1-\rho} \lambda_n(t^*) \mathrm{d}t^*\right]. \tag{4}$$

Here, each nest n corresponds to an innovator, and $\rho \equiv 1 - \theta/\sigma$ is the correlation coefficient. When $\sigma \to \theta$ so that quality and applications are equally fat-tailed, correlation is zero. As σ increases, similarity in applications across countries for ideas from a given innovator increases, creating correlation in productivity.

We obtain the marginal distribution of productivity in m by letting $z_{m'} \to \infty$ in (4) for all $m' \neq m$. This marginal distribution is Fréchet, $\mathbb{P}\left[Z_m(t) \leq z_m\right] = \exp\left[-T_m(t) z_m^{-\theta}\right]$, with shape $\theta > 0$ and scale $T_m(t) \equiv \sum_{n=1}^{N} T_{nm}(t)$, where

$$T_{nm}(t) \equiv \int_{-\infty}^{t} p_{nm}(t-t^*)^{1-\rho} \lambda_n(t^*) \mathrm{d}t^*. \tag{5}$$

We refer to this variable as shared knowledge from n to m.

When there is no international diffusion of ideas, $p_{nm}(t-t^*) = \mathbf{1}\{n = m, t \geq t^*\}$. Then, the expression in (4) collapses to

$$\mathbb{P}\left[Z_1(t) \leq z_1, \ldots, Z_N(t) \leq z_N\right] = \exp\left[-\sum_{n=1}^{N} T_n(t) z_n^{-\theta}\right], \tag{6}$$

where $T_n(t) = \Lambda_n(t)$. This special case corresponds to the case of independent Fréchet productivity, with the scale $\Lambda_n(t)$ reflecting the stock of knowledge innovated in n up to time t. Without international diffusion, productivity just reflects own innovated ideas, which are always the ones with highest efficiency. When there is international diffusion, countries can access ideas that they have not innovated themselves and that may have higher efficiency than their own ideas. The sharing probabilities $p_{nm}(t-t^*)$ (for $n \neq m$) capture how diffusion generates departures from the case of independence.

In the limiting case of $\sigma \to \infty$, the distribution of $A^*_{im}(t)$ becomes degenerate at one and, as in EK, each idea has identical efficiency across countries. This limiting case corresponds to a multivariate Fréchet distribution with perfect correlation in productivity across countries. However, in this case, the productivity distribution in (4) is not continuously differentiable. Without continuous differentiability, we cannot apply the results in Lind and Ramondo (2023b) that map max-stable Fréchet productivity to GEV expenditure shares. But for $\sigma < \infty$, applications of an idea introduce idiosyncratic differences in the efficiency of the idea across countries, break the case of perfect correlation, and allow us to obtain closed-form solutions for expenditure shares.

Finally, the productivity distribution obtained in (4) can be interpreted in the context of a model of multinational production, as in Ramondo and Rodríguez-Clare (2013), linking the efficiency losses associated with transferring productivity for production from n to m to the diffusion and adaptation of ideas discovered in n to the host market m — i.e. the variable $p_{nm}(t - t^*)$.[9] We make a closer connection with multinational production as a channel for diffusion in Section 4.

2.3 Expenditure shares

We next present the results for prices and expenditure shares across countries. We assume that trade is subject to iceberg-type trade costs, $\tau_{od}(t)$ with $\tau_{oo}(t) = 1$. The unit cost of good v in country d at time t when sourced from country o is $P_{od}(t)/Z_o(t,v)$, where $P_{od}(t) \equiv \tau_{od}(t)C_o(t)$ is the import cost index. Each destination sources good v from the origin with the lowest unit cost so that the price of the good in destination d at time t in given by

$$P_d(t,v) = \min_{o=1,\dots,N} \frac{P_{od}(t)}{Z_o(t,v)}. \tag{7}$$

Since productivity is max-stable Fréchet, aggregate expenditure shares have a closed-form solution. Assuming that $1 + \theta - \eta > 0$, the expenditure share by country d on goods from o at time t is

$$\pi_{od}(t) = \sum_{n=1}^{N} \pi_{nod}(t), \tag{8}$$

where the sum is over the innovator countries n, and

$$\pi_{nod}(t) \equiv \int_{-\infty}^{t} \pi^{\mathrm{W}}_{nod}(t^*,t)\pi^{\mathrm{B}}_{nd}(t^*,t)\mathrm{d}t^*, \tag{9}$$

[9] Cai and Xiang (2022) present a model of innovation and diffusion through multinational production where productivity is max-stable multivariate Fréchet (with symmetric correlation) and is derived using Poisson processes. Their setup is a nice extension of the structure in Buera and Oberfield (2020) and falls within the class of knowledge structures analyzed in Lind and Ramondo (2023a).

is the share of expenditure by d for goods produced in o with ideas from n. This trilateral share $\pi_{nod}(t)$ captures the impact of innovation by n and diffusion to o through two components: the within-expenditure share, $\pi^{\mathrm{W}}_{nod}(t^*, t)$, and the between-expenditure share, $\pi^{\mathrm{B}}_{nd}(t^*, t)$.

First, the expenditure by d in goods produced in o using ideas discovered in n at time t^*, relative to all other sources m that use ideas discovered in n at time t^*,

$$\pi^{\mathrm{W}}_{nod}(t^*, t) \equiv \frac{p_{no}(t-t^*)P_{od}(t)^{-\frac{\theta}{1-\rho}}}{\sum_{m=1}^{N} p_{nm}(t-t^*)P_{md}(t)^{-\frac{\theta}{1-\rho}}}, \tag{10}$$

reflects diffusion from n to o. In particular, this within-innovator share is proportional to the probability of ideas from n reaching o. Since $p_{nn}(t-t^*) = 1$,

$$\frac{\pi^{\mathrm{W}}_{nod}(t^*, t)}{\pi^{\mathrm{W}}_{nnd}(t^*, t)} = p_{no}(t-t^*)\left(\frac{P_{od}(t)}{P_{nd}(t)}\right)^{-\frac{\theta}{1-\rho}}.$$

Market d's expenditure on goods from o produced using ideas innovated by n at time t^* relative to their expenditure on n reflects how rapidly those ideas diffused from n to o between t^* and t. The within self-share $\pi^{\mathrm{W}}_{nnd}(t^*, t)$ reflects overall diffusion at time t of ideas innovated by n at time t^* to the rest of the world.

Second, the expenditure by d at time t on goods made using ideas that were discovered previously in n at time t^*,

$$\pi^{\mathrm{B}}_{nd}(t^*, t) \equiv \left(\sum_{m=1}^{N} p_{nm}(t-t^*)\left(\frac{P_{md}(t)}{P_d(t)}\right)^{-\frac{\theta}{1-\rho}}\right)^{1-\rho}\lambda_n(t^*), \tag{11}$$

reflects innovation by n at time t^* as well as overall diffusion from n to the rest of the world between t^* and t. Here, the price index is[10]

$$P_d(t) = \left[\sum_{n'=1}^{N}\int_{-\infty}^{t}\left(\sum_{m=1}^{N} p_{n'm}(t-t^*)P_{md}(t)^{-\frac{\theta}{1-\rho}}\right)^{1-\rho}\lambda_{n'}(t^*)\mathrm{d}t^*\right]^{-\frac{1}{\theta}}. \tag{12}$$

By focusing on new ideas discovered at $t^* = t$, we can isolate the effect of innovation

[10]To simplify notation throughout the paper, we impose a normalization and multiply the CES composite good by $\Gamma(\frac{1+\theta-\eta}{\theta})^{\frac{1}{\eta-1}}$.

taking the ratio of the between-shares for any two innovators,

$$\frac{\pi^{\mathrm{B}}_{nd}(t,t)}{\pi^{\mathrm{B}}_{n'd}(t,t)} = \left(\frac{P_{nd}(t)}{P_{n'd}(t)}\right)^{-\theta} \frac{\lambda_n(t)}{\lambda_{n'}(t)}.$$

The intuition for this result is simply that at the moment of innovation all applications are in the innovating country so that other countries m do not matter. Relative expenditure on new ideas from n reflects their innovation rate.

The expenditure share in (8) belongs to the GEV class, a class that generates rich patterns of substitution across countries and includes CES expenditure. In particular, when ideas cannot diffuse across countries, so that $p_{nm}(t-t^*) = 0$ for $m \neq n$, the productivity distribution reduces to (6) and (8) collapses to the familiar CES expression (e.g. Eaton and Kortum, 2002),

$$\pi_{od}(t) = \frac{\Lambda_o(t) P_{od}(t)^{-\theta}}{\sum_{o'=1}^{N} \Lambda_{o'}(t) P_{o'd}(t)^{-\theta}}, \tag{13}$$

where the denominator is related to the CES price index, $P_d(t)^{-\theta}$. CES arises in this case because the within-expenditure shares are zero except for $n = o$.

Once we add some diffusion, within-expenditure shares are no longer zero and expenditure becomes non-CES. Formally, we can calculate the cross-price elasticity of substitution for destination market d between sources o and o' at time t (see Appendix A.2 for derivations),

$$\varepsilon_{oo'd}(t) \equiv \frac{\partial \ln \pi_{od}(t)}{\partial \ln P_{o'd}(t)/P_d(t)} = \frac{\theta\rho}{1-\rho} \sum_{n=1}^{N} \int_{-\infty}^{t} \pi^{\mathrm{W}}_{nod}(t^*,t)\pi^{\mathrm{W}}_{no'd}(t^*,t) \frac{\pi^{\mathrm{B}}_{nd}(t^*,t)}{\pi_{od}(t)} \mathrm{d}t^*, \tag{14}$$

and the own price elasticity is

$$\varepsilon_{ood}(t) = -\frac{\theta}{1-\rho}\left[1 - \rho \sum_{n=1}^{N} \int_{-\infty}^{t} \pi^{\mathrm{W}}_{nod}(t^*,t)\pi^{\mathrm{W}}_{nod}(t^*,t) \frac{\pi^{\mathrm{B}}_{nd}(t^*,t)}{\pi_{od}(t)} \mathrm{d}t^*\right], \tag{15}$$

where $\sum_{o'} \varepsilon_{oo'd}(t) = -\theta$.

The expression in (14) has a simple intuition: two sources o and o' are closer substitutes when their within-expenditure is more similar, i.e. higher $\pi^{\mathrm{W}}_{nod}(t^*,t)\pi^{\mathrm{W}}_{no'd}(t^*,t)$, or correlation in productivity is higher, i.e. higher $\frac{\theta\rho}{1-\rho} \equiv \sigma - \theta$ (similar applications in terms of efficiency relative to similarity in the quality of ideas). In turn, within-expenditure similarity is directly linked to the similarity of the probability of sharing ideas coming from an innovator n, $p_{no}(t-t^*)p_{no'}(t-t^*)$, and from similarity in unit costs, $P_{od}(t)P_{o'd}(t)$. Both channels make the two sources fiercer head-to-head competitors. If there is no cross-country diffusion, or $\rho = 0$, then $\varepsilon_{oo'd}(t) = 0$ for $o \neq o'$, $\varepsilon_{ood}(t) = -\theta$, and expenditure is CES as in (13).[11]

To uncover the structure of (8) using trade data requires imposing functional forms to the import cost index, $P_{od}(t)$, the diffusion probability, $p_{nm}(t-t^*)$, and the innovation rate, $\lambda_n(t)$. We present our assumptions in Section 3 after presenting the model's equilibrium.

2.4 Equilibrium

At each moment in time, country o is populated by an exogenous measure of households $L_o(t)$ that inelastically supply labor for production in their country, and consume a non-tradable final good.

Following Alvarez and Lucas (2007), the final good in market o is produced with labor and the composite CES aggregate over traded varieties. The final good technology is Cobb-Douglas with labor share of $1-\beta \in (0,1)$. Hence, letting $W_o(t)$ denote the wage in o, the price of the final good is

$$P_o^f(t) = W_o(t)^{1-\beta} P_o(t)^{\beta}, \tag{16}$$

where $P_o(t)$ is the price of the intermediate input bundle given by (12).

In turn, each intermediate good v is produced using labor and the composite intermediate input bundle using a Cobb-Douglas production function with labor share $1-\alpha \in (0,1)$. The cost of the input bundle for each good v is then

$$C_o(t) = W_o(t)^{1-\alpha} P_o(t)^{\alpha}. \tag{17}$$

The market clearing condition for the final good market in country o implies that

$$X_o^f(t) = Y_o(t) - \xi_o(t),$$

[11] Our approach connects to Berry et al. (1995) in the IO literature, and more recently, Adao et al. (2017) in the trade literature, in that we allow for heterogenous cross-price expenditure elasticities and estimation techniques that use only aggregate data.

where $Y_o(t) = W_o(t)L_o(t)$ denotes total income in o, $X_o^f(t)$ is total expenditure on the final good, and $\xi_o(t)$ is an exogenous trade imbalance with $\sum_{j=1}^N \xi_j(t) = 0$.

Expenditure on the aggregate input in country o by final goods producers is $\beta X_o^f(t)$, while expenditure by input producers is $\alpha \sum_{d=1}^N \pi_{od}(t)X_d(t)$, so that total expenditure on traded goods in country o is $X_o(t) = \beta X_o^f(t) + \alpha \sum_{d=1}^N \pi_{od}(t)X_d(t)$. Finally, labor market clearing implies that

$$W_o(t)L_o(t) = (1-\beta)X_o^f(t) + (1-\alpha)\sum_{d=1}^N \pi_{od}(t)X_d(t).$$

For our estimation in Section 3, we invert the equilibrium conditions of the model to infer each country's expenditure on domestic production of traded inputs.

3 Estimation Procedure

We use data from the Penn World Table 10.0 (PWT) on current GDP (cgdpo), employment (emp), the value of imports and exports, as well as the price and the value of domestic absorption (pl_da and cda), for each country, from 1962 to 2019. Additionally, we use data on trade flows between countries, for the same time period, from COMTRADE. We construct self-trade shares for each country using the model equilibrium conditions in Section 2.4. Our sample contains nineteen countries and regions plus an aggregate of the rest of the world. Appendix Table B.1 reports the sample of countries.

We use estimates from the literature to calibrate the shape parameter σ. Inspecting (8), together with (9) and (10), reveals that $\sigma \equiv \theta/(1-\rho)$ plays the role of the trade elasticity in standard models of trade, i.e. the elasticity of trade shares to real import prices. Hence, we set $\sigma = 2$, as estimated by Boehm et al. (2023) from time-series trade data and most-favored-nation tariffs.

We estimate the correlation parameter ρ together with the remaining parameters of the model, as explained below. This parameter is identified from the size of departures from CES expenditure; as shown in (14), the magnitudes of all non-zero cross price elasticities are proportional to $\rho/(1-\rho)$. For the case of no diffusion, since expenditure is CES, we set $\rho = 0$ so that $\theta = \sigma = 2$.

We next present our functional form assumptions for the import cost index, diffusion rates, and innovation rates.

Import cost index. The import cost index is given by $P_{od}(t) \equiv \tau_{od}(t) C_o(t)$ where $C_o(t)$ is the cost of producing the input bundle in country o, and $\tau_{od}(t)$ is the trade cost. First, we set $\beta = \alpha = 0.5$ corresponding to the average documented in Jones (2011). Second, we construct the cost index $C_o(t)$ using the expression in (17): For wages $W_o(t)$ we use data on current (PPP-adjusted) GDP per worker, while we calculate the intermediate price index $P_o(t)$ using data on the price of domestic absorption (PPP-adjusted) and the expression in (16). Third, we assume that trade costs are a function of geographical distance, and border costs,

$$\tau_{od}(t) = \mathbf{1}\{o = d\} + \mathbf{1}\{o \neq d\} \exp\left(\kappa_d^\tau(t) + \kappa^\tau \ln \text{Dist}_{od}\right). \tag{18}$$

As in Eaton et al. (2013), the border cost, $\kappa_d^\tau(t)$, plays the role of a destination-time fixed effect and ensures that we match self-trade shares.

Diffusion probabilities. Following EK, we assume exponential diffusion,

$$p_{nm}(t - t^*) = 1 - e^{-\delta_{nm}(t-t^*)}, \tag{19}$$

where the variable δ_{nm} measures the rate at which m learns from n. To get instantaneous diffusion within the innovator, we set $\delta_{nn} = \infty$. We estimate the following cases, for all $m \neq n$:

1. No diffusion, $\delta_{nm} = 0$;
2. Homogenous diffusion, $\delta_{nm} = \delta$;
3. Gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^\delta \ln \text{Dist}_{nm})$; and
4. Fixed-effect (FE) gravity diffusion, $\delta_{nm} = \delta_m^A \delta_n^I \exp(-\kappa^\delta \ln \text{Dist}_{nm})$ with the normalization $\sum_{n=1}^N \delta_n^A = 1$.

Innovation rates. To reduce the dimensionality of the estimation procedure, we impose some assumptions on innovation rates $\lambda_n(t)$. Similarly to EK — and originally introduced in Krugman (1979) — we assume that the arrival of ideas in each country is proportional to the existing stock of ideas. Alone, this assumption would generate exponential growth with $\lambda_n(t) = \Lambda_n'(t) = \gamma_n \Lambda_n(t)$, for some $\gamma_n > 0$.

To capture periods of high innovation, we additionally assume that in each country n there are S surges of innovation. For each surge s, there is a mass λ_n^s of ideas with discovery times distributed normal with mean μ_n^s and standard deviation ν_n^s. This creates, for each country n, ideas of different "vintages." For example, in contrast to the power loom, ideas related to micro-chips would belong to a recent surge. This approach provides some flexibility to help match the trade data, while keeping the parameterization of innovation rates relatively low-dimensional. We restrict the estimation to two surges ($S = 2$), which end up corresponding to ideas created in the past and recent ideas.

The combination of proportional growth together with surges of innovation implies that the intensity of innovation in country n at time t is

$$\lambda_n(t) = \gamma_n \Lambda_n(t) + \sum_{s=1}^{S} \lambda_n^s \frac{1}{\nu_n^s} \phi\left(\frac{t - \mu_n^s}{\nu_n^s}\right), \tag{20}$$

where $\phi(x) \equiv (2\pi)^{-1/2} e^{-x^2/2}$. The first term on the right-hand side of (20) captures the creation of new ideas in proportion to the existing stock of ideas. The second term captures the exogenous discovery of ideas across surges of innovation in country n.

We show in Appendix A.3 that when $\lim_{t \to -\infty} e^{-\gamma_n t} \Lambda_n(t) = 0$, the solution to the differential equation in (20) takes the form $\Lambda_n(t) = \sum_{s=1}^{S} \Lambda_n^s(t)$, where

$$\Lambda_n^s(t) \equiv \lambda_n^s \int_{-\infty}^{t} e^{\gamma_n (t - t^*)} \frac{1}{\nu_n^s} \phi\left(\frac{t^* - \mu_n^s}{\nu_n^s}\right) \mathrm{d}t^* \quad \text{for each } s = 1, \ldots, S \tag{21}$$

captures ideas inspired by each surge of innovation.

We are left to estimate the following parameters: diffusion rates δ_{nm} (under the alternative specifications); innovation-related parameters, γ_n, λ_n^s, μ_n^s, and ν_n^s, for $s = 1, 2$; trade-cost parameters in (18), $\kappa_d^\tau(t)$ and κ^τ; and the correlation parameter ρ. We use a Poisson pseudo maximum-likelihood (PPML) procedure to match the observed bilateral trade shares over time.[12]

For the case of gravity diffusion, which will end up being our preferred specification, we have TN^2 observations to estimate: 2 parameters related to diffusion (δ and κ^δ); $N + 3SN$ related to innovation; $NT + 1$ related to trade costs; and the correlation parameter ρ. The rank restriction in this case is $N^2 T > N + 3SN + NT + 4$, which

[12] To minimize the deviance of the model from the data, we use automatic differentiation in order to compute gradients (Innes, 2018), and the ADAM algorithm (Kingma and Ba, 2014).

with $S = 2$, $N = 20$, and $T = 58$, is satisfied. For the case of no diffusion, we set $\rho = 0$ and estimate innovation and trade-cost parameters only.

The key logic for identification is that if trade shares are non-CES, then we must have international diffusion of ideas because innovation alone leads to CES demand. If international diffusion were absent ($\delta_{nm} = 0$ for $m \neq n$), or $\rho = 0$, all cross-price elasticities would be zero, and the demand system would be CES. As a consequence, any departure from CES —as the literature has documented (e.g. Adao et al., 2017) — can only be attributed by the model to international diffusion. In this case, changes over time in trade shares reflect both innovation and diffusion. Next, we make the argument more formally.

Define the following adjusted trade share,

$$\tilde{\pi}_{od}(t) \equiv \left(\frac{P_{od}(t)}{P_d(t)}\right)^{\theta} \pi_{od}(t). \tag{22}$$

Suppose that there is no diffusion so that expenditure is CES. Using (13), we get that $\tilde{\pi}_{od}(t) = \Lambda_o(t)$, which differentiating over time yields

$$\frac{\partial \tilde{\pi}_{od}(t)}{\partial t} = \lambda_o(t).$$

Hence, under CES, we can recover innovation rates $\lambda_o(t)$ from time-series data on expenditure shares and real import prices, for any destination d. For any cross-section of countries, using adjusted trade shares as in (22) is the standard way to estimate scale parameters $\Lambda_o(t)$ in trade models where import demand is CES.

Turning to the non-CES case, the intuition is analogous except that now we must account for the impact of diffusion. Using (8)-(11), the definition of $\tilde{\pi}_{od}(t)$ in (22), and the assumption on instantaneous domestic diffusion, yields (see Appendix A.4 for derivations)

$$\frac{\partial \tilde{\pi}_{od}(t)}{\partial t} = \lambda_o(t) + \sum_{n=1}^{N} \int_{-\infty}^{t} \omega_{nod}(t^*, t)\lambda_n(t^*)\mathrm{d}t^*. \tag{23}$$

The function $\omega_{nod}(t, t^*)$ captures how ideas innovated in the past in country n impact expenditure on goods produced in o and sold to d today (see Appendix Equation A.7). These weights depend on diffusion parameters (and real import prices that we control for), and are equal to zero when there is no international diffusion.[13]

Overall, the expression in (23) indicates that, after accounting for real import prices, changes in the share of expenditure on goods from o by d that cannot be attributed to ideas innovated today in o, captured by $\lambda_o(t)$, must get attributed to the diffusion of ideas innovated in the past all over the world. Put differently, changes in trade shares on goods from o that differ across destinations d cannot be attributed to $\lambda_o(t)$, which does not vary across d. Rather, they have to be attributed to the non-CES component, which varies across d, and which, through the lens of our model, indicates the presence of diffusion.

In Appendix A.5, we formally show that, for any destination d, given $\tilde{\pi}_{od}(t)$ and the diffusion parameters determining $\omega_{nod}(t^*, t)$, the condition in (23) across exporters o's and times t's admits a unique solution for $\lambda_n(t^*)$ across innovators n's and discovery times t^*'s. Because (23) holds for all destination markets d, with data on bilateral trade shares and real import prices over time, we have many more moments than necessary to pin down innovation rates. This over-identification provides enough room to also estimate diffusion rates.

To further provide intuition on the identification of the diffusion parameters, it is useful to draw a parallel to static models of trade and multinational production (MP). In those models, country n is the source of technologies (i.e. MP), country o is the location of production, and country d is the destination country of sales. Hence, the expenditure devoted to goods produced in o with ideas innovated in n, relative to all other sources of ideas (i.e. the within-share), is informative about the cost to transfer technology from n to o (i.e. the MP costs), while the share of expenditure by d in goods produced everywhere with ideas innovated in n (i.e. the between-share) is informative about n's technologies. Similarly, in our case, within expenditure shares in (10) are informative about $p_{no}(t - t^*)$, and between expenditure shares in (11) are informative about $\lambda_n(t^*)$. There are two important identifying restrictions imposed on models of trade and MP. First, there are no frictions to transfer technology within a country (i.e. the equivalent of instantaneous diffusion within the innovator country), and, second, there are no domestic trade costs.

While models of trade and MP use both observed bilateral trade and MP shares to

[13] For $\rho \to 1$ and σ fixed, it is clear that $\omega_{nod}(t^*, t)$ depends only on diffusion parameters. In that case, $\tilde{\pi}_{od}(t) = \pi_{od}(t)$, and using (A.7), $\omega_{nod}(t^*, t) = \partial \pi^{\mathrm{W}}_{nod}(t^*, t)/\partial t$; within-expenditure shares in (10) only depend on diffusion probabilities (and real import prices).

estimate MP costs, we only observe bilateral trade shares, $\pi_{od}(t) = \sum_{n=1}^{N} \pi_{nod}(t)$ to estimate diffusion. We do not observe the share of output produced in o with ideas innovated in n and destined to d, $\pi_{nod}(t)$, or to all destinations, $\sum_{d=1}^{N} X_d(t)\pi_{nod}(t)/Y_o(t)$ (i.e. the analog of an MP share); effectively, we would need to estimate these unobserved expenditures. Suppose for a moment that in order to estimate the static trade and MP model, we only had data on bilateral trade flows. Would it be possible to estimate MP frictions across countries? The answer is: yes, with enough parametric restrictions on the import cost index (such as gravity) and on MP frictions (such as symmetry). Our estimation procedure has a similar flavor, but it is enriched with dynamics, so that we use trade data both across country pairs *and* over time, as well as the parametric restrictions imposed above.

4 Estimation Results

In this section, we first present and compare results from estimating various diffusion specifications, including the case of no diffusion. Next, we show implied cross-price elasticities and shared knowledge for specifications with diffusion. Additionally, we show the implied evolution of each country's stock of innovated ideas as well as the share of their total knowledge attributable to diffusion. Finally, as a form of validation, we correlate our estimates with data on geography, income, research, and multinational production.

4.1 The Role of Diffusion

Table 1 shows statistics for the estimated model with two innovation surges ($S = 2$), and three different diffusion specifications. In the last four rows of the table, we report results from likelihood ratio tests for correct specification. Relative to no diffusion, which implies CES expenditure, all specifications with international diffusion fit the data better according to our measures of goodness of fit (R-squared and deviance). We reject the null hypothesis that CES is correctly specified relative to the alternative in column (2) with a homogenous international diffusion rate. In turn, we reject this homogenous diffusion model in favor of a gravity model for diffusion in column (3), but we fail to reject (3) in favor of gravity diffusion with fixed effects in (4). These tests lead us to adopt specification (3) as our preferred

Table 1: Estimation results: parameters and goodness of fit.

Models	(1)	(2)	(3)	(4)
Correlation coefficient, ρ	0.0	0.98	0.97	0.896
		(0.94, 0.99)	(0.93, 0.99)	(0.79, 0.96)
Trade distance elasticity κ^τ	0.423	0.933	1.73	1.73
	(0.36, 0.48)	(0.75, 1.36)	(1.46, 2.08)	(0.75, 1.99)
Diffusion distance elasticity κ^δ			1.68	1.36
			(0.13, 2.06)	(0.29, 1.75)
Number of Observations	23,200	23,200	23,200	23,200
Number of Parameters	1,301	1,303	1,304	1,342
R-squared	0.889	0.991	0.992	0.992
Deviance	620	131	116	101
Null hypothesis		(1)	(2)	(3)
χ^2		489	14.9	15.1
Degrees of freedom		2	1	38
p-value		0.00	0.00	0.99

Notes: Results from estimating the model with two innovation surges, $S = 2$, and: (1) no diffusion, $\delta_{nm} = 0$ for all $m \neq n$; (2) homogenous international diffusion, $\delta_{nm} = \delta$ for all $m \neq n$; (3) gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^\delta \ln \text{Dist}_{nm})$ for all $m \neq n$; and (4) FE gravity diffusion, $\delta_{nm} = \delta^A_m \delta^I_n \exp(-\kappa^\delta \ln \text{Dist}_{nm})$ for all $m \neq n$. In all cases, $\delta_{nn} = \infty$. We report 95% profile likelihood confidence intervals in parentheses. The last four rows show results from likelihood ratio tests of the null hypothesis that the model in the previous column is correctly specified against the alternative of that column. To be consistent with the PPML assumption that the mean equals the variance, we adjust the deviance of the model by the mean-variance ratio of observed trade shares.

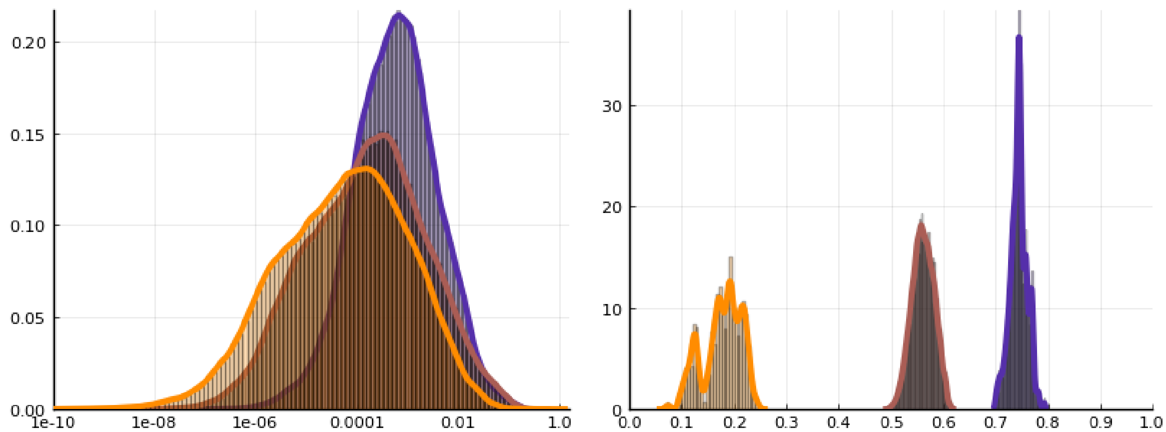
specification: $\delta_{nm} = \delta \exp(-\kappa^\delta \ln \text{Dist}_{nm})$ for all $m \neq n$.

Our estimate of the correlation parameter in (3) is around 0.97, higher than 0.65 estimated by Arkolakis et al. (2018) using multinational-production data from the United States, but much closer to the case in EK where countries have equal ability to use the same idea. If all diffusion happened through the activity of multinational firms, our model would be exactly a model of trade and multinational production. But diffusion in our framework is meant to capture other channels besides diffusion through multinational firms. Hence, it is not surprising that our estimates differ. Appendix Figure B.1 also shows the goodness of fit of the gravity-diffusion model for different values of ρ. Indeed, the best fit is achieved for our estimated value of ρ, suggesting that the data, together with our structure and parametric restrictions, have enough variation to identify this parameter.

Additionally, our estimates yield a trade-cost distance elasticity of around 1.75. This estimate is not comparable with usual gravity estimates of the distance elasticity

Figure 1: Distribution of elasticities and shared knowledge.

(a) Expenditure elasticities (b) Shared knowledge



Notes: Results from estimating the model with two innovation surges, $S = 2$, and: homogenous international diffusion, $\delta_{nm} = \delta$ for all $m \neq n$ (purple); gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$ (brown); and FE gravity diffusion, $\delta_{nm} = \delta_m^A \delta_n^I \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$ (orange). In all cases, $\delta_{nn} = \infty$. Substitution elasticities are calculated using (14), while the fraction of shared knowledge is calculated using (24).

because, precisely, our model does not have a gravity structure, so that the distance elasticity (with respect to trade flows) is not captured by a single number.

Finally, when we include bilateral distance in our specifications of diffusion, we get the expected sign — more distance, less sharing of ideas, in line with findings in Comin et al. (2012) — and a magnitude similar to the trade-cost distance elasticity.

Appendix Table B.1 shows the estimates for country-level variables related to the innovation process, while Appendix Figure B.2 shows graphically the implied estimates of $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$. Finally, Appendix Table B.3 shows the estimates of border costs in (18), by country and over time.

Elasticities and shared knowledge. First, we present some statistics related to our estimates of elasticities and shared knowledge. Figure 1 shows histograms both for bilateral elasticities and probabilities that two countries share knowledge, at each point in time. Each observation corresponds to a pair of competitors o and o' into a given destination market d at time t. Figure 1a shows that there is significant heterogeneity in cross-price elasticities. For some pairs of countries, the elasticity is essentially zero, indicating that they are not strong head-to-head competitors. But for many pairs, the elasticity is large.

Higher expenditure elasticities reflect shared knowledge. In analogy to the expression

Table 2: Estimation results: cross-price elasticities and shared knowledge. OLS.

	$\ln \varepsilon_{oo'd}(t)$				
	(1)	(2)	(3)	(4)	(5)
$\ln \text{Sharing}_{oo'}(t)$	18.739***	35.447***	15.884***	34.717***	35.933***
	(0.110)	(0.115)	(0.118)	(0.130)	(0.129)
Constant	Yes	No	No	No	No
$o \times d$ fixed effects	No	Yes	No	Yes	No
t fixed effects	No	No	Yes	Yes	No
$o \times d \times t$ fixed effects	No	No	No	No	Yes
Observations	396,720	396,720	396,720	396,720	396,720
R-squared	0.070	0.506	0.085	0.510	0.559
Within R-squared	0.070	0.231	0.046	0.185	0.211

Notes: Results from estimating the model with two innovation surges, $S = 2$, and gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$ and $\delta_{nn} = \infty$. The elasticities $\varepsilon_{oo'd}(t)$ are calculated using (14), while $\text{Sharing}_{oo'}(t)$ corresponds to (24). Robust standard errors are in parenthesis with levels of significance denoted by *** p < 0.001, and ** p < 0.01 and * p<0.05.

in (5) for knowledge from innovator n shared with m, $T_{nm}(t)$, we measure the fraction of knowledge from all innovators n shared between o and o' at time t as

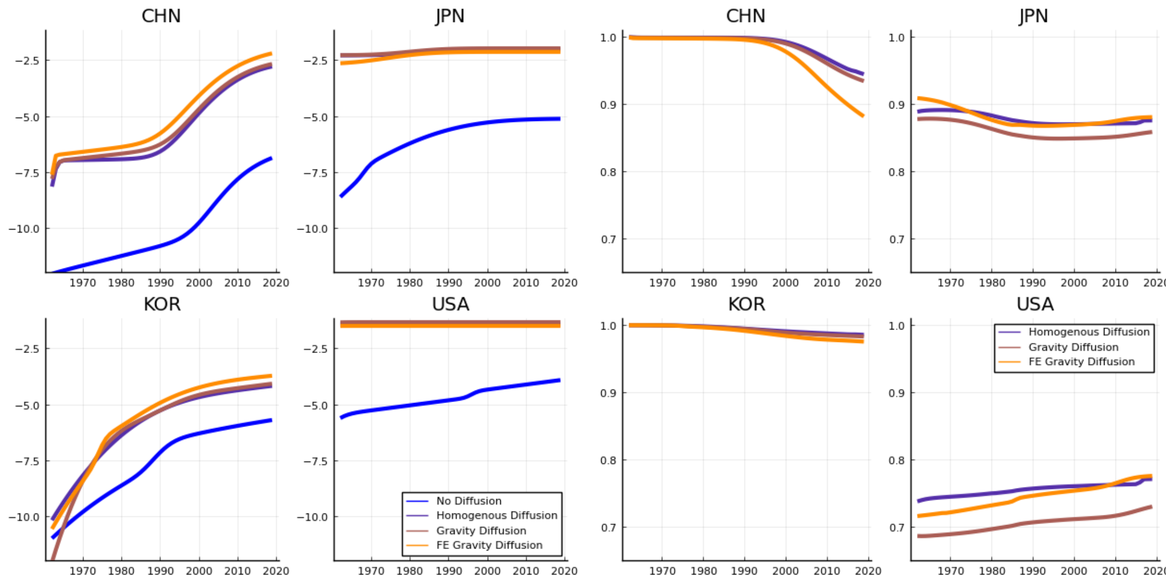
$$\text{Sharing}_{oo'}(t) \equiv \sum_{n=1}^{N} \int_{-\infty}^{t} \left(p_{no}(t-t^*)p_{no'}(t-t^*)\right)^{1-\rho} \frac{\lambda_n(t^*)}{\sum_{n=1}^{N} \Lambda_n(t)} \mathrm{d}t^*. \tag{24}$$

In Figure 1b, we show the histogram of $\text{Sharing}_{oo'}(t)$ for each specification for diffusion. Mimicking the histograms for the cross-price elasticities, diffusion specifications with less shared ideas correspond to lower cross elasticities. Overall, these estimates show that departures from CES due to diffusion are frequent and can be large, and many ideas are shared between countries.

In Table 2, we explore the link between substitution elasticities and shared knowledge. From (14), we expect cross-price elasticities to be positively related to shared knowledge after conditioning on relative unit production costs across countries. Accordingly, we regress the implied cross-price elasticities on the probabilities that pairs of countries share knowledge, using fixed effects to control for unit costs. The coefficients are not only positive and significant, indicating that higher elasticities are associated with more sharing of knowledge between country pairs, but they also survive the inclusion of a battery of fixed effects.

Figure 2: Evolution of knowledge, selected countries.

(a) Innovation. (b) Diffusion.



Notes: Results from estimating the model with two innovation surges, $S = 2$, and: homogenous international diffusion, $\delta_{nm} = \delta$ for all $m \neq n$; gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \mathrm{Dist}_{nm})$ for all $m \neq n$; and FE gravity diffusion, $\delta_{nm} = \delta_m^A \delta_n^I \exp(-\kappa^{\delta} \ln \mathrm{Dist}_{nm})$ for all $m \neq n$. In all cases, $\delta_{nn} = \infty$. Innovation refers to estimates of $\ln \Lambda_n(t)$, while diffusion refers to the fraction of productivity coming from using ideas from other countries, $1 - \Lambda_m(t)/T_m(t)$.

Knowledge dynamics. We next turn to the dynamics of knowledge. Our estimates yield predictions on the evolution of knowledge over time in each country, both coming from innovation and diffusion. Figure 2 shows our estimates for a selected number of countries, for each of the model specifications for diffusion. Figure 2a shows that the model picks up the surge of China as a source of innovated knowledge by the end of the 1990's and beginning of the 21st century. As a share of total knowledge in each year, China goes from having innovated almost a zero share of global knowledge to around ten percent by 2019. The estimates for all the diffusion specifications pick up this increase. Estimates for Japan show a surge in innovation, which begins near the start of the sample in 1962, only if diffusion is shut off. Once the model allows for diffusion, Japan's stock of innovated ideas increases steadily throughout the sample. Estimates for the United States indicate that this country contributes a large amount to global knowledge (around 25 percent), and does not experience a single large surge of innovation within the sample. Finally, all the specifications for diffusion pick up an innovation surge in Korea that is particularly pronounced in the sixties and seventies. Appendix Figure B.4 shows the stock of innovated knowledge, for each of the 20 countries in our sample.

Figure 2b shows the evolution of the share of each country's knowledge that diffused from other countries, $1-\Lambda_m(t)/T_m(t)$. While in China the share of foreign knowledge decreased — as domestic sources of knowledge were surging — the opposite was true for the United States, where foreign sources of knowledge went from 70 to around 80 percent of their total knowledge. In general, our estimates deliver very high shares of foreign knowledge for most countries in the sample over the entire period — including the richest countries. This result is also consistent with the insight of Nelson and Phelps (1966) that long-run cross-country differences in income relate to diffusion of ideas from global innovators to the rest of the world. Appendix Figure B.5 shows the stock of diffused knowledge, for each of the 20 countries in our sample, while Appendix Figure B.6 shows diffusion curves for each country (adopter) in our sample from each other country (innovator). For each country, there is always a partner — in general, one geographically close-by — for which the diffusion rates are higher than for the remaining locations.

4.2 Gravity, Research, and Multinational Production

We now contrast the estimates coming from our preferred specification for diffusion with data. These comparisons provide some form of validation to our estimates of diffusion. Our main findings show that these estimates follow a spatial pattern and capture the cross-country patterns of multinational production flows.

Gravity and research intensity. Are diffusion patterns related to geography, country size, and measures of research intensity? Table 3 shows the results of regressing our estimate of $T_{nm}(t)$ as defined in (5), on bilateral distance, GDP, number of researchers, and researchers as a share of population, for the origin and receiving country. As expected, the further away the countries are, the lower the number of ideas from n contributing to productivity in m (column 1), the lower the share of knowledge shared with m from n (column 2), and the lower the bilateral diffusion rate (column 3). Country size of the adopter increases $T_{nm}(t)$ (column 1) and the share of their ideas that gets adopted by m (column 2), while the size of the innovator decreases these two variables.

Turning to the diffusion rates, we consider $\Delta \ln T_{nm}(t)$ in column 3. Since we

Table 3: Estimation results: bilateral diffusion patterns. OLS.

	$\ln T_{nm}(t)$	$\ln \frac{T_{nm}(t)}{\Lambda_n(t)}$	$\Delta \ln T_{nm}(t)$
	(1)	(2)	(3)
$\ln \text{Dist}_{nm}$	-0.923***	-0.725***	-0.002***
	(0.028)	(0.029)	(0.000)
$\log Y_m(t)$	1.129***	1.236***	
	(0.024)	(0.025)	
$\log Y_n(t)$	-0.071**	-1.126***	
	(0.023)	(0.024)	
$\ln \frac{T_{nm}(t-1)}{\Lambda_n(t-1)}$			-0.004***
			(0.000)
$\log \text{Researchers}_m(t)$			0.006***
			(0.000)
$\log \text{Researchers}_n(t)$			-0.003***
			(0.000)
$\log \text{Researchers pc}_m(t)$			-0.010***
			(0.001)
$\log \text{Researchers pc}_n(t)$			-0.001
			(0.000)
Constant	Yes	Yes	No
t fixed effects	No	No	Yes
Observations	22,040	22,040	7,128
R-squared	0.122	0.180	0.383
Within R-squared			0.355

Notes: Results from estimating the model with two innovation surges, $S = 2$, and gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$, with $\delta_{nn} = \infty$. $T_{nm}(t)$ is defined in (5) and we include observations with $m \neq n$. Bilateral distance is from CEPII, GDP is from PWT(10.0), and Researchers are from OECD. Robust standard errors are in parenthesis with levels of significance denoted by *** p < 0.001, and ** p < 0.01 and * p<0.05.

anticipate that diffusion will slow down when an adopter has already learned all the ideas from n, we control for the previous period share, $T_{nm}(t-1)/\Lambda_n(t-1)$, so that we take into account the gap to full diffusion. As expected, the sign is negative. The higher the number of researchers in the adopter country, the higher the diffusion rate. This result suggests that diffusion needs some absorptive capacity at the receiving end of knowledge to materialize, an idea put forward by the early work of Nelson and Phelps (1966). In contrast, more researchers in the innovator country decreases the diffusion rates of their ideas, while a higher research intensity in either the innovator or adopter country (as measured by researchers per capita) slows down diffusion.

Multinational production. We now compare estimates of our model with data on multinational production (MP) — i.e. sales of affiliates operating in country m with parents in country n, from Ramondo et al. (2015). Indeed, MP is an important channel through which ideas innovated in a country diffuse — and get adopted — abroad, albeit not the only channel. Using the expenditure share in (9), we can calculate the global expenditure on goods from country m produced with ideas from country n,

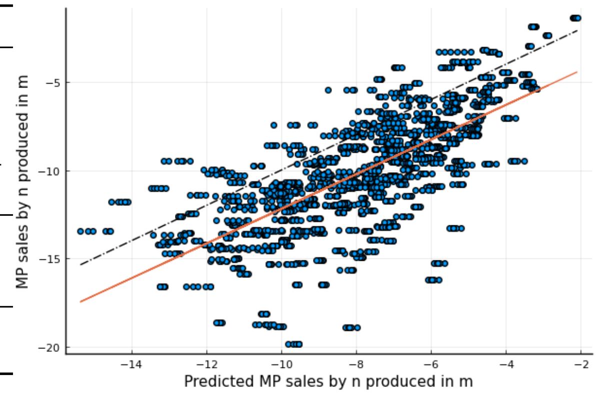
$$X_{nm}(t) = \sum_d \pi_{nmd}(t) X_d(t). \tag{25}$$

If all diffusion from n to m happened through MP, this variable would be the model counterpart of bilateral MP in the data. Results in Table 3 show that bilateral MP is correlated with our measure of shared knowledge $T_{nm}(t)$, distance, and our predicted expenditure in goods from m produced with ideas from n, $X_{nm}(t)$. Figure 3 further shows that the model variable $X_{nm}(t)$ predicts higher expenditure in goods produced in m with ideas from n than the MP data do. This result is as expected when the activity of multinational firms is not the only source of international technology transfer. Appendix Figure B.7 shows the same scatter but using MP shares instead of flows, with unchanged results.

Innovation and Research. Our final result links growth in our estimates of innovated knowledge stocks to the number of researchers in a country over time. First, the more innovative the country, captured by a higher level of knowledge $\Lambda_n(t-1)$, the lower the innovation growth rate. Second, a higher number of researchers in a country is associated with a significantly higher growth of innovated knowledge,

Figure 3: Diffusion patterns and multinational production. OLS.

	$\ln MP\ sales_{nm}(t)$			
$\ln X_{nm}(t)$	0.982***			0.984***
	(0.057)			(0.123)
$\ln T_{nm}(t)$		0.198***		
		(0.038)		
$\ln Dist_{nm}$			-1.801***	-1.523***
			(0.182)	(0.098)
$m \times t$ FE	No	No	Yes	Yes
$n \times t$ FE	No	No	Yes	Yes
Constant	Yes	Yes	No	No
Observations	2,022	2,022	2,022	2,022
R-squared	0.496	0.072	0.794	0.869



Notes: Results from estimating the model with two innovation surges, $S = 2$, and gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$ and $\delta_{nn} = \infty$. $T_{nm}(t)$ is defined in (5) while $X_{nm}(t)$ is defined in (25). Bilateral distance is from CEPII. $MP\ sales_{nm}$ are sales of affiliates operating in country m with parents in country n, from Ramondo et al. (2015). In the table, robust standard errors are in parenthesis with levels of significance denoted by *** p < 0.001, and ** p < 0.01 and * p<0.05. In the figure, the dashed line is $45°$ line, while the solid line is best fit.

Table 4: Estimation results: growth patterns. OLS.

	$\Delta \ln \Lambda_n(t)$	
$\ln \Lambda_n(t-1)$	-0.027***	-0.026***
	(0.002)	(0.002)
$\ln$ Researchers$_n(t)$	0.026***	0.027***
	(0.002)	(0.002)
$\ln$ Researchers per capita$_n(t)$		-0.013***
		(0.002)
Observations	537	537
R-squared	0.465	0.523

Notes: Results from estimating the model with two innovation surges, $S = 2$, and gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$ and $\delta_{nn} = \infty$. The dependent variable is changes in $\ln \Lambda_n(t)$. Researches data are from OECD and population data are from PWT(10.0). All specifications include year fixed effects. Robust standard errors are in parenthesis with levels of significance denoted by *** p < 0.001, and ** p < 0.01 and * p<0.05.

while research intensity, captured by the share of researchers in the population, is negatively associated with knowledge growth. Combined with the negative estimate on the innovation lag, this result is consistent with models of semi-endogenous growth where the accumulated level of knowledge depends on scale, but inconsistent with models of endogenous growth where growth rates exhibit scale effects (see Jones, 2005, for a detailed discussion).

5 Conclusion

In this paper, we propose to use easily-available data on trade flows across countries and over time to uncover the global dynamics of knowledge. To such end, we build a model of trade, innovation, and diffusion, inspired by the work of Eaton and Kortum (1999). The model fits the general framework we developed in our previous work (Lind and Ramondo, 2023a), ensuring that the global productivity distribution is max-stable Fréchet. Under Ricardian trade, max-stability ensures closed-form solutions for expenditure shares and prices across countries. In this way, we obtain a transparent mapping between observable expenditure flows and unobservable knowledge flows across countries, which allows us to implement an estimation procedure based on international trade data. Our estimates suggest that although innovation is highly correlated with economic growth, there are many high income countries that primarily produce using diffused ideas. Even though we only use data on bilateral trade shares to estimates cross-country diffusion flows, these flows are strongly correlated with multinational production flows, which are one of the main channels for international technology diffusion.

References

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

Akcigit, U., J. Grigsby, and T. Nicholas. Immigration and the rise of american ingenuity. *American Economic Review*.

Alvarez, F. and R. E. Lucas (2007). General equilibrium analysis of the eaton-kortum model of international trade. *Journal of Monetary Economics 54*(6), 1726–1768.

Arkolakis, C., N. Ramondo, A. Rodríguez-Clare, and S. Yeaple (2018). Innovation and production in the global economy. *American Economic Review Forthcoming*.

Bas, M., T. Mayer, and M. Thoenig (2017). From micro to macro: Demand, supply, and heterogeneity in the trade elasticity. *Journal of International Economics 108*(C), 1–19.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Bloom, N., T. A. Hassan, A. Kalyani, J. Lerner, and A. Tahoun (2021). The diffusion of disruptive technologies. *NBER Working Paper 28999*.

Boehm, C. E., N. Pandalai-Nayar, and A. A. Levchenko (2023). The long and short (run) of trade elasticities. *American Economic Review 113*(4), 861–905.

Brooks, W. and P. Pujolas (2019). Gains from trade with variable trade elasticities. *International Economic Review 60*(4), 1619–1646.

Buera, F. J. and E. Oberfield (2020). The global diffusion of ideas. *Econometrica 88*(1).

Cai, S. and W. Xiang (2022). Multinational production, technology diffusion, and economic growth. *Manuscript (Yale)*.

Caron, J., T. Fally, and J. Markusen (2014). International trade puzzles: a solution linking production factors and demand. *The Quarterly Journal of Economics 129*(3), 1501–1552.

Chen, Z., Z. Liu, J. C. Suarez-Serrato, and D. Y. Xu (2021). Notching r&d investment with corporate income tax cuts in china. *American Economic Review 111*(7), 2065–2100.

Comin, D. and B. Hobijn (2004). Cross-country technology adoption: making the theories face the facts. *Journal of Monetary Economics 51*, 39–83.

Comin, D. A., M. Dmitriev, and E. Rossi-Hansberg (2012). The spatial diffusion of technology. Working Paper 18534, National Bureau of Economic Research.

Eaton, J. and S. Kortum (1996a). Measuring technology diffusion and the international sources of growth. *Eastern Economic Journal 22*(4), 401–410.

Eaton, J. and S. Kortum (1996b). Trade in ideas patenting and productivity in the oecd. *Journal of International Economics 40*(3), 251–278.

Eaton, J. and S. Kortum (1997). Engines of growth: Domestic and foreign sources of innovation. *Japan and the World Economy 9*, 235–259.

Eaton, J. and S. Kortum (1999). International technology diffusion: Theory and measurement. *International Economic Review 40*(3), 537–570.

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

Eaton, J. and S. Kortum (2012). Putting ricardo to work. *Journal of Economic Perspectives 26*(2), 65–90.

Eaton, J., S. Kortum, and S. Sotelo (2013). International trade: Linking micro and macro. *Advances in Economics and Econometrics: Tenth World Congress, Volume II, Applied Economics*, 329–370.

Feenstra, R. C., P. Luck, M. Obstfeld, and K. N. Russ (2018). In search of the armington elasticity. *The Review of Economics and Statistics 100*(1), 135–150.

Innes, M. (2018). Don't unroll adjoint: Differentiating ssa-form programs. *arXiv preprint arXiv:1810.07951*.

Jones, C. I. (2005). Growth and ideas. *Handbook of Economic Growth, P. Aghion and S. Durlauf, eds. 1A*(1/2), 1064–1108.

Jones, C. I. (2011). Intermediate goods and weak links in the theory of economic development. *American Economic Journal: Macroeconomics 3*(2), 1–28.

Keller, W. (2010). International trade, foreign direct investment, and technology spillovers. *Handbook of the Economics of Innovation 2*, 793–829.

Kingma, D. P. and J. Ba (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kingman, J. (1992). *Poisson Processes*. Oxford Studies in Probability. Clarendon Press.

Krugman, P. (1979). A model of innovation, technology transfer, and the world distribution of income. *Journal of Political Economy 87*(2), 253–266.

Lashkari, D. and M. Mestieri (2016). Gains from trade with heterogeneous income and price elasticities. *Mimeo, Harvard University*.

Lind, N. and N. Ramondo (2023a). Global innovation and knowledge diffusion. *AER:Insights 5*(4), 494–510.

Lind, N. and N. Ramondo (2023b). Trade with correlation. *American Economic Review 103*(2), 317–353.

Ma, X. (2023). College expansion, trade, and innovation: Evidence from china. *International Economic Review forthcoming*.

Nelson, R. R. and E. S. Phelps (1966). Investment in humans, technological diffusion, and economic growth. *The American Economic Review 56*(1/2), 69–75.

Parente, S. L. and E. C. Prescott (1994). Barriers to technology adoption and development. *Journal of political Economy 102*(2), 298–321.

Ramondo, N. and A. Rodríguez-Clare (2013). Trade, multinational production, and the gains from openness. *Journal of Political Economy 121*(2), 273–322.

Ramondo, N., A. Rodríguez-Clare, and F. Tintelnot (2015). Multinational production: Data and stylized facts. *American Economic Review, Papers & Proceedings 105*(5), 530–36.

A Derivations

A.1 Distribution of Applications

We show that $A^*_{im}(t)$ in (1) is independent Fréchet. For $t < t^*$, $A^*_{im}(t) = 0$ for all m. For $t \geq t^*$ we have

$$
\begin{aligned}
&\mathbb{P}\left[A^*_{i1}(t) \leq a_1, \ldots A^*_{iN}(t) \leq a_N \mid Q_i = q, n_i = n, t^*_i = t^*\right] \\
&= \mathbb{P}\left[\max_{j=1,2,\ldots} A_{ij}\mathbf{1}\{m_{ij} = m, t_{ij} \leq t\} \leq a_m \ \ \forall m = 1, \ldots, N \mid Q_i = q, n_i = n, t^*_i = t^*\right] \\
&= \mathbb{P}\left[A_{ij}\mathbf{1}\{m_{ij} = m, t_{ij} \leq t\} \leq a_m \ \ \forall m = 1, \ldots, N \ \ \forall j \mid Q_i = q, n_i = n, t^*_i = t^*\right] \\
&= \mathbb{P}\left[A_{ij} \leq a_{m_{ij}} \ \ \forall j \text{ s.t. } t_{ij} \leq t \mid Q_i = q, n_i = n, t^*_i = t^*\right] \\
&= \mathbb{P}\left[A_{ij} > a_{m_{ij}} \ \text{ for no } j \text{ s.t. } t_{ij} \leq t \mid Q_i = q, n_i = n, t^*_i = t^*\right]
\end{aligned}
\tag{A.1}
$$

The third equality follows because for any given j such that $t_{ij} \leq t$, whenever $m_{ij} \neq m$ then $A_{ij}\mathbf{1}\{m_{ij} = m, t_{ij} \leq t\} = 0$ and $a_m \geq 0$ and whenever $m_{ij} = m$ we have $A_{ij}\mathbf{1}\{m_{ij} = m, t_{ij} \leq t\} \leq a_m$ if and only if $A_{ij} \leq a_{m_{ij}}$.

Next, because $\{A_{ij}, m_{ij}, t_{ij}\}_{j=1,2,\ldots}$ conditional on $Q_i = q$, $n_i = n$, and $t^*_i = t^*$ are the points of a Poisson process with intensity $\Gamma(1-\theta/\sigma)^{-\sigma/\theta}\sigma a^{-\sigma-1}p'_{nm}(t-t^*)\mathrm{d}a\mathrm{d}t$, we can calculate the void probability in (A.1) as

$$
\begin{aligned}
&\mathbb{P}\left[A^*_{i1}(t) \leq a_1, \ldots A^*_{iN}(t) \leq a_N \mid Q_i = q, n_i = n, t^*_i = t^*\right] \\
&= \exp\left[-\sum_{m=1}^{N}\int_{t^*}^{t}\int_{a_m}^{\infty}\Gamma(1-\theta/\sigma)^{-\sigma/\theta}\sigma a^{-\sigma-1}p'_{nm}(x-t^*)\mathrm{d}a\mathrm{d}x\right] \\
&= \exp\left[-\Gamma(\rho)^{-\frac{1}{1-\rho}}\sum_{m=1}^{N}p_{nm}(t-t^*)a_m^{-\frac{\theta}{1-\rho}}\right] \equiv M_n(a_1, \ldots, a_N; t-t^*)
\end{aligned}
\tag{A.2}
$$

where $\rho \equiv 1 - \theta/\sigma$.

A.2 Global Productivity Distribution

The joint distribution for productivity, defined in (3) as $Z_m(t) \equiv \max_{i=1,2,\dots} Q_i A^*_{im}(t)$, is

$$
\begin{aligned}
F(z_1,\dots,z_N;t) &\equiv \mathbb{P}\left[Z_1(t)\le z_1,\dots,Z_N(t)\le z_N\right] = \mathbb{P}\left[\max_{i=1,2,\dots} Q_i A^*_{im}(t) \le z_m \;\; \forall m=1,\dots,N\right] \\
&= \mathbb{P}\left[Q_i A^*_{im}(t)\le z_m \;\; \forall m=1,\dots,N \;\; \forall i\right] = \mathbb{P}\left[Q_i A^*_{im}(t) z_m^{-1} \le 1 \;\; \forall m=1,\dots,N \;\; \forall i\right] \\
&= \mathbb{P}\left[Q_i \max_{m=1,\dots,M} A^*_{im}(t) z_m^{-1} \le 1 \;\; \forall i\right] = \mathbb{P}\left[Q_i \max_{m=1,\dots,M} A^*_{im}(t) z_m^{-1} > 1 \;\; \text{for no } i\right]
\end{aligned}
\tag{A.3}
$$

Next, since $\{Q_i, n_i, t_i^*\}_{i=1,2,\dots}$ are the points of a Poisson process with intensity of $\theta q^{-\theta-1}\lambda_n(t^*)\mathrm{d}q\mathrm{d}t^*$ and $(A^*_{i1}(t),\dots,A^*_{iN}(t))$ is a random vector whose distribution conditional on $Q_i = q$, $n_i = n$, and $t_i^* = t^*$ is $M_n(a_1,\dots,a_N;t-t^*)$, then, by the marking theorem for Poisson processes (Kingman, 1992), $\{Q_i, n_i, t_i^*, (A^*_{i1}(t),\dots,A^*_{iN}(t))\}_{i=1,2,\dots}$ are the points of a Poisson process with intensity $\mathrm{d}M_n(a_1,\dots,a_N;t-t^*)\theta q^{-\theta-1}\lambda_n(t^*)\mathrm{d}q\mathrm{d}t^*$. We can then calculate the void probability in (A.3) using this result to get

$$
\begin{aligned}
&F(z_1,\dots,z_N;t) \\
&= \exp\left[-\int_{-\infty}^{t}\sum_{n=1}^{N}\int_0^\infty\int_{\mathbb{R}_+^N} \mathbf{1}\left\{q\max_{m=1,\dots,M} a_m z_m^{-1} > 1\right\}\mathrm{d}M_n(a_1,\dots,a_N \mid t-t^*)\theta q^{-\theta-1}\mathrm{d}q\lambda_n(t^*)\mathrm{d}t^*\right] \\
&= \exp\left[-\int_{-\infty}^{t}\sum_{n=1}^{N}\int_{\mathbb{R}_+^N}\int_{(\max_{m=1,\dots,M} a_m z_m^{-1})^{-1}}^\infty \theta q^{-\theta-1}\mathrm{d}q\mathrm{d}M_n(a_1,\dots,a_N \mid t-t^*)\lambda_n(t^*)\mathrm{d}t^*\right] \\
&= \exp\left[-\int_{-\infty}^{t}\sum_{n=1}^{N}\int_{\mathbb{R}_+^N}\max_{m=1,\dots,M} a_m^\theta z_m^{-\theta}\mathrm{d}M_n(a_1,\dots,a_N \mid t-t^*)\lambda_n(t^*)\mathrm{d}t^*\right].
\end{aligned}
\tag{A.4}
$$

From (A.2), $A^*_{im}(t)$ is distributed Fréchet independent across m, and so

$$
\begin{aligned}
\int_{\mathbb{R}_+^N}\max_{m=1,\dots,M} a_m^\theta z_m^{-\theta}\mathrm{d}M_n(a_1,\dots,a_N \mid t-t^*) &= \mathbb{E}\left[\max_{m=1,\dots,M}(A^*_{im}(t))^\theta z_m^{-\theta} \mid Q_i = q, n_i = n, t_i^* = t^*\right] \\
&= \Gamma(\rho)\left(\Gamma(\rho)^{-\frac{1}{1-\rho}}\sum_{m=1}^{N} p_{nm}(t-t^*) z_m^{-\frac{\theta}{1-\rho}}\right)^{1-\rho} \\
&= \left(\sum_{m=1}^{N} p_{nm}(t-t^*) z_m^{-\frac{\theta}{1-\rho}}\right)^{1-\rho},
\end{aligned}
$$

which replacing back in (A.4) gives the expression in (4).

A.3 Innovation Process

Since the innovation rate in country n at time t is

$$\lambda_n(t) = \Lambda_n'(t) = \gamma_n \Lambda_n(t) + \sum_{s=1}^{S} \lambda_n^s \frac{1}{\nu_n^s} \phi\left(\frac{t-\mu_n^s}{\nu_n^s}\right),$$

we have that

$$\frac{\partial}{\partial t}\left[e^{-\gamma_n t}\Lambda_n(t)\right] = e^{-\gamma_n t}\Lambda_n'(t) - \gamma_n e^{-\gamma_n t}\Lambda_n(t) = \sum_{s=1}^{S} \lambda_n^s e^{-\gamma_n t} \frac{1}{\nu_n^s} \phi\left(\frac{t-\mu_n^s}{\nu_n^s}\right).$$

Integrating from $-\infty$ to t and using $\lim_{t\to-\infty} e^{-\gamma_n t}\Lambda_n(t) = 0$ yields

$$e^{-\gamma_n t}\Lambda_n(t) = \sum_{s=1}^{S} \lambda_n^s \int_{-\infty}^{t} e^{-\gamma_n t^*} \frac{1}{\nu_n^s} \phi\left(\frac{t^*-\mu_n^s}{\nu_n^s}\right) \mathrm{d}t^*$$

so that

$$\Lambda_n(t) = \sum_{s=1}^{S} \Lambda_n^s(t), \quad \text{where} \quad \Lambda_n^s(t) \equiv \lambda_n^s \int_{-\infty}^{t} e^{\gamma_n (t-t^*)} \frac{1}{\nu_n^s} \phi\left(\frac{t^*-\mu_n^s}{\nu_n^s}\right) \mathrm{d}t^*. \tag{A.5}$$

A.4 Derivation of Equation (23)

Combining (8), (9), (10), (11), and (12), and defining $r_{od}(t) \equiv P_{od}(t)/P_d(t)$, we can rewrite expenditure shares as

$$\pi_{od}(t) = \sum_{n=1}^{N} \int_{-\infty}^{t} p_{no}(t-t^*) r_{od}(t)^{-\frac{\theta}{1-\rho}} \left(\sum_{m=1}^{N} p_{nm}(t-t^*) r_{md}(t)^{-\frac{\theta}{1-\rho}}\right)^{-\rho} \lambda_n(t^*) \mathrm{d}t^*. \tag{A.6}$$

Multiplying (A.6) by $r_{od}(t)^{\theta}$, taking the time derivative, and using Leibniz's rule, yields

$$\frac{\partial}{\partial t}\left[r_{od}(t)^{\theta}\pi_{od}(t)\right] \equiv \frac{\partial \tilde{\pi}_{od}(t)}{\partial t} = \sum_{n=1}^{N} p_{no}(0) \left(\sum_{m=1}^{N} p_{nm}(0) \left(\frac{r_{md}(t)}{r_{od}(t)}\right)^{-\frac{\theta}{1-\rho}}\right)^{-\rho} \lambda_n(t)$$
$$+ \sum_{n=1}^{N} \int_{-\infty}^{t} \omega_{nod}(t^*, t) \lambda_n(t^*) \mathrm{d}t^*,$$

where

$$\omega_{nod}(t^*,t) \equiv \frac{\partial}{\partial t}\left[p_{no}(t-t^*)\left[\sum_{m=1}^{N} p_{nm}(t-t^*)\left(\frac{r_{md}(t)}{r_{od}(t)}\right)^{-\frac{\theta}{1-\rho}}\right]^{-\rho}\right]. \tag{A.7}$$

Using $p_{nm}(0) = \mathbf{1}\{m=n\}$, we get (23).

A.5 Proposition A.1

Proposition A.1. *Fix some $T \in \mathbb{R}$ and assume that there exists $g > 0$ such that $e^{-gt}\omega_{nod}(t^*,t)$, with $\omega_{nod}(t^*,t)$ defined in (A.7), is bounded on $(-\infty,T]^2$. Then, there is a unique innovation process, $\lambda_o(t^*)$, that is consistent with observed trade flows, $\pi_{od}(t)$, and import cost indices, $P_{od}(t)$, for any individual destination d.*

We prove that $\{\lambda_n(t^*)\}_{n\in\{1,\dots,N\},t^*\in(-\infty,T]} \mapsto \{\pi_{od}(t)\}_{(o,d)\in\{1,\dots,N\}^2,t\in(-\infty,T]}$, the map from innovation rates to trade shares, is injective. Note that, in (A.6), the map from innovation rates to trade shares is a linear integral operator. As a result, to show injectivity, we show that $\pi_{od}(t) = 0$ implies that $\lambda_n(t^*) = 0$.

Next, suppose that $\pi_{od}(t) = 0$. Then, $\frac{\partial}{\partial t}\left[r_{od}(t)^{\theta}\pi_{od}(t)\right] = 0$, so (23) reduces to

$$\lambda_o(t) = -\sum_{n=1}^{N}\int_{-\infty}^{t}\omega_{nod}(t^*,t)\lambda_n(t^*)\mathrm{d}t^*.$$

Defining $x_o(t) = e^{-gt}\lambda_o(t)$, we have

$$x_o(t) = -\sum_{n=1}^{N}\int_{-\infty}^{t}e^{-g(t-t^*)}\omega_{nod}(t^*,t)x_n(t^*)\mathrm{d}t^*.$$

Let $\bar{x}(t) = \max_{o=1,\dots,N}|x_o(t)|$ and $\bar{\omega}_d = \sup_{(t^*,t)\in(0,T]^2}\max_{(o,n)\in\{1,\dots,N\}} e^{-gt}|\omega_{nod}(t^*,t)|$, which is finite by assumption. Then,

$$\begin{aligned}\bar{x}(t) = \max_{o=1,\dots,N}|x_o(t)| &\le \max_{o=1,\dots,N}\sum_{n=1}^{N}\int_{-\infty}^{t}e^{-g(t-t^*)}|\omega_{nod}(t^*,t)||x_n(t^*)|\mathrm{d}t^* \\ &\le \bar{\omega}_d\sum_{n=1}^{N}\int_{-\infty}^{t}e^{gt^*}|x_n(t^*)|\mathrm{d}t^* \le \bar{\omega}_d N\int_{-\infty}^{t}e^{gt^*}\max_{n=1,\dots,N}|x_n(t^*)|\mathrm{d}t^* = \bar{\omega}_d N\int_{-\infty}^{t}e^{gt^*}\bar{x}(t^*)\mathrm{d}t^*.\end{aligned}$$

Let $y(t) \equiv \bar{\omega}_d N \int_{-\infty}^{t} e^{gt^*} \bar{x}(t^*) \mathrm{d}t^* e^{-\bar{\omega}_d N g^{-1} e^{gt}}$. Note that $\lim_{t \to -\infty} y(t) = 0$. Then

$$\begin{aligned} y'(t) &= \bar{\omega}_d N e^{gt} \bar{x}(t) e^{-\bar{\omega}_d N g^{-1} e^{gt}} - \bar{\omega}_d N \int_{-\infty}^{t} e^{gt^*} \bar{x}(t^*) \mathrm{d}t^* e^{-\bar{\omega}_d N g^{-1} e^{gt}} \bar{\omega}_d N e^{gt} \\ &= \bar{\omega}_d N e^{gt} e^{-\bar{\omega}_d N g^{-1} e^{gt}} \left[\bar{x}(t) - \bar{\omega}_d N \int_{-\infty}^{t} e^{gt^*} \bar{x}(t^*) \mathrm{d}t^* \right] \leq 0, \end{aligned}$$

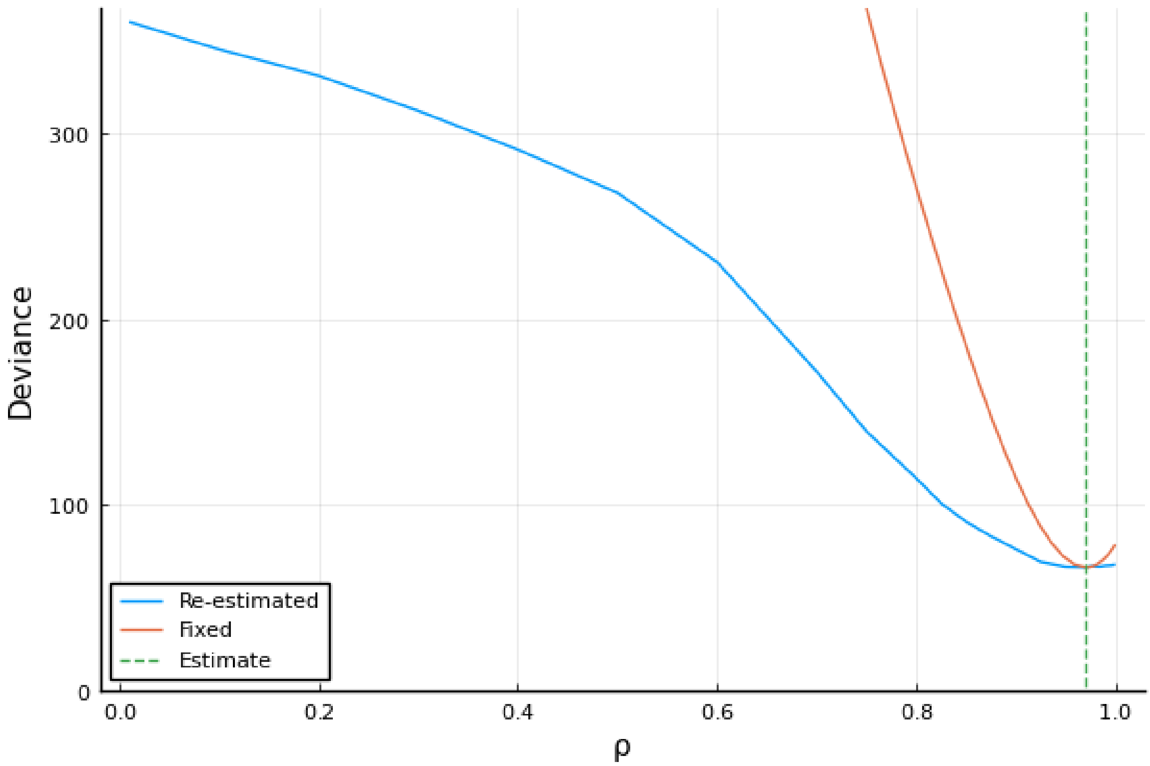
and hence, $y(t) - y(t_0) = \int_{t_0}^{t} y'(t^*) \mathrm{d}t^* \leq 0$. Then, $y(t) \leq y(t_0) \to 0$ as $t_0 \to -\infty$, so $y(t) \leq 0$. Due to the definition of $y(t)$, we then have

$$\bar{\omega}_d N \int_{-\infty}^{t} e^{gt^*} \bar{x}(t^*) \mathrm{d}t^* \leq 0 \quad \Longrightarrow \quad \bar{x}(t) \leq 0 \quad \Longrightarrow \quad x_o(t) = 0 \;\; \forall o = 1, \ldots, N.$$

Since $x_o(t) = e^{-gt} \lambda_o(t)$, we must have $\lambda_o(t) = 0$ for all $o = 1, \ldots, N$. Therefore, $\{\lambda_n(t^*)\}_{n \in \{1,\ldots,N\}, t^* \in (-\infty, T]} \mapsto \{\pi_{od}(t)\}_{(o,d) \in \{1,\ldots,N\}^2, t \in (-\infty, T]}$ is injective.

B Additional Results

Figure B.1: Estimation of the correlation coefficient ρ, robustness.



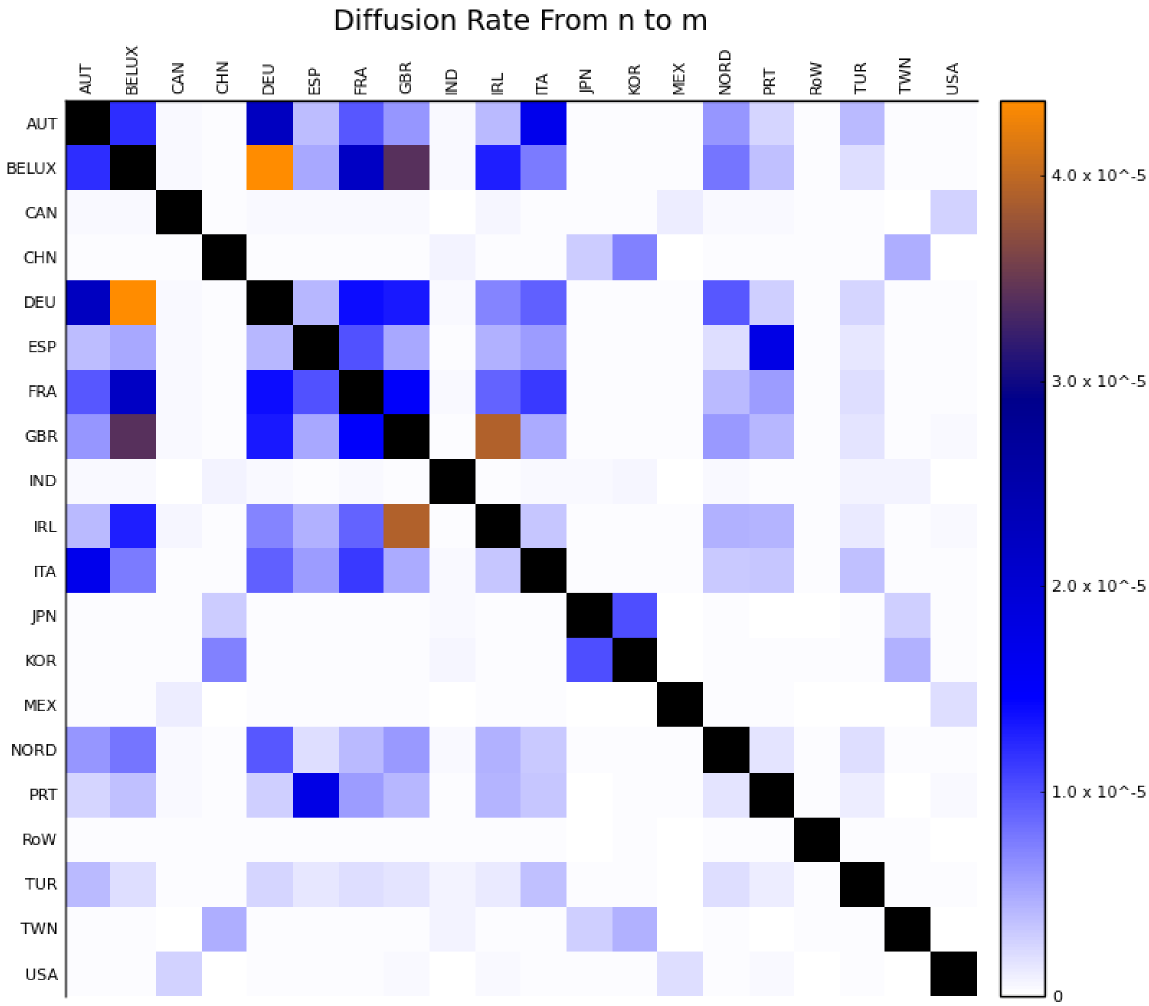
Notes: Results from estimating the model with two innovation surges, $S = 2$, and gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \mathrm{Dist}_{nm})$ for all $m \neq n$ and $\delta_{nn} = \infty$. In "Re-estimated" (blue line), we fix the value of ρ and re-estimate the remaining parameters; in "Fixed" (red line), we fix the value of the remaining parameters at the baseline values and change ρ. The green line shows our baseline estimate of $\rho = 0.97$.

Table B.1: Estimation results: country-level parameters.

Country name	Country code	γ_n	λ_n^1	μ_n^1	ν_n^1	λ_n^2	μ_n^2	ν_n^2
Austria	AUT	0.002	0.004	-84.69	52.73	0.001	24.52	0.468
Benelux	BLX	1e-10	0.019	-89.46	7.035	0.032	34.05	66.33
Canada	CAN	1e-12	0.018	-1.678	4.740	0.001	47.98	1.352
China	CHN	0.019	0.001	1.128	1.196	0.024	47.22	7.724
Germany	DEU	0.001	0.156	-213.2	171.8	0.000	24.71	0.182
Spain	ESP	2e-13	0.023	-285.5	276.8	0.005	32.76	6.128
France	FRA	6e-13	0.050	-95.86	91.42	0.001	24.75	0.595
Great Britain	GBR	9e-14	0.104	-268.3	315.8	3e-13	36.27	2.290
India	IND	5e-11	0.002	0.120	0.649	0.004	46.66	5.190
Ireland	IRL	3e-11	0.002	23.32	7.445	0.002	36.68	1.982
Italy	ITA	7e-13	0.037	-99.48	94.60	0.002	24.80	0.776
Japan	JPN	2e-13	0.102	-43.98	7.037	0.036	19.50	7.482
Korea	KOR	0.022	0.002	16.33	5.561	0.003	32.46	6.073
Mexico	MEX	0.000	0.019	-287.9	13.98	0.001	50.01	0.139
Nordic countries	NORD	1e-12	0.039	-52.5	32.85	0.001	25.50	0.578
Portugal	PRT	1e-11	0.002	-75.95	11.16	0.000	26.29	1.187
Rest of the world	RoW	1e-13	0.206	-5.094	1.469	0.064	54.96	4.462
Turkey	TUR	5e-12	0.003	-251.3	21.60	0.002	42.43	1.639
Taiwan	TWN	1e-11	0.005	-8.154	206.6	0.006	26.11	5.948
United States	USA	1e-13	0.265	-9.039	4.294	1e-12	36.50	0.745

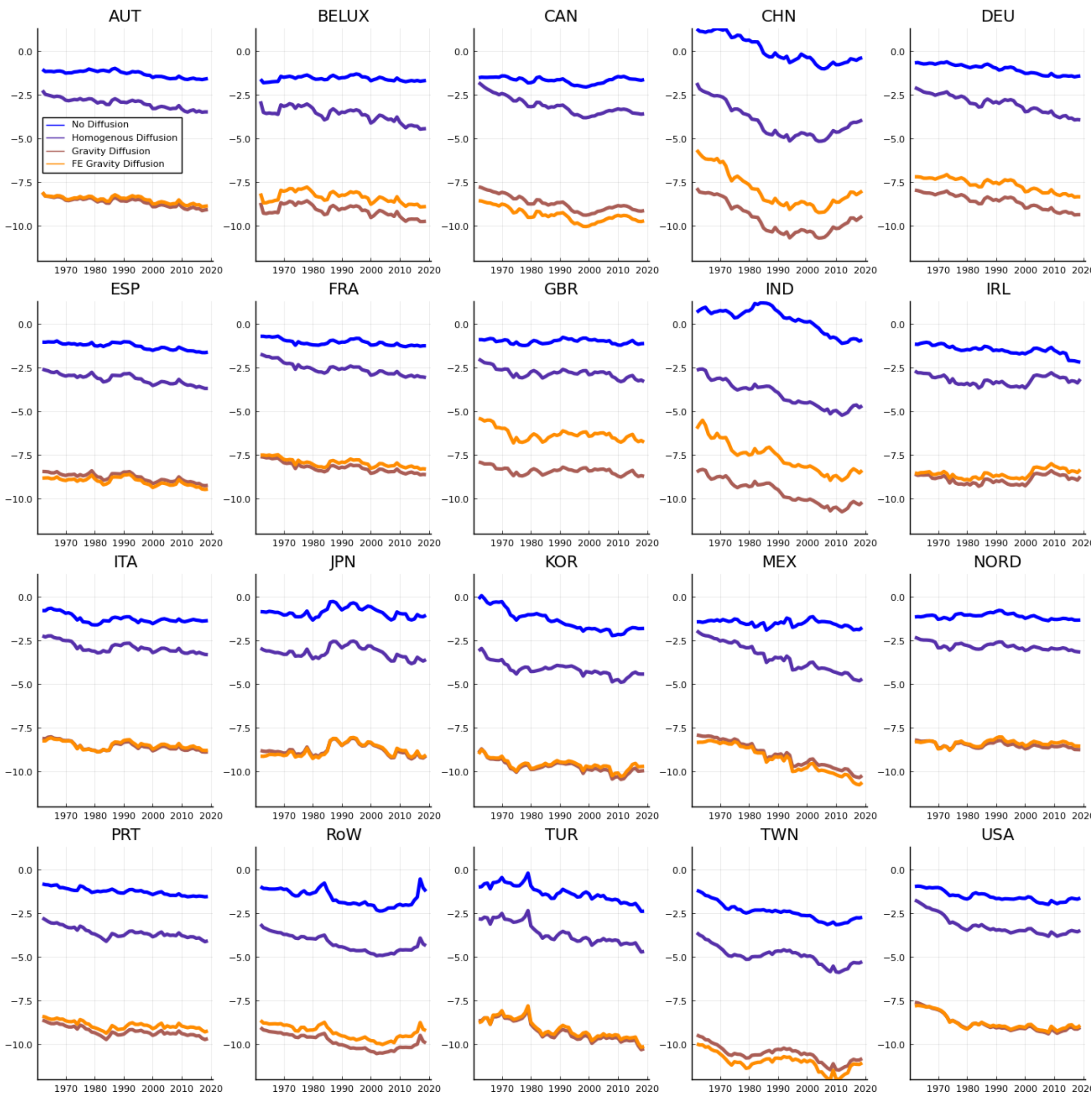
Notes: Benelux includes Belgium, Luxembourg, and Netherlands, while Nordic countries include to Denmark, Sweden, and Finland. Estimated parameters of the innovation process in (21), for $S = 2$.

Figure B.2: Bilateral diffusion parameters, by country pair.



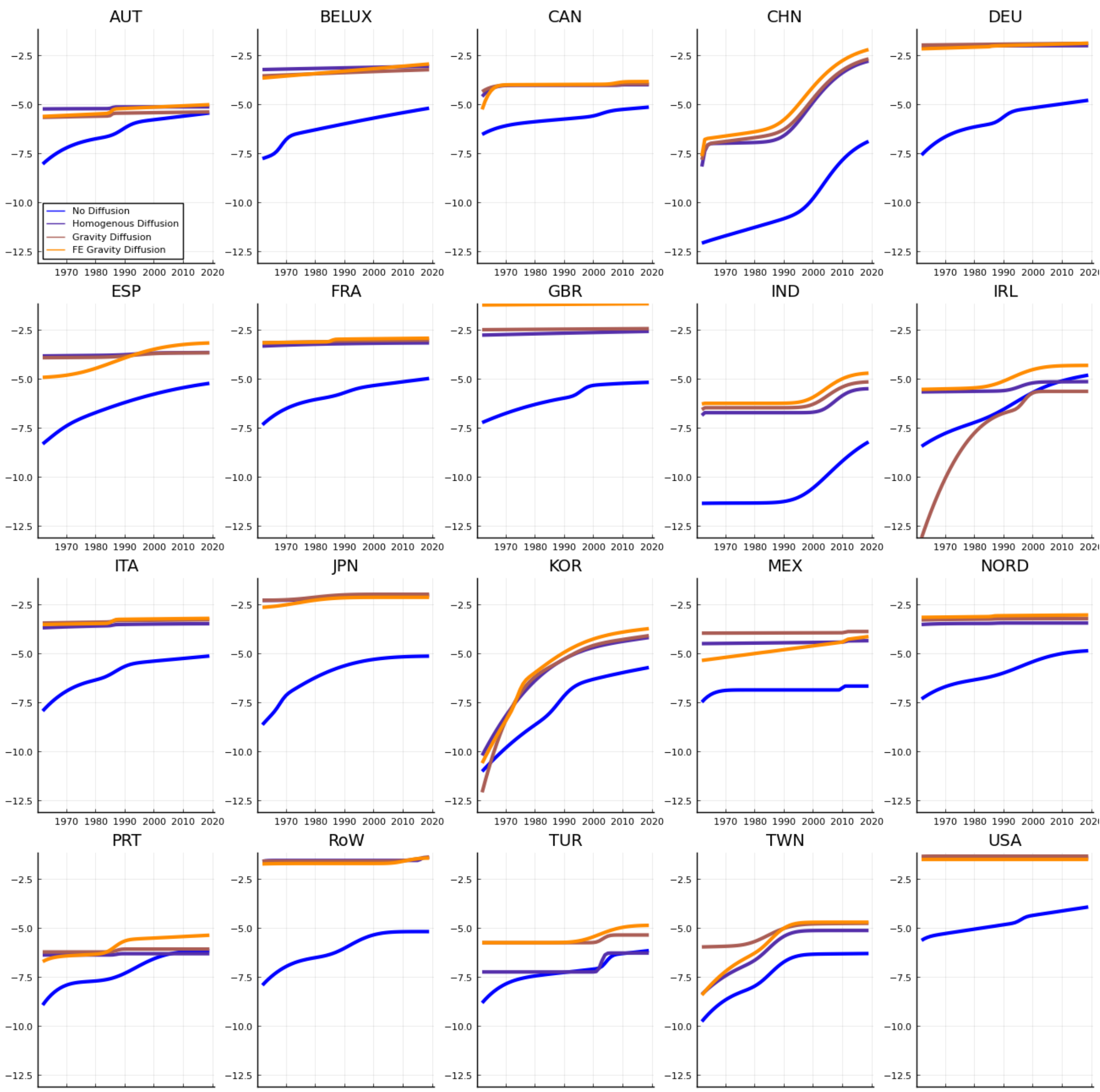
Notes: Results from estimating the model with two innovation surges, $S = 2$, and $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$. Note that $\delta_{nn} = \infty$ (represented by the black diagonal). The variable shown is the implied δ_{nm} from innovator n (column) to adopter m (row).

Figure B.3: Evolution of border costs, by country.



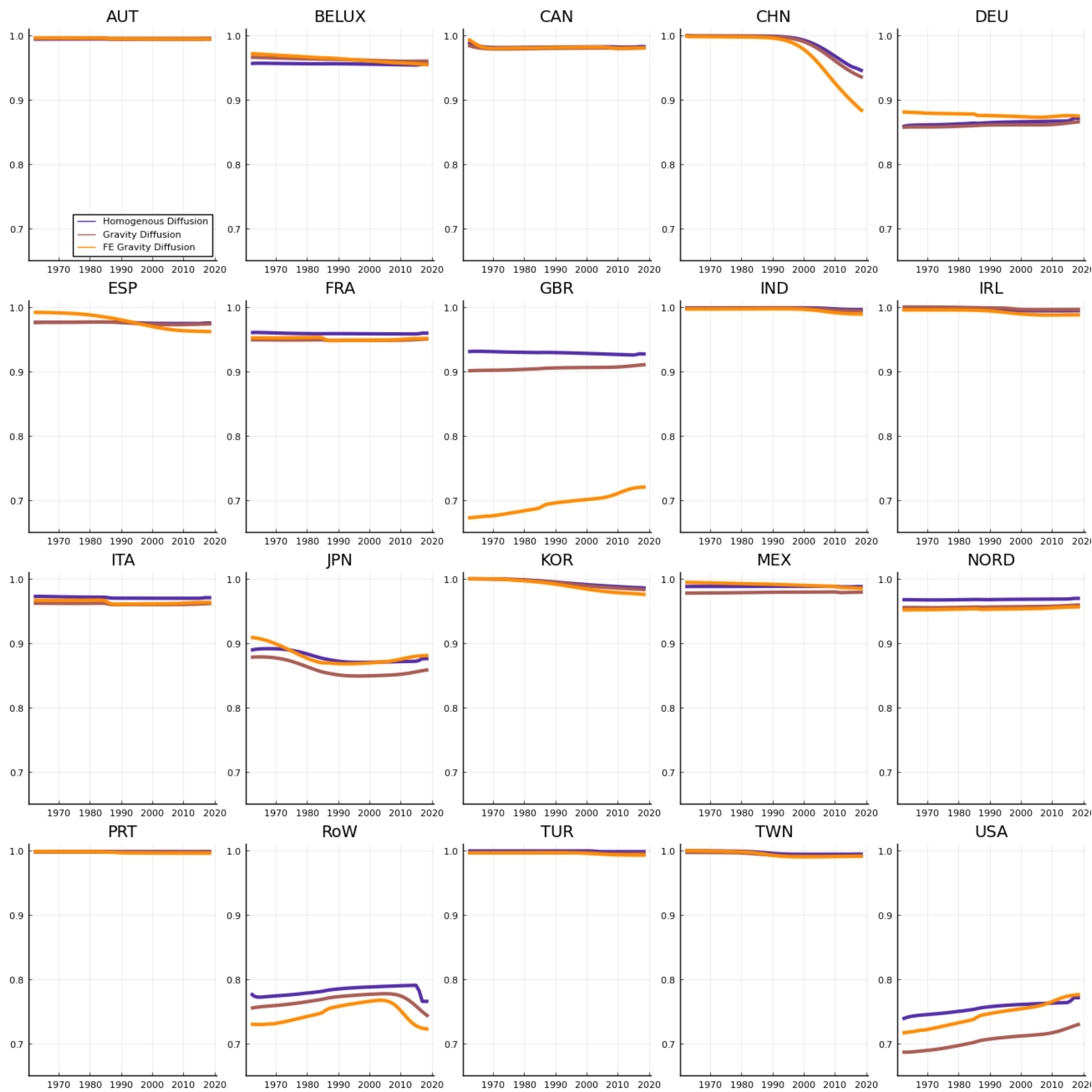
Notes: Results from estimating the model with two innovation surges, $S = 2$, and: no diffusion, $\delta_{nm} = 0$ for all $m \neq n$; homogenous international diffusion, $\delta_{nm} = \delta$ for all $m \neq n$; gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$; and FE gravity diffusion, $\delta_{nm} = \delta_m^A \delta_n^I \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$. In all cases, $\delta_{nn} = \infty$. The variable shown is the fixed effect $\kappa_d^{\tau}(t)$ in (18).

Figure B.4: Evolution of log innovated knowledge, by country.



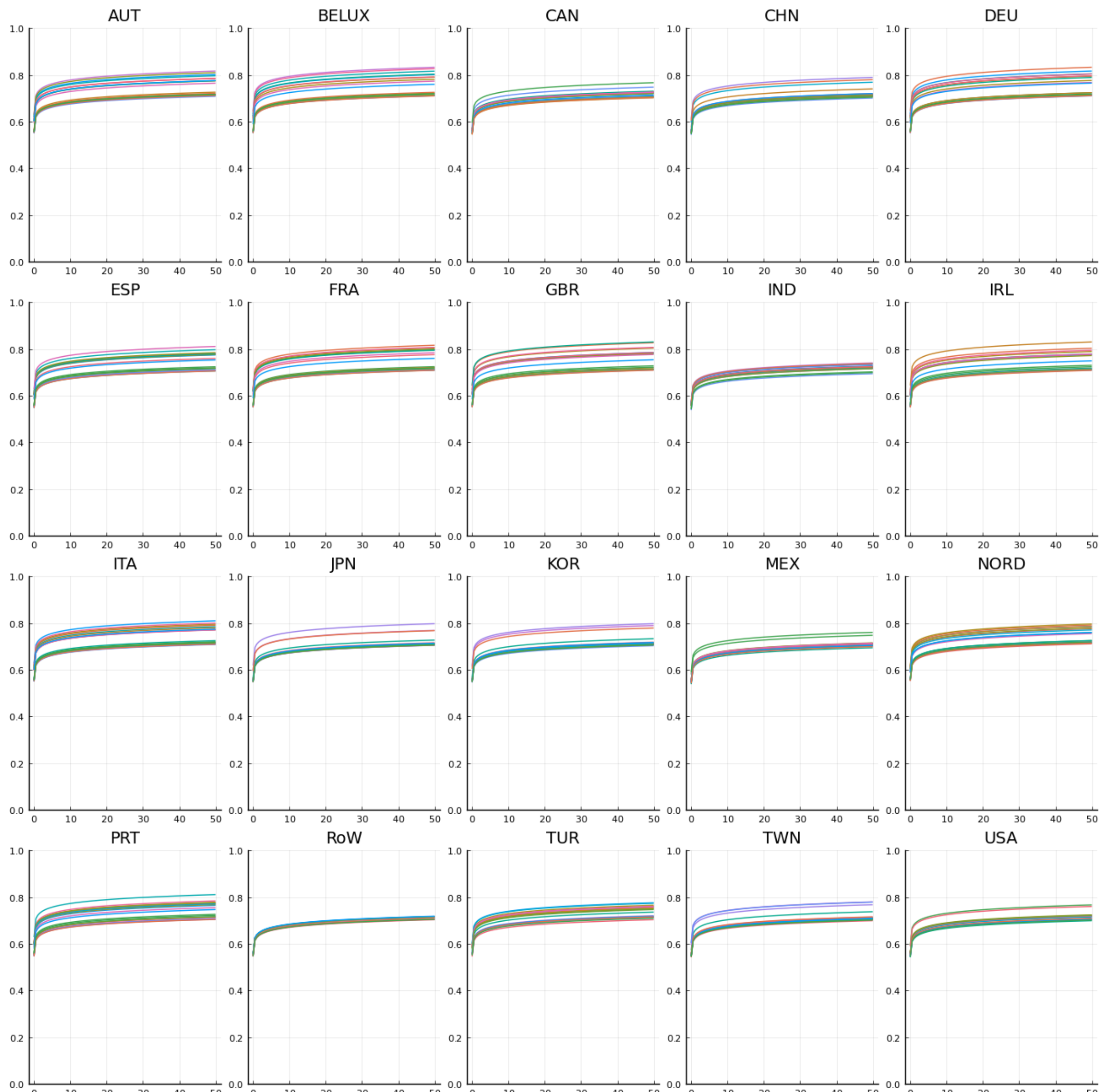
Notes: Results from estimating the model with two innovation surges, $S = 2$, and: no diffusion, $\delta_{nm} = 0$ for all $m \neq n$; homogenous international diffusion, $\delta_{nm} = \delta$ for all $m \neq n$; gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$; and FE gravity diffusion, $\delta_{nm} = \delta_m^A \delta_n^I \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$. In all cases, $\delta_{nn} = \infty$. The variable shown is $\ln \Lambda_n(t)$.

Figure B.5: Evolution of share of knowledge from diffusion, by country.



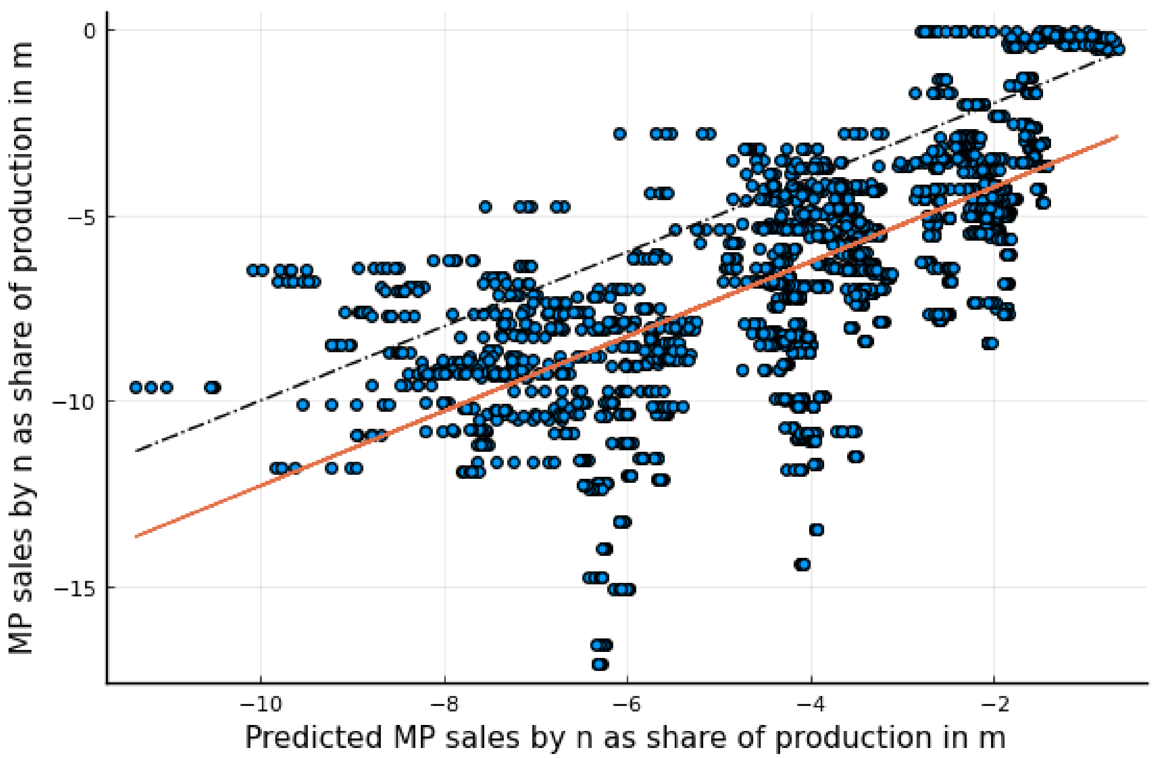
Notes: Results from estimating the model with two innovation surges, $S = 2$, and: homogenous international diffusion, $\delta_{nm} = \delta$ for all $m \neq n$; gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$; and FE gravity diffusion, $\delta_{nm} = \delta^{A}_{m}\delta^{I}_{n} \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$. In all cases, $\delta_{nn} = \infty$. The variable shown is $1 - \Lambda_m(t)/T_m(t)$.

Figure B.6: Diffusion curves, by country.



Notes: Results from estimating the model with two innovation surges, $S = 2$, and gravity diffusion $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$ and $\delta_{nn} = \infty$. Each subfigure corresponds to an innovator, n, and shows $p_{nm}(t - t^*)^{1-\rho}$ over time, for each adopter, m.

Figure B.7: Diffusion patterns and multinational production, robustness.



Notes: Results from estimating the model with two innovation surges, $S = 2$, and gravity diffusion, $\delta_{nm} = \delta \exp(-\kappa^{\delta} \ln \text{Dist}_{nm})$ for all $m \neq n$ and $\delta_{nn} = \infty$. $X_{nm}(t)$ is defined in (25) as a share of output in m. $\ln MP\ sales_{nm}$ are sales of affiliates operating in country m with parents in country n, from Ramondo et al. (2015). The dashed line is $45°$ line, while the solid line is best fit.