Global Knowledge and Trade Flows: Theory and Measurement*

Nelson Lind
Emory University

Natalia Ramondo
BU, NBER, and CEPR

September 28, 2022

Abstract

We study the global innovation and diffusion of ideas by introducing trade into the model in Eaton and Kortum (1999) (EK). This extension allows us to use international trade flows and country-level factor costs to estimate both the intensity of innovation within countries over time and diffusion rates across countries. We find significant specialization across the globe: some countries have high innovation rates, while other countries rely on diffusion. Although innovation is correlated with economic growth, there are many high income countries that primarily produce using diffused ideas. Additionally, these patterns shift over time — we estimate that a wave of innovation began in China during the early-2000's, reducing its reliance on diffused technology.

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. For instance, Edmund Cartwright sparked the industrial revolution by inventing the power loom in the United Kingdom during 1786, and over the next 100 years other places adopted his idea. And in the United States during 1959, Robert Noyce sparked the computer era by creating the first silicon microchip, which, nowadays, is widely spread across the globe. Although new ideas may ultimately transform day-to-day lives and bring economic growth to both creators and adopters, it is often hard to know about ideas until we observe their impact. Nevertheless, researchers have devised various direct and indirect ways of measuring knowledge. Some efforts relate the creation — and spread — of ideas to patent creation and citations (e.g. Keller, 2010; Akcigit et al., 2017); other 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). The early work by Eaton and Kortum (Eaton and Kortum, 1996a,b, 1997, 1999) belongs to the first strand of the literature since they use patent data — and the structure of their model — to measure the contribution of innovation and diffusion to growth.

In this paper, we use readily-available data on trade flows across countries and over time to uncover the global dynamics of knowledge. To such end, we extend the model of innovation and diffusion in Eaton and Kortum (1999) (henceforth EK) to incorporate international trade. Adding trade to the model allows us to use cross-country expenditure substitution patterns to detect shared knowledge and differentiate growth due to diffusion from growth due to innovation.

Specifically, we apply a result from our previous work that links a structure of innovation and diffusion to max-stable Fréchet productivity distributions (Lind and Ramondo, 2022). In the context of a Ricardian model of trade where sources compete head-to-head for markets, this result leads to closed-form solutions for trade shares across countries and ties substitution elasticities to diffusion patterns. Expenditure shares belong to the Generalized-Extreme-Value (GEV) class (McFadden, 1978), which does not impose Independence of Irrelevant Alternatives (IIA) and allows for rich substitution patterns across countries.[1] In this way, we obtain a transparent mapping between observable expenditure patterns and unobservable

[1] See, for instance, Adao et al. (2017) for evidence on departures of IIA in the data. Lind and Ramondo (2018) show the mapping between max-stable Fréchet and GEV expenditure.

knowledge flows across countries, which allows us to implement an estimation procedure based on international trade data.

We start by presenting the model. Ideas get discovered according to a Poisson process, are specific to each good, and have a unique discovery time and location. The efficiency of each idea is characterized by a global time-invariant component, "quality", and a location-specific component, "applicability." While quality among ideas discovered in a country up to a given time is distributed Pareto, we assume that applicability is independently and identically distributed Fréchet. By adding this random applicability component to the efficiency of each idea, we introduce some idiosyncratic differences in how countries can use the same idea. In the limit, if all countries had the same applicability for each idea, all countries would have equal ability to use ideas after they diffuse — which is the case in EK. But because we assume that the tail of the applicability distribution is thinner than the Pareto-tail for quality, applicability introduces small differences in productivity within each idea across countries relative to differences in productivity across ideas. In this way, we capture the essential spirit of EK where productivity levels are largely determined by the quality of ideas, and countries have similar ability to use diffused ideas.

After discovery, each idea may diffuse to other countries. As in EK, diffusion occurs within an idea, and, therefore, the strength of head-to-head competition in international trade depends on the extent of diffusion.[2] We further follow EK and assume that ideas diffuse exponentially over time and independently across countries, and that diffusion rates are innovator-adopter specific. While at the moment of discovery only the innovator has knowledge of the idea, over time, more countries learn about each idea. This creates a complex combinatorial problem as one needs to keep track of the sets of "who knows what" when calculating the global distribution of knowledge. However, under the exponential assumption, and conditional on an idea's discovery location and time, it turns out that the variable indicating whether or not a country knows an idea is a Bernoulli random variable independent across countries. This property allows us to get a simple integral formula for the distribution of productivity, rather than keeping track of the distribution of ideas across all possible subsets of countries.

Finally, we make the following assumption to capture short-run knowledge dy-

[2] A standard assumption is that diffusion occurs across ideas so that it does not increases head-to-head competition (e.g. Cai et al., 2022).

namics — and have more flexibility in the estimation later on. In addition to the EK assumption that the arrival of ideas in each country is proportional to the existing stock of ideas — an assumption that generates exponential growth — we introduce an additional exogenous source of discoveries in the form of waves of innovative activity. For each wave, there is a mass of ideas with discovery times distributed normal so that in each country, at each point in time, there are ideas of different "vintages."

Productivity in each country is the result of using, at each point in time, the most efficient idea available to them to produce each good. With this structure for knowledge, we can apply Theorem 1 in Lind and Ramondo (2022), which characterizes the conditions under which the global distribution of productivity is max-stable Fréchet with arbitrary correlation across countries.[3]

Thanks to productivity being distributed max-stable Fréchet, we are able to introduce Ricardian trade, get closed-form expressions for expenditure shares, and analyze how innovation and diffusion shape the observed patterns of trade across countries.[4] The key insight is that the sharing of ideas makes head-to-head competition between sources fiercer, resulting in more substitutable 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 use bilateral trade data and aggregate cost indices over time, together with standard geography data, to estimate the parameters of the EK-type knowledge model. In particular, we estimate bilateral diffusion rates, innovation rates, trade costs, and the parameters characterizing short-run waves of innovative activity.[5] Crucially, we do not need to use patent creation, patent citation, or any other R&D

[3]The assumption on a random country-specific applicability component, together with its independence from the quality component is key to this characterization — that is, it is equally likely that a low-quality ideas gets high applicability as a high-quality idea. Moreover, the result is an "if-and-only-if" statement so that one can always find a distribution of applicability, together with Poisson qualities, that by choosing the most efficient idea available results in a given max-stable Fréchet productivity distribution.

[4]As first presented by Eaton and Kortum (2002), an independent Fréchet distribution for productivity in the context of Ricardian trade leads to a closed-form solution for expenditure shares, which belong to the CES class. Lind and Ramondo (2018) show that max-stable Fréchet productivity with arbitrary correlation over space also leads to a closed-form solution for expenditure shares belonging to the much larger GEV class. The crucial property to obtain these results is not independence but max-stability.

[5]We treat these innovative waves as latent dimensions of the data, and estimate the number of waves that best fit the data.

data in the estimation.

In a nutshell, the estimation procedure boils down to estimating a non-CES import demand system, as many other papers in the trade literature do.[6] But thanks to the result in Lind and Ramondo (2022), we are able to connect substitution elasticities to primitive knowledge parameters and estimate them directly.

Using a multinomial maximum-likelihood procedure, similar to the one used in Eaton et al. (2013), our estimates suggest that over the last 60 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.[7] Interestingly, the trade data uncover a pattern of global knowledge where most rich (European) countries have a large share of knowledge coming from diffusion, with the United States acting as a primary innovator during all the sample period and Japan undergoing a wave of innovation that starts around 1960 and ends by 1980. Additionally, our estimates of innovated knowledge correlate with increases in researchers and researchers per capita in a country over time. Diffusion also correlates with more researchers over time — even if less so — suggesting that countries need some absorption capacity to adopt foreign knowledge, as in Nelson and Phelps (1966).

The paper is structured as follows. Section 2 presents the model of global knowledge and trade flows, and shows how a max-stable Fréchet global productivity distribution as well as GEV expenditure shares are obtained. Section 3 describes the estimation procedure, and Section 4 presents the results on the dynamics of knowledge implied by the observed trade flows across countries. Section 5 concludes.

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

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

2 Model

2.1 Preliminaries

The global economy consists of N countries that produce and trade a continuum of goods $v \in [0,1]$. Time is continuous and indexed by t. We generally index countries by n, but also o when they are the origin location where goods are produced, and d when they are a destination market. At each moment in time, country n is populated by an exogenous measure of households $L_n(t)$ 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 destination market n is produced with labor and a composite input using a Cobb-Douglas technology with labor share of $1-\beta \in (0,1)$. Hence, the price of the final good is

$$P_n^f(t) = W_n(t)^{1-\beta} P_n(t)^{\beta}. \tag{1}$$

Here, $W_n(t)$ is the wage and $P_n(t)$ is the price index for the composite input.

The composite input is produced from a continuum of tradable intermediate goods $v \in [0,1]$ using a CES technology with elasticity of substitution $\eta \geq 1$. Hence, the price of this good is

$$P_n(t) = \zeta \left(\int_0^1 P_n(t,v)^{1-\eta} \mathrm{d}v \right)^{\frac{1}{1-\eta}}. \tag{2}$$

Here, ζ is a constant that we will use for a global scale normalization, and $P_n(t,v)$ is the price of good v in n.

In turn, each intermediate v is produced under perfect competition using labor and the composite input. The production function is Cobb-Douglas with productivity $Z_n(v,t)$ specific to each good, and labor share $1-\alpha \in (0,1)$. The marginal cost to produce input v is then $C_n(t)/Z_n(v,t)$ where

$$C_n(t) = W_n(t)^{1-\alpha} P_n(t)^{\alpha} \tag{3}$$

indexes the cost to produce intermediates.

Productivity, $Z_n(t,v)$, results from the adoption of ideas. We next describe how the innovation and diffusion of ideas determines productivity over space and time.

2.2 Innovation and diffusion

As in EK, for each good v, there exists an infinite, but countable, set of ideas $i = 1, 2, \ldots$. Idea i applied to the production of good v gets discovered in a unique country at a unique moment in time. We denote the innovator country where the idea is first discovered by $n_i(v)$ and its discovery time by $t_i^*(v)$. We assume that 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 $\theta > 0$. The arrival of ideas in n over time is controlled by $\lambda_n(t^*)$, while $\theta q^{-\theta-1}\mathrm{d}q$ controls the arrival rate of low versus high quality ideas. The Poisson assumption means that the expected number of ideas discovered in n up to time t with quality above $\underline{q}$ is given by $\underline{q}^{-\theta}\Lambda_n(t)$ 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 $\underline{q}$ and tail parameter θ.

Similarly to EK, we assume that the arrival of ideas in each country is proportional to the existing stock of ideas, with the time-invariant country-specific proportion denoted by γ_n. This assumption generates arrival of new ideas over time inspired by old ideas, some potentially from the distant past. For example, the idea that gave rise to the power loom was discovered long before our sample, but may have inspired many ideas over time. Alone, this assumption generates exponential growth.

However, to capture periods of high innovation, we introduce an additional exogenous source of discoveries. We assume that in each country n there are S surges of innovation. For each surge s, there is a mass λ_{ns} of ideas with discovery times distributed normal with mean μ_{ns} and standard deviation ν_{ns}. 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 additional assumption is made to add flexibility in matching the trade data in our estimation procedure, and introduces a "latent" dimension underlying the data.

With this assumption, the intensity of innovation in country n at time t is

$$\lambda_n(t) = \Lambda_n'(t) = \gamma_n \Lambda_n(t) + \sum_{s=1}^{S} \lambda_{ns} \frac{1}{\nu_{ns}} \phi\left(\frac{t - \mu_{ns}}{\nu_{ns}}\right), \tag{4}$$

where $\phi(x) \equiv (2\pi)^{-1/2}e^{-x^2/2}$. The first term on the right-hand side of (4) 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. If there are no innovation surges, we are back to the model with only exponential growth and $\lambda_n(t) = \gamma_n \Lambda_n(t)$.

Given an initial condition, $\Lambda_n(0)$, we show in the Appendix that the solution to the differential equation in (4) takes the form

$$\Lambda_n(t) = \Lambda_{n0}(t) + \sum_{s=1}^{S} \Lambda_{ns}(t) \tag{5}$$

where $\Lambda_{n0}(t) = \lambda_{n0} e^{\gamma_n t}$ captures the stock of ideas inspired by the distant past, and

$$\Lambda_{ns}(t) \equiv \lambda_{ns} \int_{-\infty}^{t} e^{\gamma_n (t-t^*)} \frac{1}{\nu_{ns}} \phi\left(\frac{t^* - \mu_{ns}}{\nu_{ns}}\right) \mathrm{d}t^* \quad \text{for each } s = 1, \dots, S \tag{6}$$

captures ideas inspired by each surge of innovation.[8] For the rest of the paper, we indicate ideas inspired by the distant past by $s = 0$, and have $s \in \{0, \dots, S\}$.

After discovery, each idea may diffuse to other countries. Let $\mathcal{N}_i(t, v)$ denote the set of countries with knowledge of idea i for good v at time t. At the moment of discovery, only the innovator has knowledge of the idea so that $\mathcal{N}_i(t, v) = \{n_i(v)\}$. Over time, as more countries learn about each idea, the set $\mathcal{N}_i(t, v)$ expands.

For each country in the set $\mathcal{N}_i(t, v)$, the idea's efficiency is given by $Q_i(v) A_{io}(v)$, where $Q_i(v)$ is the global quality component and $A_{io}(v)$ is the country-specific (random) component to which we refer to as the idea's applicability in o.[9] When $\mathcal{N}_i(t, v) = \{n_i(v)\}$, $A_{io}(v)$ is positive only for $o = n$ (i.e. the innovator). As ideas diffuse, applicability $A_{io}(v)$ turns positive for other countries as well, which get included in the set $\mathcal{N}_i(t, v)$.

We assume that applicability is independently and identically distributed unit Fréchet with shape $\sigma > \theta$ across goods and countries. As $\sigma \to \infty$, $A_{io}(v)$ becomes degenerate at 1 and productivity is simply equal to $Q_i(v)$. This limiting case corresponds to the setup in EK where all countries have equal ability to use ideas after they diffuse. By allowing $\sigma < \infty$, we introduce some idiosyncratic differences in how countries can use the same idea. Note that because we assume applicability

[8] Here, $\lambda_{n0} \equiv \Lambda_n(0) - \lambda_{ns} \int_{-\infty}^{0} e^{-\gamma_n t^*} \frac{1}{\nu_{ns}} \phi\left(\frac{t^* - \mu_{ns}}{\nu_{ns}}\right) \mathrm{d}t^*$ is the stock of ideas that exist at $t = 0$ and were inspired by the distant past. This stock is simply all ideas at $t = 0$ net of those ideas that exist at $t = 0$ but are associated with surges in innovation.

[9] This is a special case of Lind and Ramondo (2022) where we assume more generally that applicability is not only country specific, but also that it can change over time.

is distributed Fréchet, it has a Pareto tail with shape σ, and, because we assume that $\sigma > \theta$, this tail is thinner than the Pareto tail for quality. Relative to differences in productivity across ideas, applicability introduces small differences in productivity within an idea across countries. In this way, our setup captures the essential spirit of EK where productivity differences are largely determined by the quality of ideas, and countries have similar ability to use diffused ideas.

We further follow EK and assume that ideas innovated in n diffuse exponentially over time and independently across countries. The probability that o has learned an idea by time t discovered in surge s in country n at time t^* is

$$p_{ons}(t - t^*) \equiv 1 - e^{-\delta_{ons}(t-t^*)}, \tag{7}$$

and $p_{nns}(t - t^*) = 1$. Under this assumption, and conditional on an idea's discovery location, surge, and time, the indicator variable of o knowing the idea, $\mathbf{1}\{o \in \mathcal{N}_i(t, v)\}$, is a Bernoulli random variable with success probability $p_{ons}(t-t^*)$. As we show next, this property will allow us to get a simple integral formula for the distribution of productivity, rather than keeping track of the distribution of ideas across all possible subsets of countries.

2.3 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 o at time t for production of good v is

$$Z_o(t, v) = \max_{\{i=1,2,\cdots|o\in\mathcal{N}_i(t,v),t_i^*(v)\leq t\}} Q_i(v)A_{io}(v). \tag{8}$$

When a country gains access to an idea, they enter the set $\mathcal{N}_i(t, v)$. 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 high applicability in their country. In this way, the dynamics of productivity are entirely driven by two forces: innovation can increase productivity through the introduction of new ideas in their discovery location, and diffusion can increase productivity by expanding access to previously discovered ideas.

This structure 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 (2022), which provides a constructive method to derive the productivity distribution. Appendix A shows the derivation in detail. In what follows, we provide a sketch of the result and some intuition.

First, from the assumption in (8) on the technology adoption process, we can write the joint distribution of productivity as

$$F(z_1, \ldots, z_N; t) = \mathbb{P}\left[Q_i(v) > \min_{o \in \mathcal{N}_i(t)} \frac{z_o}{A_{io}(v)} \text{ for no } i \text{ s.t. } t_i^*(v) \leq t\right]. \tag{9}$$

This expression states that the quality of each idea cannot be too high. If a country had knowledge of a single idea with quality above $z_o/A_{io}(v)$, then they could achieve productivity above z_o. The probability that productivity lies below z_o in all countries is then the probability that there are no ideas with too-high quality.

Next, because innovation follows a Poisson process, we can calculate this probability using the expected number of ideas with quality above the lower-bound in (9).[10] It is useful to first calculate the marginal distribution of productivity in o. Letting $z_{o'} \to \infty$ in (9) for all $o' \neq o$ yields an expression for this marginal distribution:

$$\mathbb{P}[Z_o(t) \leq z_o] = \mathbb{P}\left[Q_i(v)A_{io}(v) > z_o \text{ for no } i \text{ s.t. } o \in \mathcal{N}_i(t, v) \text{ and } t_i^*(v) \leq t\right]. \tag{10}$$

Simply, productivity is below z_o if there are no ideas known to o with higher idea-level productivity. As a consequence, we can focus on the collection of ideas known to o with productivity above z_o,

$$\mathcal{I}_o(z_o, t, v) \equiv \{i = 1, 2, \ldots \mid Q_i(v)A_{io}(v) > z_o, o \in \mathcal{N}_i(t, v), t_i^*(v) \leq t\}. \tag{11}$$

If an idea either has a sufficiently high quality or applicability in o, it enters this set after o learns it.

We can then calculate the marginal distribution in (10) using the expected number of ideas in $\mathcal{I}_o(z_o, t)$:[11]

$$\mathbb{P}[Z_o(t) \leq z_o] = \exp\left(-\mathbb{E}\sum_{i=1}^{\infty} \mathbf{1}\{i \in \mathcal{I}_o(z_o, t, v)\}\right) = \exp\left[-T_o(t) z_o^{-\theta}\right],$$

[10] For a Poisson process with a countable number of points, $\{x_i\}_{i=1,2,\ldots}$, and any set X, the void probability $\mathbb{P}[x_i \in X \text{ for no } i] = \exp\left(-\mathbb{E}\sum_{i=1}^{\infty} \mathbf{1}\{x_i \in X\}\right)$.

[11] Letting $s_i(v)$ denote the surge in which i was discovered, we can calculate this expectation by

for

$$T_o(t) \equiv \Gamma\left(1 - \frac{\theta}{\sigma}\right) \sum_{s=o}^{S} \sum_{n=1}^{N} \int_{-\infty}^{t} p_{ons}(t - t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*. \tag{12}$$

That is, the marginal distribution is Fréchet with shape θ and scale $T_o(t)$, which is simply the expected number of ideas known to o with productivity above 1.

We can also use the collections of high productivity ideas in (11) to calculate the joint distribution of productivity. In particular, because applicability is continuously distributed, there is a unique country that attains the minimum in (9), and we can re-write (9) as

$$F(z_1, \ldots, z_N; t) = \exp\left[-\mathbb{E}\sum_{o=1}^{N} \sum_{i \in \mathcal{I}_o(z_o,t,v)} \mathbf{1}\left\{A_{io'}(v) \leq \frac{z_{o'}}{z_o} A_{io}(v), \;\; \forall o' \in \mathcal{N}_i(t,v)\right\}\right]. \tag{13}$$

Since a unique country attains the minimum, we can partition all ideas by conditioning on whether a country, say o, out-competes all other countries. Among ideas known to o with high productivity, they attain the minimum in (9) if all other countries with knowledge of the idea have applicability below $(z_{o'}/z_o)A_{il}(v)$. Note that competition within each idea happens purely in terms of applicability because the quality of an idea is common across countries. After counting up the number of ideas for which o beats all competitors within $\mathcal{I}_o(z_o, t)$, we can simply sum over o and get the expected number of ideas with quality above the bound in (9).

As a consequence, the joint distribution of productivity follows from answering: Given the ideas with high productivity that a country knows, how often does the country out-compete the rest of the world due to cross-country differences in applicability? In general, the answer depends on calculating the probability that an idea is common knowledge among every possible group of countries, $\mathcal{N}_i(t,v)$, and then summing over all groups. As the number of countries grows, this sum has an exponentially increasing number of terms.

To overcome this curse of dimensionality, we leverage the structure of diffusion

summing the following result over surges and innovators:

$$\mathbb{E}\sum_{i=1}^{\infty} \mathbf{1}\left\{Q_i(v) > \frac{z_o}{A_{io}(v)}, s_i(v) = s, n_i(v) = n, t_i(v) \leq t\right\} = \int_{-\infty}^{t} \mathbb{E}\int_{z_o/A_{io}(v)} \theta q^{-\theta-1}\mathrm{d}q p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*$$

$$= \int_0^\infty a_o^\theta e^{-a_o^{-\sigma}} \sigma a_o^{-\sigma-1}\mathrm{d}a_o \int_{-\infty}^{t} p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^* z_o^{-\theta} = \Gamma(1-\theta/\sigma)\int_{-\infty}^{t} p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^* z_o^{-\theta}.$$

proposed in (7). In particular, after conditioning on an idea's discovery location, surge, and time, the probability that o knows the idea is independent across o and equal to $p_{o'ns}(t-t^*)$ in (7). Then, given applicability in o, the chance they outcompete o' is given by $\exp[-(\frac{z_{o'}}{z_o}A_{io}(v))^{-\sigma}]$ since applicability is independent across countries and distributed unit-Fréchet with shape σ. All together, among ideas with productivity above z_o that were discovered at time t^* in surge s of n and known to o by time t, the probability that o beats all other countries is

$$H_{ons}(z_1,\ldots,z_N;t-t^*) = \int_0^\infty \prod_{o'\neq o}\left[1-p_{o'ns}(t-t^*)\left(1-e^{-\left(\frac{z_{o'}}{z_o}a_o\right)^{-\sigma}}\right)\right]\mathrm{d}F_A(a_o). \quad (14)$$

Here, the variable a_o represents a level of applicability in o and F_A denotes the distribution of applicability among high-quality ideas.[12] The term in brackets is the probability that o beats o', which equals the probability that either o' does not know the idea or has applicability below the threshold in (13). Due to independence of diffusion across countries, the product of these terms integrated over applicability levels gives the probability that no country beats o.

Using this result, we can calculate the expected value in (13) to get the joint distribution of productivity,

$$F(z_1,\ldots,z_N;t) = \exp\left[-\sum_{o=1}^N H_o(z_1,\ldots,z_N;t)T_o(t)z_o^{-\theta}\right], \quad (15)$$

where

$$H_o(z_1,\ldots,z_N;t) \equiv \frac{\sum_{s=0}^S\sum_{n=1}^N\int_{-\infty}^t H_{ons}(z_1,\ldots,z_N;t-t^*)p_{ons}(a)\Lambda'_{ns}(t^*)\mathrm{d}t^*}{\sum_{s=0}^S\sum_{n=1}^N\int_{-\infty}^t p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*} \quad (16)$$

is the the probability that, among all ideas known to o with productivity above z_o, no country has higher productivity.[13]

[12]The distribution of applicability among ideas known to o with productivity above z_o is

$$\mathbb{P}\left[A_{io}(v)\leq a_o \mid i\in\mathcal{I}_o(z_o,t)\right] = \frac{\sum_{s=0}^S\sum_{n=1}^N\int_{-\infty}^t\int_0^{a_o}\int_{z_o/x}\theta q^{-\theta-1}\mathrm{d}q\mathrm{d}e^{-x^{-\sigma}}p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*}{T_o(t)z_o^{-\theta}}$$

$$= \frac{\int_0^{a_o}x^\theta e^{-x^{-\sigma}}\sigma x^{-\sigma-1}\mathrm{d}x z_o^{-\theta}\sum_{s=0}^S\sum_{n=1}^N\int_{-\infty}^t p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*}{T_o(t)z_o^{-\theta}} = \int_0^{a_o}\frac{e^{-x^{-\sigma}}x^{\theta-\sigma-1}}{\Gamma(1-\theta/\sigma)}\mathrm{d}x \equiv F_A(a_o).$$

[13]The implied correlation function for productivity, as in Lind and Ramondo (2018), is $G(x_1,\ldots,x_N) = \sum_{o=1}^N H_o\left((x_1/T_1(t))^{-1/\theta},\ldots,(x_N/T_N(t))^{-1/\theta};t\right)x_o$. See Appendix A for details.

To gain some intuition for the expression in (15), suppose first that there is no diffusion of ideas so that $p_{ons}(a) = \mathbf{1}\{o = n\}$. In this case, the expression in (16) collapses to $H_o(z_1, \ldots, z_N; t) = 1$. When ideas are specific to each country, a country always has higher productivity among the ideas they know. The joint distribution in (15) reduces to $\exp(-\sum_{o=1}^{N} T_o(t) z_o^{-\theta})$, which is the special case of independent Fréchet productivity.

The function $H_o(z_1, \ldots, z_N; t)$ captures how diffusion generates departures from the case of independence because other countries sometimes will have higher productivity for a diffused idea. As an example, suppose that for each innovator n there is a group of technological peers $\mathcal{N}_n$ to which ideas instantaneously diffuse: $p_{ons}(t - t^*) = \mathbf{1}\{o \in \mathcal{N}_n\}$. Then the joint productivity distribution is cross-nested CES,

$$F(z_1, \ldots, z_N; t) = \exp\left[-\Gamma(\rho) \sum_{n=1}^{N} \left(\sum_{o \in \mathcal{N}_n} z_o^{\frac{\theta}{1-\rho}}\right)^{1-\rho} \Lambda_n(t)\right],$$

with correlation parameter $\rho \equiv 1 - \theta/\sigma$ and each nest corresponding to an innovator.[14] In this example, correlation in productivity arises from similarity in applicability across countries that learn ideas from the same innovator.[15]

Finally, consider the limiting case when $\sigma \to \infty$, so that the distribution of applicability becomes degenerate at 1 and, as in EK, each idea has identical productivity across countries. This limiting case corresponds to a multivariate Fréchet distribution with perfect correlation in productivity among countries with knowledge of an idea. In this case, the productivity distribution in (15) is not continuously differentiable. For instance, in the previous cross-nested CES example, we get $\rho \to 1$, and the CES aggregator inside each nest converges to a max function. Without continuous differentiability, we cannot apply the results in Lind and Ra-

[14] The scale is $T_o(t) = \Gamma(1 - \theta/\sigma) \sum_{n=1}^{N} \mathbf{1}\{o \in \mathcal{N}_n\} \Lambda_n(t)$,

$$H_{ons}(z_1, \ldots, z_N; t - t^*) = \int_0^\infty \prod_{o' \in \mathcal{N}_n} e^{\left(\frac{z_{o'}}{z_o} x\right)^{-\sigma}} \frac{e^{-x^{-\sigma}} x^{\theta - \sigma - 1}}{\Gamma(1 - \theta/\sigma)} \mathrm{d}x = z_o^{\theta - \sigma} \left(\sum_{o' \in \mathcal{N}_n} z_{o'}^{-\sigma}\right)^{\frac{\theta}{\sigma} - 1},$$

and

$$H_o(z_1, \ldots, z_N; t) T_o(t) z_o^{-\theta} = \Gamma(1 - \theta/\sigma) \sum_{n=1}^{N} \left(\sum_{o' \in \mathcal{N}_n} z_{o'}^{-\sigma}\right)^{\frac{\theta}{\sigma} - 1} \mathbf{1}\{o \in \mathcal{N}_n\} \Lambda_n(t) z_o^{-\sigma}.$$

[15] This example corresponds to a special case of the multinational production model in Ramondo and Rodríguez-Clare (2013) when multinational production costs between each home country (n) and production location (o) are either absent or infinite.

mondo (2018) that map max-stable Fréchet productivity to Generalized-Extreme-Value (GEV) expenditure shares. But for $\sigma < \infty$, applicability introduces small idiosyncratic differences in productivity across countries within each idea, breaks the case of perfect correlation, and allows us to obtain closed-form solutions for expenditure shares.

2.4 Expenditure shares

We next present results for prices and expenditure shares across countries. We assume that trade is subject to iceberg-type trade costs: $\tau_{od}(t) \geq 1$ with $\tau_{oo}(t) = 1$. Define the import cost index $P_{od}(t) \equiv \tau_{od}(t)C_o(t)$. Then, the unit cost of good v in d at time t when sourced from o is $P_{od}(t)/Z_o(t,v)$. 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

$$P_d(t,v) = \min_{o=1,\ldots,N} \frac{P_{od}(t)}{Z_o(t,v)}. \tag{17}$$

Since productivity is max-stable multivariate Fréchet, aggregate expenditure shares have a closed-form solution (Eaton and Kortum, 2002). Following the results in Lind and Ramondo (2018) and using (15), the expenditure share by d on goods from o at time t is

$$\pi_{od}(t) = \frac{H_o(P_{1d}(t),\ldots,P_{Nd}(t);t)T_o(t)P_{od}(t)^{-\theta}}{\sum_{o'=1}^{N} H_{o'}(P_{1d}(t),\ldots,P_{Nd}(t);t)T_{o'}(t)P_{o'd}(t)^{-\theta}}, \tag{18}$$

where the denominator is $P_d(t)^{-\theta}$ (after using the normalization $\zeta = \Gamma(\frac{1+\theta-\eta}{\theta})^{\frac{1}{\eta-1}}$ and assuming that $1+\theta-\eta > 0$). These expenditures belong to the GEV class, a class that allows for departures from IIA and generates rich patterns of substitution across countries.

As shown in Lind and Ramondo (2022), diffusion determines these substitution patterns. When ideas cannot diffuse across countries, $H_o(z_1,\ldots,z_N;t) = 1$, and the expenditure share in (18) is CES:

$$\pi_{od}(t) = \frac{T_o(t)P_{od}(t)^{-\theta}}{\sum_{o'=1}^{N} T_{o'}(t)P_{o'd}(t)^{-\theta}}. \tag{19}$$

Once we add some diffusion, expenditure becomes non-CES.

To see this clearly, we next calculate the elasticities of substitution for destination market d between sources o and o' at time t, $\varepsilon_{oo'd}(t) \equiv \frac{\partial \ln \pi_{od}(t)}{\partial \ln(P_{o'd}(t)/P_d(t))}$ (see Appendix A for derivations):

$$\varepsilon_{oo'd}(t) = \frac{\sigma - \theta}{\pi_{od}(t)} \frac{P_{od}(t)^{-\sigma} P_{o'd}(t)}{(P_{od}(t)^{-\sigma} + P_{o'd}(t)^{-\sigma})^2} \left(\frac{P_{od}(t)^{-\sigma} + P_{o'd}(t)^{-\sigma}}{P_d(t)^{-\sigma}} \right)^{\frac{\theta}{\sigma}} Q_{oo'd}(t) K_{oo'}(t). \tag{20}$$

Starting from the right, the term

$$K_{oo'}(t) \equiv \Gamma\left(\frac{\sigma-\theta}{\sigma}\right) \sum_{s=0}^{S} \sum_{n=1}^{N} p_{ons}(a) p_{o'ns}(a) \Lambda'_{ns}(t-a) \mathrm{d}a \tag{21}$$

captures the number of ideas that both o and o' know. Not surprisingly, this quantity increases with the bilateral diffusion rates of the two countries. In the Appendix, we show that we can fully solve this expression in terms of the parameters of the model. The next term,

$$Q_{oo'd}(t) \equiv \sum_{s=0}^{S} \sum_{n=1}^{N} \int_0^\infty \int_0^\infty \left[\prod_{m \notin \{o,o'\}} \left(1 - p_{mns}(a)\left(1 - e^{-\frac{P_{md}(t)^{-\sigma}}{P_{od}(t)^{-\sigma} + P_{o'd}(t)^{-\sigma}} x}\right)\right)\right] \frac{e^{-x} x^{1-\frac{\theta}{\sigma}}}{\Gamma\left(2 - \frac{\theta}{\sigma}\right)} \mathrm{d}x$$
$$\times \frac{p_{ons}(a) p_{o'ns}(a) \Lambda'_{ns}(t-a)}{\sum_{s'=0}^{S} \sum_{n'=1}^{N} p_{on's'}(a) p_{o'n's'}(a) \Lambda'_{n's'}(t-a) \mathrm{d}a} \mathrm{d}a$$

is the probability that no other country m can compete with either o or o' among those ideas. The term $\left(\frac{P_{od}(t)^{-\sigma} + P_{o'd}(t)^{-\sigma}}{P_d(t)^{-\sigma}}\right)^{\frac{\theta}{\sigma}}$ is the probability that there are no other ideas that can be used to supply market d at lower cost. Finally, $\frac{P_{od}(t)^{-\sigma} P_{o'd}(t)}{(P_{od}(t)^{-\sigma} + P_{o'd}(t)^{-\sigma})^2}$ captures the strength of head-to-head competition between o and o' in market d in terms of input costs.

2.5 Equilibrium

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.

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

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

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 $\sum_{o=1}^N \xi_o(t) = 0$.

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

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

Given labor endowments $\{L_o(t)\}_{o=1,\ldots,N,t\in[0,T]}$, and trade imbalances $\{\xi_d(t)\}_{d=1,\ldots,N,t\in[0,T]}$, an equilibrium consists of paths for wages, cost indices, price levels, incomes, and intermediate-good expenditures, $\{W_o(t), C_o(t), P_o(t), Y_o(t), X_o(t)\}_{o=1,\ldots,N,t\in[0,T]}$, and trade shares $\{\pi_{od}(t)\}_{o,d=1,\ldots,N,t\in[1,T]}$, such that

1. Prices satisfy (1), (2) and (17), for each o and t;
2. Cost indices satisfy (3) for each o and t;
3. Trade shares satisfy (18) for each o, d, and t;
4. Labor market, final-good market, and intermediate-good market clear for each o and t.

3 Estimation Procedure

We use data from the Penn World Table 10.0 (PWT) on current GDP (CGDPO), employment, the value of imports and exports, as well as the price and value of domestic absorption for each country, from 1962 to 2019. Additionally, we use data on trade flows between countries and the same time period from COMTRADE. We construct self-trade shares for each country using the model equilibrium conditions in Section 2.5. Our sample contains 19 countries plus an aggregate of the rest of the world. Appendix Table B.1 reports the country names.

First, we set $\beta = \alpha = 0.5$. Second, we construct the cost index $C_o(t)$ using the expression in (3). For wages $W_o(t)$, we use data on current GDP per worker, while we calculate the intermediate price index $P_o(t)$ using data on the price of domestic absorption and the expression in (1). Third, we assume that trade costs are a function of geographical distance, and border costs that follow quadratic time trends,

$$\tau_{od}(t) = \mathbf{1}\{o \neq d\} \exp\left(\kappa_d^0 + \kappa_d^1 t + \kappa_d^2 t^2 + \kappa^3 \ln Dist_{od}\right). \tag{22}$$

Finally, we use numerical quadrature to approximate the integrals in (14) and (16).

We estimate bilateral diffusion rates, δ_{ons}, innovation rates λ_{ns} for each $s = 0, \ldots, S$, (Normal) distribution parameters, μ_{ns} and ν_{ns}, for each $s = 1, \ldots, S$, the trade cost parameters κ's in (22), and the parameters θ and σ. We estimate the parameters of the model using a multinomial pseudo maximum-likelihood procedure, similar to Eaton et al. (2013), to match bilateral trade shares in the data. This procedure minimizes the following criterion:

$$\sum_{t=1}^{T} \sum_{d=1}^{N} \sum_{o=1}^{N} \pi_{od}(t) \ln \frac{\pi_{od}(t)}{\hat{\pi}_{od}(t)}.$$

We also target the composite price for intermediates by adding a penalty for the log-squared distance between data and model,

$$\sum_{t=1}^{T} \sum_{d=1}^{N} \left[\sum_{o=1}^{N} \left(\ln P_d(t) - \ln \hat{P}_d(t) \right) \right]^2.$$

We perform the estimation for the model with one innovation surge, $S = 1$, and two innovation surges, $S = 2$. For comparison purposes, we also estimate a version of the model that corresponds to gravity estimation of CES expenditure (i.e. the no diffusion case), treating the scales, $T_o(t)$, as parameters. This CES case corresponds to multinomial pseudo maximum likelihood gravity estimates, as in Eaton et al. (2013), with origin-time fixed effects.[16]

Table 1 shows statistics on fit for the CES model, and models with one innovation surge, $S = 1$, and two innovation surges, $S = 2$. Relative to CES, the model with $S = 1$ is more parsimonious — 923 vs. 1,221 parameters — yet still fits the data better according to all the measures of goodness of fit. In contrast, the model with $S = 2$ is less parsimonious than CES — 1,363 vs. 1,221 parameters — and it only provides a modest additional improvement in fit over the one-surge model. For

[16] Due to the result in Sotelo (2019) on the numerical equivalence of multinomial pseudo maximum likelihood and poisson pseudo maximum likelihood with destination fixed effects, this case also corresponds to standard gravity estimates with origin-time and destination-time fixed effects using poisson pseudo maximum likelihood, as in Silva and Tenreyro (2006).

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

Models	CES	$S = 1$	$S = 2$
θ		0.201	0.239
σ		1.405	1.574
Distance Elasticity, κ^3	0.787	1.578	1.433
Number of Observations	23,200	23,200	23,200
Number of Parameters	1,221	923	1,363
Trade Share R-Squared	0.983	0.991	0.992
Price Level R-Squared	0.964	0.983	0.986
Trade Share KL Divergence	148.26	35.61	34.80
Price Level Mean Squared Error	0.0127	0.0061	0.0048
Total Loss	163.0	42.67	40.38

Notes: Results from estimating the CES model, and the models with one ($S = 1$) and two ($S = 2$) surges. Note that the distance elasticity, κ^3, in (22) is the elasticity of trade costs to distance. The elasticity of trade shares to distance is not constant and depends on the values of the other parameters.

this reason, our preferred model is the one with one innovation surge, $S = 1$. This estimation yields a high correlation coefficient in productivity, $\rho \equiv 1 - \theta/\sigma = 0.92$, while the elasticities of trade costs with respect to distance are estimated at 1.6. Appendix Table B.1 shows the estimates of country-level variables related to the innovation process (γ_n, λ_{n0}, λ_{n1}, μ_{n1} and ν_{n1}).

How does the one-surge model fit the data better than CES despite having fewer parameters? There are two key differences between the models. First, the one-surge model is more parsimonious because the CES model uses origin-time fixed effects to capture the scale of productivity for each country, $T_o(t)$. In contrast, these variables are restricted by the structure of innovation and diffusion in the one-surge model as indicated by (12). Second, the one-surge model allows for departures of IIA through the diffusion of ideas, while the CES model imposes IIA.

To examine this mechanism further, we use the following index of exposure to third-party unit costs to detect departures from IIA,

$$\text{Exposure}_{od}(t) = \sum_{o' \neq o} \ln \frac{C_{o'}(t)}{C_o(t)} \frac{\pi_{o'd}(t)}{\sum_{n \neq o} \pi_{nd}(t)}. \tag{23}$$

This index is an expenditure share-weighted average of relative unit costs between o and each of its competitors. If the CES model is correctly specified, this measure should be insignificant when included in a regression of observed trade shares on

Table 2: CES and departures from IIA. OLS.

	$\ln \pi_{od}(t)$			
	(1)	(2)	(3)	(4)
$\ln \hat{\pi}_{od}^{\text{CES}}(t)$	1.109***	1.097***	1.014***	-0.029
	(0.032)	(0.033)	(0.016)	(0.031)
$\text{Exposure}_{od}(t)$		-0.308**	-0.041	-0.041
		(0.115)	(0.050)	(0.050)
$\ln \frac{\hat{\pi}_{od}^{S=1}(t)}{\hat{\pi}_{od}^{\text{CES}}(t)}$			1.043***	
			(0.024)	
$\ln \hat{\pi}_{od}^{S=1}(t)$				1.043***
				(0.024)
Observations	22,601	22,601	22,601	22,601
R-squared	0.532	0.537	0.872	0.872

Notes: $\hat{\pi}_{od}^{m}(t)$ denotes predicted trade shares by model m, with $m = CES, S = 1$. The index of exposure to third-party unit costs $\text{Exposure}_{od}(t)$ is given by (23). Standard errors clustered at the origin-destination level are in parenthesis with levels of significance denoted by *** p < 0.001, and ** p < 0.01 and * p<0.05.

the CES model predicted shares.

Table 2 shows regressions of (log) observed trade shares on predictions across the three models, as well as the third-party exposure index. First, the CES model explains about 53 percent of the variation in the data (column 1).[17] After adding the index of third-party exposure in column 2, the R-squared increases to 53.7 percent, and the exposure index is significant. This provides evidence that the CES model is misspecified due to departures from IIA present in the data.[18]

Column 3 adds the difference between the predictions of the one-surge model and the CES model. After conditioning on this difference, the index of third-party exposure becomes insignificant, while the CES model continues to be significant with essentially the same coefficient. That is, the difference between the two models explains the departures from IIA associated with the index of third-party exposure. This result suggests that the additional explanatory power of the one-surge model comes from its ability to fit departures from IIA. Additionally, the R-squared increases to more than 87 percent. Finally, in column 4, we just include the predictions for trade shares of the model with $S = 1$. After conditioning on the predic-

[17]Note that, the R-squared values reported in Table 1 are for the level of trade shares rather than the log of trade shares.

[18]For further evidence, see Adao et al. (2017).

tions of this model, both the CES model prediction and the index of third-party exposure are insignificant. The one-surge model captures all the variation that the CES model can explain as well as the departures from IIA associated with the index of exposure to third-parties.

These results show that the departures from CES generated by diffusion in the one-surge model provide a larger improvement in fit than the flexibility arising from including origin-time fixed effects in the CES model.

4 Estimation Results

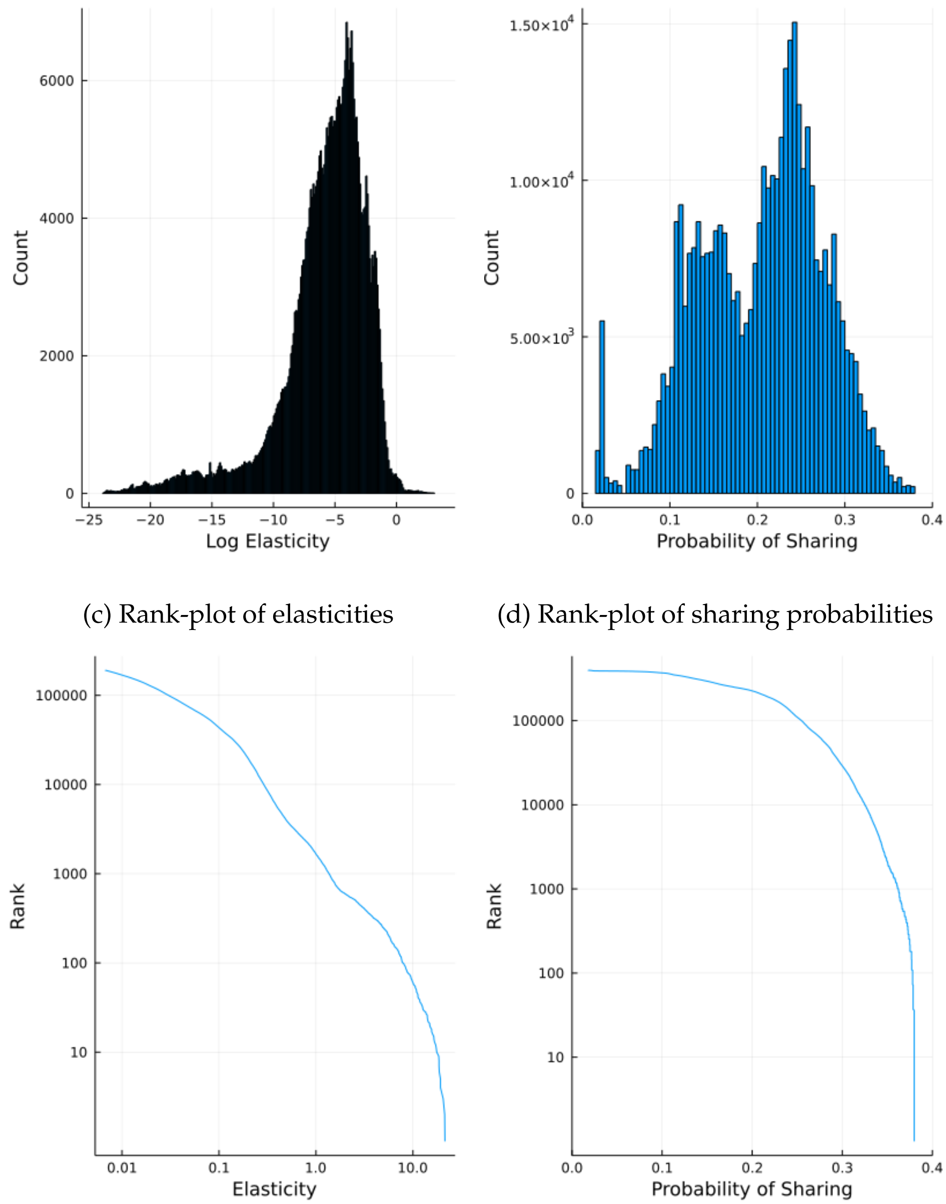
We start by exploring the one-surge model predictions for cross-price elasticities, which allow this model to get a better fit of the trade expenditure data relative to the CES model. Inspecting the cross-price elasticity in (20) makes clear that the model uses diffusion patterns to generate departures from CES. In particular, the cross-price elasticity between any two origins o and o' is proportional to the measure of ideas that those two countries share, given by the expression in (21).

First, we present some statistics related to our estimates of elasticities and shared knowledge. Figure 1 shows both histograms and rank-plots of bilateral elasticities and probabilities that two countries share knowledge, at each point in time. Here, each observation corresponds to a pair of competitors o and o' within a given destination market d at time t. The histogram of the log cross-price elasticity in 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. Similarly, Figure 1d shows that there are some countries with almost no common knowledge, but the majority of competitors have an overlap in knowledge that makes up between 10 and 30 percent of global knowledge. We can get a clearer sense of the extreme cases using the log-rank plots in Figures 1c and 1d. There are about one hundred cases with a cross-elasticity over ten, and about one thousand cases with a cross-elasticity above one. In over ten thousand cases, two countries share about 30 percent of all knowledge. These estimates show that departures from CES due to diffusion are frequent and can be very large, and that many ideas are shared between countries.

Figure 1: Distribution of elasticities and shared knowledge.

(a) Histogram of elasticities (b) Histogram of sharing probabilities

(c) Rank-plot of elasticities (d) Rank-plot of sharing probabilities



Notes: Estimates from the model with $S = 1$. Substitution elasticities are calculated using (20), while sharing probabilities are calculated using (21).

In Table 3, we explore the link between substitution elasticities and shared knowledge as estimated by the one-surge model. From (20), we expect that cross-price elasticities to be proportional to shared knowledge after conditioning on the effect of 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 positive coefficients

Table 3: Estimation results: cross-price elasticities and shared knowledge. OLS.

	$\ln \varepsilon_{oo'd}(t)$				
	(1)	(2)	(3)	(4)	(5)
$\ln \frac{K_{oo'}(t)}{\sum_{n=1}^{N} \Lambda_n(t)}$	1.342***	0.947***	1.361***	0.963***	0.957***
	(0.011)	(0.013)	(0.012)	(0.013)	(0.013)
$o \times d$ fixed effects		Yes		Yes	
t fixed effects			Yes	Yes	
$o \times d \times t$ fixed effects					Yes
Observations	396,720	396,720	396,720	396,720	396,720
R-squared	0.031	0.229	0.032	0.229	0.248
Within-R-squared		0.013	0.031	0.013	0.013

Notes: Estimates from the model with $S = 1$. $\varepsilon_{oo'd}(t)$ are calculated using (20), while sharing probabilities are calculated using the ratio of (21) to total knowledge in a year, $\sum_{n=1}^{N} \Lambda_n(t)$. Robust standard errors are in parenthesis with levels of significance denoted by *** p < 0.001, and ** p < 0.01 and * p<0.05.

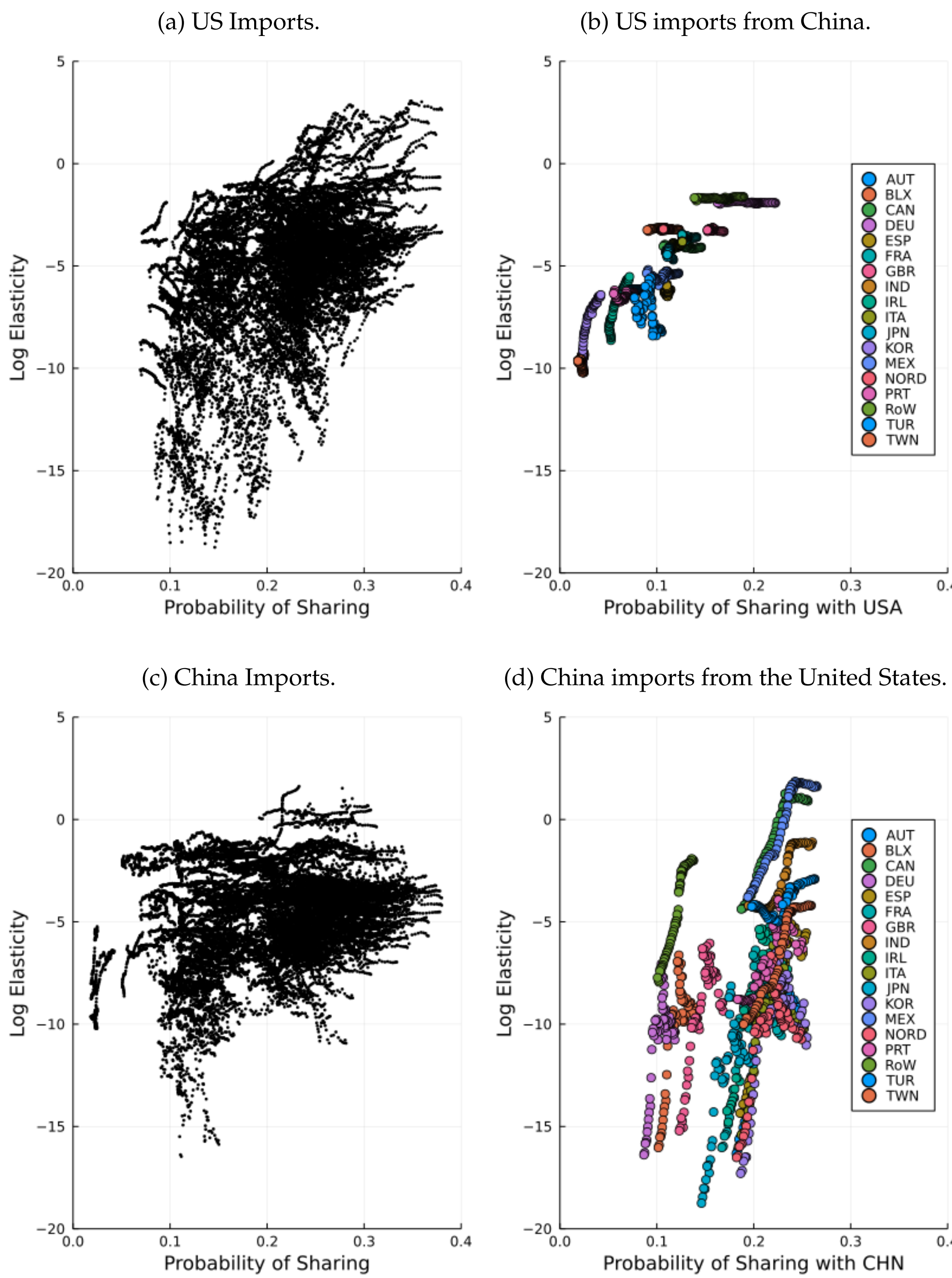
are not only are significant and around one, indicating that higher elasticities are associated with more sharing of knowledge between country pairs, but they also survive the inclusion of a battery of fixed effects.

To better visualize the estimates behind the results in Table 3, in Figure 2, we zoom into the link between cross-price elasticities and probabilities of shared knowledge when the United States (upper panels) and China (lower panels), respectively, are the destination countries. Appendix Figure B.3 shows results by destination country.[19] In the left panels, we plot the elasticities for each exporter o and competitor o' into the US market and China, while in the right panel, we focus on specific exporters — China into the United States and the United States into China–and plot the elasticities and probabilities across competitors o', for each year between 1962 and 2019. The positive coefficient reported in Table 3 clearly emerges in these figures: more elastic expenditure is linked to more shared knowledge between an exporter and each of its competitors.

Are diffusion patterns related to geography, country size, and measures of research intensity? Table 4 shows the results of regressing an estimate of average bilateral

[19]Additionally, Appendix Figures B.1 and B.2 show heat maps for the elasticities across exporters and their competitors into the United States and China, respectively, as well as heat maps for the bilateral probabilities of sharing knowledge from country n in o, for 1962 and 2019.

Figure 2: Cross-price elasticities and shared knowledge, examples.



Notes: Estimates from the model with $S = 1$. $\varepsilon_{oo'd}(t)$ are calculated using (20), while sharing probabilities are calculated using (21). Left panels show the elasticities and sharing probabilities for import into the United States (China) by each exporter and each of their competitors, for each year. Right panels show US imports from China for each competitor of China, and China imports from the United States for each competitor of the United States.

diffusion rates, $\delta_{not} = \sum_{s=0}^{S} \delta_{ons}\Lambda_{ns}(t)/\Lambda_n(t)$, on bilateral distance, GDP, population, 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 bilateral diffusion rate. Country size has an opposite effect in the origin country n and receiving country o: while diffusion rates decrease with the size of the origin country, either measure by GDP or population, they increase with the size of the receiving country. The effect of the number of researchers and the number of researchers relative to population present a similar pattern: They positively affect the diffusion rate for the receiving country and negatively the origin country. This result suggest that diffusion needs some absorption 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 country where knowledge originates slow down diffusion.

We next turn to the dynamics of knowledge. Our estimates yield predictions on the evolution of knowledge over time in each country. Figure 3 shows the stock of knowledge innovated by each of the 20 countries in our sample, $\Lambda_n(t)$. 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 total knowledge within each year, this country goes from having innovated almost a zero share of global knowledge to ten percent by 2019. Japan also experiences a surge of innovation, which begins near the start of the sample in 1962, and ends by around 1980. In contrast, the United States consistently contributes a large amount to global knowledge, and does not experience a single large surge of innovation within the sample. Due to surges of innovation in countries like Japan and China, the share of total knowledge innovated in the United States decreases from 25 to around 17 percent over the course of the sample. Still, they remain the individual country with the largest contribution.

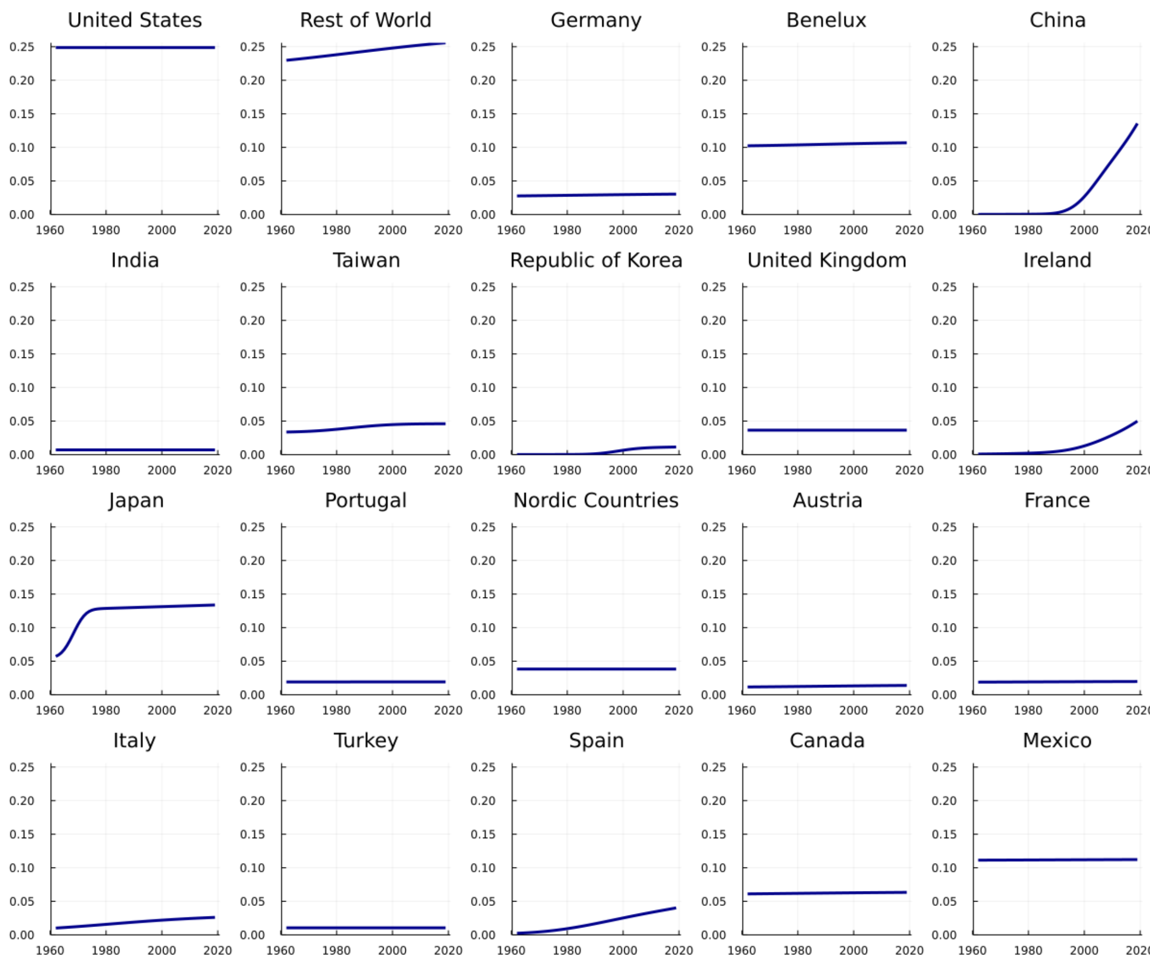
Figure 4 shows, for each country, the evolution of the share of each country's knowledge that diffused from all other countries, $1 - \frac{\Lambda_n(t)}{\Gamma(\rho)^{-1}T_n(t)}$, where $T_n(t)$ is given by (12). While in China the share of foreign knowledge decreased — as domestic sources of knowledge were surging — the opposite was true for the United States, where foreign sources of knowledge went from 40 to around 55 percent of their total knowledge. In general, our estimates deliver very high shares of foreign

Table 4: Estimation results: bilateral diffusion patterns. OLS.

	δ_{ont}				
	(1)	(2)	(3)	(4)	(5)
$\ln \text{Distance}_{on}$	-0.001**	-0.001**	-0.001**	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
$\ln \text{GDP}_n(t)$	-0.006***				
	(0.001)				
$\ln \text{GDP}_o(t)$	0.006***				
	(0.001)				
$\ln \text{Population}_n(t)$			-0.001		-0.057***
			(0.001)		(0.016)
$\ln \text{Population}_o(t)$			0.014***		-0.021
			(0.003)		(0.018)
$\ln \text{Researchers}_n(t)$				-0.016***	
				(0.003)	
$\ln \text{Researchers}_o(t)$				0.013***	
				(0.003)	
$\ln \frac{\text{Researchers}_n(t)}{\text{Population}_n(t)}$					-0.013***
					(0.003)
$\ln \frac{\text{Researchers}_o(t)}{\text{Population}_o(t)}$					0.015***
					(0.003)
n fixed effects	Yes		Yes	Yes	Yes
o fixed effects	Yes		Yes	Yes	Yes
t fixed effects	Yes		Yes	Yes	Yes
$n \times t$ fixed effects		Yes			
$o \times t$ fixed effects		Yes			
Observations	22,040	22,040	22,040	7,128	7,128
R-squared	0.292	0.303	0.291	0.383	0.384

Notes: Estimates from the model with $S = 1$. $\delta_{ont} \equiv \sum_{s=0}^{S} \delta_{ons}\Lambda_{ns}(t)/\Lambda_n(t)$ for $n \neq o$. Bilateral distance is from CEPII, GDP and Population are 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.

Figure 3: Evolution of innovated knowledge, by country.

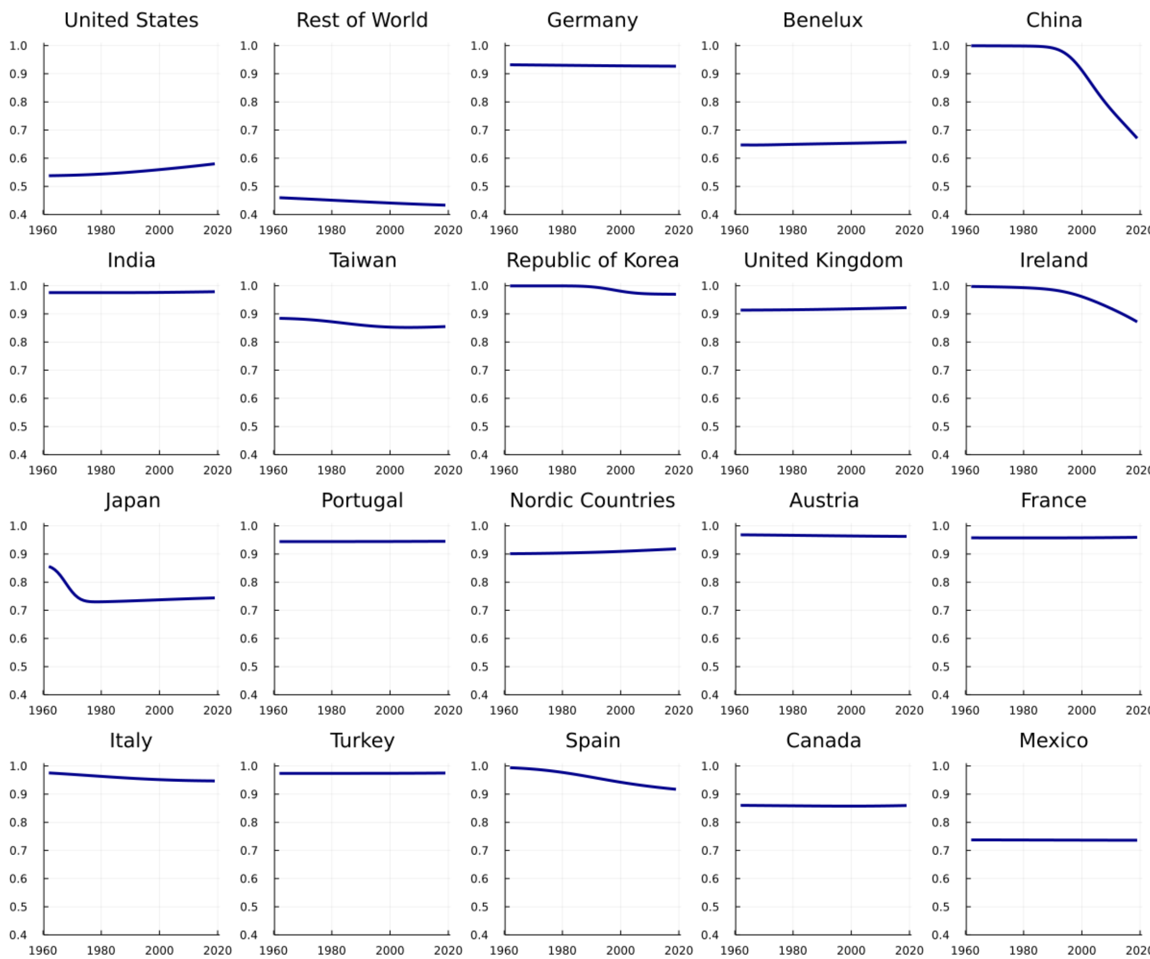


Notes: Estimates from the model with $S = 1$. Innovated knowledge is $\Lambda_n(t)$.

knowledge for most countries — including the richest countries — in the sample over the entire period. 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.

How do our knowledge estimates correlate with observable variables such as income, number of researchers and researchers per capita? Figure 5 compares the knowledge coming from innovation and diffusion with income per worker in each country over time. The top left panel shows that innovation increases are associated to increases in income per worker over time, a link predicted by standard growth theory. This is the case of China, Korea, Ireland, and Japan. Interestingly, in the top right panel, more knowledge from diffusion is also associated with higher income per worker, for rich countries such as the United States, France, and

Figure 4: Evolution of share of knowledge from diffusion, by country.



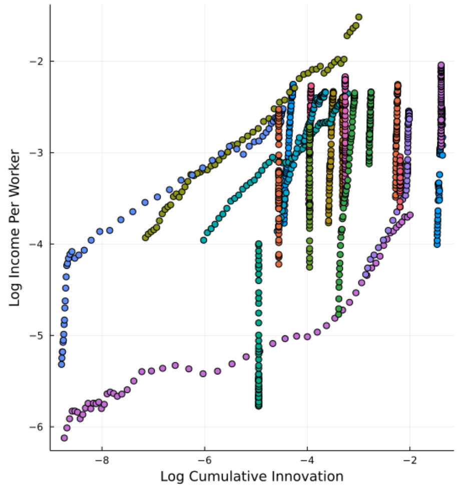
Notes: Estimates from the model with $S = 1$. Share of each country's knowledge coming from diffusion from all other countries in the sample is $1 - \frac{\Lambda_n(t)}{\Gamma(\rho)^{-1} T_n(t)}$, where $T_n(t)$ is given by (12).

Benelux, over time. The middle and bottom panels of the figure show similar patterns for the evolution of the two types of knowledge and measures of research intensity over time in each country.

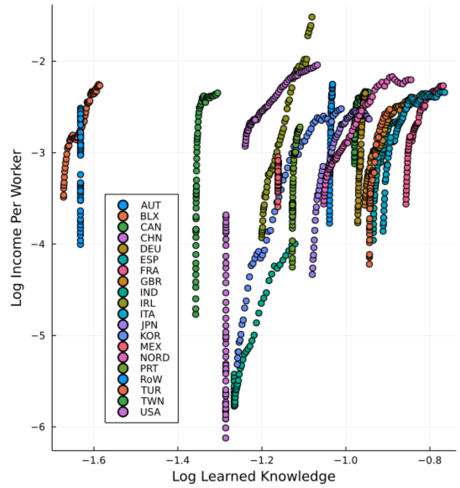
Our final result links growth in our estimates of innovated knowledge stocks to the number of researchers in a country over time. First, the more innovative the country, captured by a higher level of knowledge $\Lambda_n(t-1)$, the lower the innovation growth rate. Second, a higher number of researchers in a country is associated with a significantly higher growth of innovated knowledge, while the share of researchers in the population is significantly associated with growth only when time fixed effects are not included. Combined with the negative estimate on the innovation lag, this result is consistent with models of semi-endogenous growth where

Figure 5: Knowledge estimates vs income per worker and research, by country.

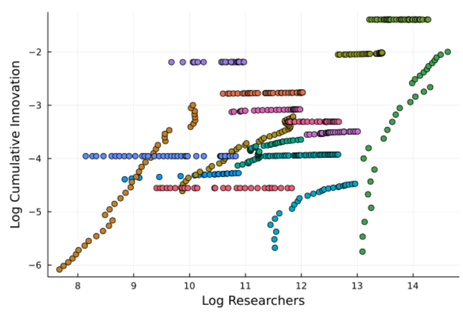
(a) Innovation vs income per worker.



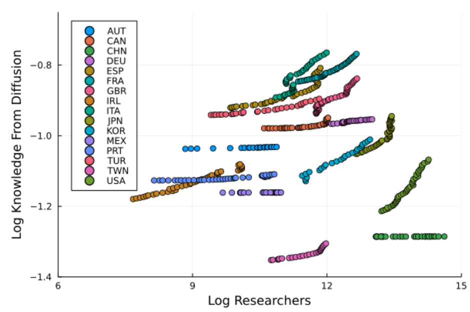
(b) Diffusion vs income per worker.



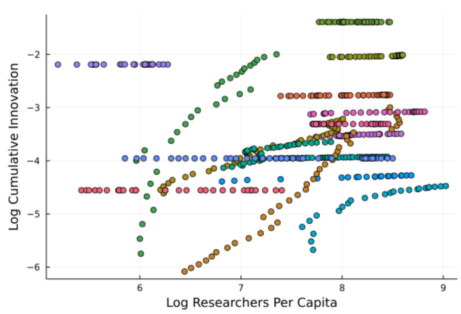
(c) Innovation vs number of researchers.



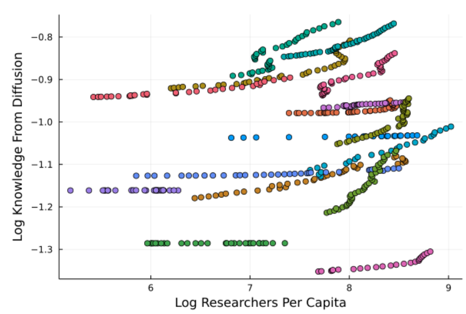
(d) Diffusion vs number of researchers.



(e) Innovation vs researchers per capita.



(f) Diffusion vs researchers per capita.



Notes: Estimates from the model with $S = 1$. Knowledge from innovation is $\Lambda_n(t)$, while knowledge from diffusion is $\Gamma(\rho)^{-1}T_n(t) - \Lambda_n(t)$, where $T_n(t)$ is given by (12). Income per worker is from PWT(10.0) and Researchers is from OECD.

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).

Table 5: Estimation results: growth patterns. OLS.

	$\Delta \ln \Lambda_n(t)$			
	(1)	(2)	(3)	(4)
$\ln \Lambda_n(t-1)$	-0.045***	-0.044***	-0.044***	-0.044***
	(0.005)	(0.005)	(0.005)	(0.005)
$\ln \text{Researchers}_n(t)$	0.003*		0.005*	
	(0.001)		(0.002)	
$\ln \frac{\text{Researchers}_n(t)}{\text{Population}_n(t)}$		0.002*		0.003
		(0.001)		(0.002)
t fixed effects			Yes	Yes
Observations	537	537	537	537
R-squared	0.886	0.886	0.898	0.897

Notes: Estimates from the model with $S = 1$. The dependent variable denotes changes in the stock of innovated knowledge by country and over time with $\Lambda_n(s) = \sum_s \Lambda_{ns}(t)$. Researches data are from OECD and population data are from PWT(10.0). All specifications include country 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.

5 Conclusion

In this paper, we propose to use easily-available data on trade flows across countries and country-level factor costs over time to uncover the global dynamics of knowledge. To such end, we build a model of innovation and diffusion based on Eaton and Kortum (1999) and extend it to incorporate international trade. Using results developed in our previous work (Lind and Ramondo, 2022), we are able to link the proposed structure of innovation and diffusion to a max-stable Fréchet distribution for productivity across countries. In the context of a Ricardian model of trade where sources compete head-to-head for markets, we obtain closed-form solutions for trade shares across countries. These expenditure shares belong to the Generalized-Extreme-Value (GEV) class, allow for rich substitution patterns across countries, and depart from Independence of Irrelevant Alternatives (IIA). 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 estimation suggests significant specialization across the globe: some countries have high innovation rates, while other countries rely on diffusion. Although innovation is highly correlated with economic growth, there are many high income countries that primarily produce using diffused ideas.

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 (2017). The rise of american ingenuity: Innovation and inventors of the golden age. *NBER Working Paper 23047*.

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.

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.

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

Brooks, W. and P. Pujolas (2017). Non linear gravity. *Mimeo, Notre-Dame University*.

Cai, J., N. Li, and A. M. Santacreu (2022). Knowledge diffusion, trade, and innovation across countries and sectors. *American Economic Journal: Macroeconomics 14*(1), 104–45.

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., L. Z. S. J. C. S. and D. Y. Xu (2021). Notching r&d investment with corporate income tax cuts in china. *American Economic Review 111*(7).

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

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., 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.

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

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

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

Lind, N. and N. Ramondo (2018). Trade with correlation. *NBER WP 24380*.

Lind, N. and N. Ramondo (2022). Global innovation and knowledge diffusion. *NBER WP 29629*.

Ma, X. (2021). College expansion, trade, and innovation: Evidence from china. *Manuscript*.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record* (673).

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

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

Silva, J. S. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and statistics 88*(4), 641–658.

Sotelo, S. (2019). Practical aspects of implementing the multinomial pml estimator. *Ann Arbor: University of Michigan, mimeo*.

A Derivations

A.1 Global Productivity Distribution

By Theorem 1 in Lind and Ramondo (2022) the joint distribution of productivity across countries at each moment in time is max-stable multivariate Fréchet. Since the proof is constructive, we can calculate the distribution directly as follows.

First, we can express the joint distribution of productivity as

$$\begin{aligned} F(z_1,\ldots,z_N;t) &\equiv \mathbb{P}\left[Z_{1t}(v)\le z_1,\ldots,Z_{Nt}(v)\le z_N\right] \\ &= \mathbb{P}\left[Q_i(v)A_{io}(v)\le z_o\ \ \forall o\in\mathcal{N}_i(t,v) \text{ for all } i \text{ s.t. } t_i^*(v)\le t\right] \\ &= \mathbb{P}\left[Q_i(v)\le \min_{o\in\mathcal{N}_i(t,v)}\frac{z_o}{A_{io}(v)} \text{ for all } i \text{ s.t. } t_i^*(v)\le t\right] \\ &= \mathbb{P}\left[Q_i(v)> \min_{o\in\mathcal{N}_i(t,v)}\frac{z_o}{A_{io}(v)} \text{ for no } i \text{ s.t. } t_i^*(v)\le t\right], \end{aligned}$$

which yields the result in (9). Next, since applicability is continuously distributed and $n_i(v)\in\mathcal{N}_i(t,v)$, there is a unique origin, $o_i^*(v,t)$, that attains the minimum in this expression for each idea known at time t. We then have

$$\begin{aligned} F(z_1,\ldots,z_N;t) &= \mathbb{P}\left[Q_i(v)>\frac{z_{o_i^*(v)}}{A_{io_i^*(v)}(v)} \text{ for no } i \text{ s.t. } t_i^*(v)\le t\right] \\ &= \exp\left[-\mathbb{E}\sum_{i=1}^{\infty}\mathbf{1}\left\{Q_i(v)>\frac{z_{o_i^*(v)}}{A_{io_i^*(v)}(v)}, t_i^*(v)\le t\right\}\right] \\ &= \exp\left[-\mathbb{E}\sum_{i=1}^{\infty}\sum_{o=1}^{N}\mathbf{1}\left\{Q_i(v)>\frac{z_o}{A_{io}(v)}, o_i^*(v)=o, t_i^*(v)\le t\right\}\right]. \end{aligned}$$

Next, the expression in (13) follows from

$$
\begin{aligned}
&\mathbb{E}\sum_{i=1}^{\infty}\sum_{o=1}^{N}\mathbf{1}\left\{Q_i(v) > \frac{z_o}{A_{io}(v)}, o_i^*(v) = o, t_i^*(v) \le t\right\} \\
&= \mathbb{E}\sum_{i=1}^{\infty}\sum_{o=1}^{N}\mathbf{1}\left\{Q_i(v) > \frac{z_o}{A_{io}(v)}, o \in \mathcal{N}_i(t,v), \frac{z_o}{A_{io}(v)} \le \frac{z_{o'}}{A_{io'}(v)} \;\; \forall o' \in \mathcal{N}_i(t,v), t_i^*(v) \le t\right\} \\
&= \mathbb{E}\sum_{i=1}^{\infty}\sum_{o=1}^{N}\mathbf{1}\left\{i \in \mathcal{I}_o(z_i,t,v), \frac{z_o}{A_{io}(v)} \le \frac{z_{o'}}{A_{io'}(v)} \;\; \forall o' \in \mathcal{N}_i(t,v)\right\} \\
&= \mathbb{E}\sum_{o=1}^{N}\sum_{i\in\mathcal{I}_o(z_o,t,v)}\mathbf{1}\left\{A_{io'}(v) \le \frac{z_{o'}}{z_o}A_{io}(v) \;\; \forall o' \in \mathcal{N}_i(t,v)\right\},
\end{aligned}
$$

for $\mathcal{I}_o(z_o,t,v) \equiv \{i = 1,2,\ldots \mid Q_i(v)A_{io}(v) > z_o, o \in \mathcal{N}_i(t,v), t_i^*(v) \le t\}$ as defined in (11). Note that

$$
\begin{aligned}
&\mathbb{E}\sum_{i=1}^{\infty}\mathbf{1}\{i \in \mathcal{I}_o(z_o,t,v), n_i(v) = n, s_i(v) = s, t_i^*(v) \le x\} \\
&= \mathbb{E}\sum_{i=1}^{\infty}\mathbf{1}\left\{Q_i(v)A_{io}(v) > z_o, o \in \mathcal{N}_i(t,v), n_i(v) = n, s_i(v) = s, t_i^*(v) \le \min\{t,x\}\right\} \\
&= \mathbb{E}\sum_{i=1}^{\infty}\mathbf{1}\left\{Q_i(v) > \frac{z_o}{A_{io}(v)}, o \in \mathcal{N}_i(t,v), n_i(v) = n, s_i(v) = s, t_i^*(v) \le t, t_i^*(v) \le x\right\} \\
&= \int_{-\infty}^{\min\{t,x\}}\int_0^{\infty}\int_{z_o/a_o}^{\infty}\theta q^{-\theta-1}\mathrm{d}q\mathrm{d}e^{-a_o^{-\sigma}}p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^* \\
&= \int_{-\infty}^{\min\{t,x\}}\int_0^{\infty}\left(\frac{z_o}{a_o}\right)^{-\theta}\mathrm{d}e^{-a_o^{-\sigma}}p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^* \\
&= \Gamma(1-\frac{\theta}{\sigma})\int_{-\infty}^{\min\{t,x\}}p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^* z_o^{-\theta},
\end{aligned}
$$

and hence,

$$
\begin{aligned}
&\mathbb{E}\sum_{i=1}^{\infty}\mathbf{1}\{i \in \mathcal{I}_o(z_o,t,v)\} \\
&= \mathbb{E}\sum_{i=1}^{\infty}\sum_{s=0}^{S}\sum_{n=1}^{N}\mathbf{1}\{i \in \mathcal{I}_o(z_o,t,v), n_i(v) = n, s_i(v) = s, t_i^*(v) \le t\} \\
&= \Gamma(1-\frac{\theta}{\sigma})\sum_{s=0}^{S}\sum_{n=1}^{N}\int_{-\infty}^{t}p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^* z_o^{-\theta} = T_o(t)z_o^{-\theta},
\end{aligned}
$$

for $T_o(t)$ defined in (12).

Together, we then have

$$\begin{aligned}
&\mathbb{P}[n_i(v) = n, s_i(v) = s, t_i^*(v) \leq x \mid i \in \mathcal{I}_o(z_o, t, v)] \\
&= \frac{\int_{-\infty}^{\min\{t,x\}} p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^* z_o^{-\theta}}{\sum_{s=0}^{S}\sum_{n=1}^{N}\int_{-\infty}^{t} p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^* z_o^{-\theta}} \\
&= \frac{\int_{-\infty}^{\min\{t,x\}} p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*}{\sum_{s=0}^{S}\sum_{n=1}^{N}\int_{-\infty}^{t} p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*},
\end{aligned}$$

which means that the probability that an idea known to o at time t is from surge s of innovator n is

$$\mathbb{P}[n_i(v) = n, s_i(v) = s \mid i \in \mathcal{I}_o(z_o, t, v)] = \frac{\int_{-\infty}^{t} p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*}{\sum_{s=0}^{S}\sum_{n=1}^{N}\int_{-\infty}^{t} p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*},$$

while the density of discovery times among those ideas is

$$\frac{\partial}{\partial t^*}\mathbb{P}[t_i^*(v) \leq t^* \mid i \in \mathcal{I}_o(z_o, t, v), n_i(v) = n, s_i(v) = s] = \frac{p_{ons}(t-t^*)\Lambda'_{ns}(t^*)}{\int_{-\infty}^{t} p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*}.$$

Below, we use these two results to take expectations with respect to $n_i(v)$, $s_i(v)$, and $t_i^*(v)$ conditional on $i \in \mathcal{I}_o(z_o, t, v)$.

Conditional on $n_i(v)$, $s_i(v)$, and $t_i^*(v)$, $\mathbf{1}\{o \in \mathcal{N}_i(t, v)\}$ is a Bernoulli random variable with success probability $p_{ons}(t-t^*)$ independent across o. That is, for $(x_1, \ldots, x_N) \in \{0, 1\}^N$,

$$\mathbb{P}_{nst^*}\left[\mathbf{1}\{o \in \mathcal{N}_i(t, v)\} = x_o \ \ \forall o = 1, \ldots, N\right] = \prod_{o=1}^{N}(1 - p_{ons}(t-t^*))^{1-x_o} p_{ons}(t-t^*)^{x_o},$$

where $\mathbb{P}_{nst^*}[\cdot] \equiv \mathbb{P}[\cdot \mid n_i(v) = n, s_i(v) = s, t_i^*(v) = t^*]$. Then,

$$\begin{aligned}
&\mathbb{P}\left[A_{io'}(v) \leq \frac{z_{o'}}{z_o}A_{io}(v) \ \ \forall o' \in \mathcal{N}_i(t, v) \mid A_{io}(v) = a_o, i \in \mathcal{I}_o(z_i, t, v), n_i(v) = n, s_i(v) = s, t_i^*(v) = t^*\right] \\
&= \mathbb{E}_{nst^*}\left[\prod_{o' \neq o}\mathbf{1}\left\{A_{io'}(v) \leq \frac{z_{o'}}{z_o}A_{io}(v)\right\}^{\mathbf{1}\{o' \in \mathcal{N}_i(t,v)\}} \mid A_{io}(v) = a_o, i \in \mathcal{I}_o(z_i, t, v)\right] \\
&= \prod_{o' \neq o}\mathbb{E}_{nst^*}\left[\mathbf{1}\left\{A_{io'}(v) \leq \frac{z_{o'}}{z_o}a_o\right\}^{\mathbf{1}\{o' \in \mathcal{N}_i(t,v)\}}\right] \\
&= \prod_{o' \neq o}\left[1 - p_{o'ns}(t-t^*) + p_{o'ns}(t-t^*)e^{-\left(\frac{z_{o'}}{z_o}a_o\right)^{-\sigma}}\right],
\end{aligned}$$

which uses independence of applicability and conditional independence of $\mathbf{1}\{o \in \mathcal{N}_i(t,v)\}$ across countries.

We now average this result over applicability levels. The distribution of applicability in o among ideas in $\mathcal{I}_o(z_i,t,v)$ conditional on $n_i(v)$, $s_i(v)$, and $t_i^*(v)$ is

$$\mathbb{P}_{nst^*}\left[A_{io}(v) \leq x \mid i \in \mathcal{I}_o(z_o,t)\right] = \frac{\int_0^x \int_{z_o/a_o} \theta q^{-\theta-1} \mathrm{d}q \mathrm{d}e^{-a_o^{-\sigma}} p_{ons}(t-t^*)\Lambda'_{ns}(t^*)}{\int_0^\infty \int_{z_o/a_o} \theta q^{-\theta-1} \mathrm{d}q \mathrm{d}e^{-a_o^{-\sigma}} p_{ons}(t-t^*)\Lambda'_{ns}(t^*)}$$
$$= \frac{\int_0^x a_o^\theta e^{-a_o^{-\sigma}} \sigma a_o^{-\sigma-1} \mathrm{d}x z_o^{-\theta}}{\int_0^\infty a_o^\theta e^{-a_o^{-\sigma}} \sigma a_o^{-\sigma-1} \mathrm{d}a_o z_o^{-\theta}} = \int_0^x \frac{e^{-a_o^{-\sigma}} \sigma a_o^{\theta-\sigma-1}}{\Gamma(1-\theta/\sigma)} \mathrm{d}a_o.$$

Therefore,

$$H_{ons}(z_1,\ldots,z_N;t-t^*) \equiv \mathbb{P}_{nst^*}\left[A_{io'}(v) \leq \frac{z_{o'}}{z_o} A_{io}(v) \;\; \forall o' \in \mathcal{N}_i(t,v) \mid i \in \mathcal{I}_o(z_i,t,v)\right]$$
$$= \int_0^\infty \prod_{o' \neq o} \left[1 - p_{o'ns}(t-t^*) + p_{o'ns}(t-t^*) e^{-\left(\frac{z_{o'}}{z_o} a_o\right)^{-\sigma}}\right] \frac{e^{-a_o^{-\sigma}} \sigma a_o^{\theta-\sigma-1}}{\Gamma(1-\theta/\sigma)} \mathrm{d}a_o,$$

as in (14). Using the previous results for the distribution of ideas across surges, innovators, and discovery times within $\mathcal{I}_o(z_o,t,v)$ yields

$$H_o(z_1,\ldots,z_N;t) \equiv \mathbb{P}\left[A_{io'}(v) \leq \frac{z_{o'}}{z_o} A_{io}(v) \;\; \forall o' \in \mathcal{N}_i(t,v) \mid i \in \mathcal{I}_o(z_i,t,v)\right]$$
$$= \mathbb{E}\left[\mathbb{P}_{nst^*}\left[A_{io'}(v) \leq \frac{z_{o'}}{z_o} A_{io}(v) \;\; \forall o' \in \mathcal{N}_i(t,v) \mid i \in \mathcal{I}_o(z_i,t,v)\right] \mid i \in \mathcal{I}_o(z_i,t,v)\right]$$
$$= \mathbb{E}\left[H_{on_i(v)s_i(v)}(z_1,\ldots,z_N;t-t_i^*(v)) \mid i \in \mathcal{I}_o(z_i,t,v)\right]$$
$$= \frac{\sum_{s=0}^S \sum_{n=1}^N \int_{-\infty}^t H_{ons}(z_1,\ldots,z_N;t-t^*) p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*}{\sum_{s=0}^S \sum_{n=1}^N \int_{-\infty}^t, p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*} \mathrm{d}t^*,$$

which results in (16).

Finally, this result and the previous result on the expected number of ideas in

$\mathcal{I}_o(z_o, t, v)$ allow us to compute

$$\begin{aligned}
&\mathbb{E} \sum_{o=1}^{N} \sum_{i \in \mathcal{I}_o(z_o,t,v)} \mathbf{1}\left\{ A_{io'}(v) \leq \frac{z_{o'}}{z_o} A_{io}(v) \ \ \forall o' \in \mathcal{N}_i(t,v) \right\} \\
&= \sum_{o=1}^{N} \mathbb{P}\left[A_{io'}(v) \leq \frac{z_{o'}}{z_o} A_{io}(v) \ \ \forall o' \in \mathcal{N}_i(t,v) \mid i \in \mathcal{I}_o(z_o,t,v) \right] \mathbb{E} \sum_{i=1}^{\infty} \mathbf{1}\{ i \in \mathcal{I}_o(z_o,t,v) \} \\
&= \sum_{o=1}^{N} H_o(z_1, \ldots, z_N; t) T_o(t) z_o^{-\theta}.
\end{aligned}$$

Therefore, the joint distribution of productivity is

$$F(z_1, \ldots, z_N; t) \equiv \mathbb{P}[Z_1(t,v) \leq z_1, \ldots, Z_n(t,v) \leq z_N] = \exp\left[-\sum_{o=1}^{N} H_o(z_1, \ldots, z_N; t) T_o(t) z_o^{-\theta} \right],$$

as in (15).

Note that as $z_{o'} \to \infty$ for all $o' \neq o$, we have $H_{ons}(z_1, \ldots, z_N; t - t^*) \to 1$, $H_o(z_1, \ldots, z_N; t) \to 1$, and so

$$\mathbb{P}[Z_o(t,v) \leq z_o] = \lim_{z_{o'} \to \infty \ \forall o' \neq o} F(z_1, \ldots, z_N; t) = e^{-T_o(t) z_o^{-\theta}}.$$

That is, the marginal distribution in o is Fréchet with shape θ and scale $T_o(t)$.

A.2 Trade Shares and Elasticities

To get trade shares and elasticities, we apply the results in Proposition 2 of Lind and Ramondo (2018). To do so, first note that the (time-dependent) correlation function implied by (15) is

$$G^d(x_1, \ldots, x_N; t) = \sum_{o=1}^{N} H_o\left((x_1/T_1(t))^{-1/\theta}, \ldots, (x_N/T_N(t))^{-1/\theta}; t \right) x_o.$$

Then,

$$
\begin{aligned}
G_o^d(x_1,\ldots,x_N;t) \equiv{}& \frac{\partial G^d(x_1,\ldots,x_N)}{\partial x_o} \\
={}& H_o\left((x_1/T_1(t))^{-1/\theta},\ldots,(x_N/T_N(t))^{-1/\theta};t\right) \\
&\quad + x_o\frac{\partial}{\partial x_o}\left[H_o\left((x_1/T_1(t))^{-1/\theta},\ldots,(x_N/T_N(t))^{-1/\theta};t\right)\right] \\
&\quad + \sum_{o'\neq o} x_{o'}\frac{\partial}{\partial x_o}\left[H_{o'}\left((x_1/T_1(t))^{-1/\theta},\ldots,(x_N/T_N(t))^{-1/\theta};t\right)\right] \\
={}& H_o\left((x_1/T_1(t))^{-1/\theta},\ldots,(x_N/T_N(t))^{-1/\theta};t\right) \\
&\quad - \frac{1}{\theta}\sum_{o'=1}^{N} x_{o'}H_{o'o}\left((x_1/T_1(t))^{-1/\theta},\ldots,(x_N/T_N(t))^{-1/\theta};t\right)x_o^{-1/\theta-1}T_o(t)^{1/\theta},
\end{aligned}
$$

where $H_{o'o}\equiv \partial H_{o'}/\partial z_o$. Then,

$$
\begin{aligned}
&T_o(t)z_o^{-\theta}G_o^d(T_1(t)z_1^{-\theta},\ldots,T_N(t)z_N^{-\theta};t) \\
&= T_o(t)z_o^{-\theta}H_o\left(z_1,\ldots,z_N;t\right) - \frac{1}{\theta}\sum_{o'=1}^{N} T_{o'}(t)z_{o'}^{-\theta}H_{o'o}\left(z_1,\ldots,z_N;t\right)z_o.
\end{aligned}
$$

Define

$$
\begin{aligned}
M(z_1,\ldots,z_N;t) \equiv \sum_{s=0}^{S}\sum_{n=1}^{N}\int_{-\infty}^{t}\int_0^{\infty}\prod_{o=1}^{N}\left[1-p_{ons}(t-t^*)+p_{ons}(t-t^*)e^{-\left(\frac{z_o}{x}\right)^{-\sigma}}\right] \\
x^{-\theta-1}\mathrm{d}x\Lambda'_{ns}(t^*)\mathrm{d}t^*.
\end{aligned}
$$

Then, by a change of variables of $a_o = z_o/x$ in (14),

$$
\begin{aligned}
H_{ons}(z_1,\ldots,z_N;t-t^*) = \int_0^{\infty}\prod_{o'\neq o}\left[1-p_{o'ns}(t-t^*)+p_{o'ns}(t-t^*)e^{-(z_{o'}/x)^{-\sigma}}\right] \\
\times\frac{e^{-(z_o/x)^{-\sigma}}\sigma z_o^{\theta-\sigma}x^{\sigma-\theta-1}}{\Gamma(1-\theta/\sigma)}\mathrm{d}x,
\end{aligned}
$$

we have

$$\frac{\partial M(z_1,\ldots,z_N;t)}{\partial z_o} = \sum_{s=0}^{S}\sum_{n=1}^{N}\int_{-\infty}^{t}\int_{0}^{\infty} p_{ons}(t-t^*)e^{-\left(\frac{z_o}{x}\right)^{-\sigma}}\sigma z_o^{-\sigma-1}x^{\sigma}$$
$$\times\prod_{o'\neq o}\left[1-p_{o'ns}(t-t^*)+p_{o'ns}(t-t^*)e^{-\left(\frac{z_{o'}}{x}\right)^{-\sigma}}\right]x^{-\theta-1}\mathrm{d}x\Lambda'_{ns}(t^*)\mathrm{d}t^*$$
$$=\sum_{s=0}^{S}\sum_{n=1}^{N}\int_{-\infty}^{t}\Gamma(1-\theta/\sigma)p_{ons}(t-t^*)z_o^{-\theta-1}H_{ons}(z_1,\ldots,z_N;t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*$$
$$=\Gamma(1-\theta/\sigma)z_o^{-\theta-1}H_o(z_1,\ldots,z_N;t)\sum_{s=0}^{S}\sum_{n=1}^{N}\int_{0}^{\infty}p_{ons}(t-t^*)\Lambda'_{ns}(t^*)\mathrm{d}t^*$$
$$=T_o(t)z_o^{-\theta-1}H_o(z_1,\ldots,z_N;t).$$

Consequently,

$$\frac{\partial^2 M(z_1,\ldots,z_N;t)}{\partial z_{o'}\partial z_o} = \frac{\partial}{\partial z_o}T_{o'}(t)z_{o'}^{-\theta-1}H_{o'}(z_1,\ldots,z_N;t)$$
$$= T_{o'}(t)z_{o'}^{-\theta-1}H_{o'o}(z_1,\ldots,z_N;t).$$

Then, replacing yields

$$T_o(t)z_o^{-\theta}G_o^d(T_1(t)z_1^{-\theta},\ldots,T_N(t)z_N^{-\theta})$$
$$= T_o(t)z_o^{-\theta}H_o\left(z_1,\ldots,z_N;t\right)-\frac{1}{\theta}\sum_{o'=1}^{N}T_{o'}(t)z_{o'}^{-\theta}H_{o'o}\left(z_1,\ldots,z_N;t\right)z_o$$
$$= T_o(t)z_o^{-\theta}H_o\left(z_1,\ldots,z_N;t\right)-\frac{1}{\theta}\sum_{o'=1}^{N}T_o(t)z_o^{-\theta}H_{oo'}\left(z_1,\ldots,z_N;t\right)z_{o'}.$$

Note that $H_o(z_1,\ldots,z_N;t)$ is homogenous of degree zero because $H_{nos}(z_1,\ldots,z_N;t-t^*)$ is homogenous degree zero. Then, by Euler's theorem for homogenous functions, $\sum_{o'=1}^{N} z_{o'}H_{oo'}(z_1,\ldots,z_N;t)=0$, implying that

$$T_o(t)z_o^{-\theta}G_o^d(T_1(t)z_1^{-\theta},\ldots,T_N(t)z_N^{-\theta}) = T_o(t)z_o^{-\theta}H_o\left(z_1,\ldots,z_N;t\right).$$

Therefore,

$$G_o^d(x_1,\ldots,x_N;t) = H_o\left((x_1/T_1(t))^{-1/\theta},\ldots,(x_N/T_N(t))^{-1/\theta};t\right)$$
$$= T_o^{1/\theta}x_o^{-1/\theta-1}M_o\left((x_1/T_1(t))^{-1/\theta},\ldots,(x_N/T_N(t))^{-1/\theta};t\right),$$

since $M_o(z_1,\ldots,z_N;t) = T_o(t)z_o^{-\theta-1}H_o(z_1,\ldots,z_N;t)$ where $M_o \equiv \frac{\partial M}{\partial z_o}$. Then,

$$\begin{aligned} G^d(x_1,\ldots,x_N;t) &= \sum_{o=1}^{N} x_o G_o^d(x_1,\ldots,x_N;t) \\ &= \sum_{o=1}^{N} (x_o/T_o(t))^{-1/\theta} M_o\left((x_1/T_1(t))^{-1/\theta},\ldots,(x_N/T_N(t))^{-1/\theta};t\right), \end{aligned}$$

and

$$\begin{aligned} P_d(t) &= G^d(T_1(t)P_{1d}(t)^{-\theta},\ldots,T_N(t)P_{Nd}(t)^{-\theta})^{-\frac{1}{\theta}} \\ &= \left[\sum_{o=1}^{N} H_o\left(P_{1d}(t),\ldots,P_{Nd}(t);t\right)T_o(t)P_{od}(t)^{-\theta}\right]^{-\frac{1}{\theta}}. \end{aligned}$$

Note that, here, $\zeta = \Gamma(1-(\eta-1)/\theta)^{\frac{1}{\eta-1}}$, which is the inverse of γ in Proposition 2 of Lind and Ramondo (2018), and $T_o(t)P_{od}(t)^{-\theta}$ corresponds to $P_{od}^{-\theta}$ in that same proposition.

Next, trade shares are

$$\begin{aligned} \pi_{od}(t) &= \frac{T_{od}(t)P_{od}(t)^{-\theta}G_o^d(T_{1d}(t)P_{1d}(t)^{-\theta},\ldots,T_{Nd}(t)P_{Nd}(t)^{-\theta};t)}{G^d(T_{1d}(t)P_{1d}(t)^{-\theta},\ldots,T_{Nd}(t)P_{Nd}(t)^{-\theta};t)} \\ &= \frac{T_{od}(t)P_{od}(t)^{-\theta}H_o(P_{1d}(t),\ldots,P_{Nd};t)}{\sum_{o'=1}^{N} T_{o'd}(t)P_{o'd}(t)^{-\theta}H_{o'}(P_{1d}(t),\ldots,P_{Nd};t)}, \end{aligned}$$

yielding (18).

We next use this result to derive cross-price elasticities of substitution. Since H_o is homogenous of degree zero,

$$\begin{aligned} \pi_{od}(t) &= \frac{T_o(t)H_o(P_{1d}(t),\ldots,P_{Nd}(t);t)P_{od}(t)^{-\theta}}{\sum_{o'=1}^{N} T_{o'}(t)H_{o'}(P_{1d}(t),\ldots,P_{Nd}(t);t)P_{o'd}(t)^{-\theta}} \\ &= T_o(t)\left(\frac{P_{od}(t)}{P_d(t)}\right)^{-\theta} H_o\left(\frac{P_{1d}(t)}{P_d(t)},\ldots,\frac{P_{Nd}(t)}{P_d(t)};t\right). \end{aligned}$$

Then, for $o' \neq o$,

$$\frac{\partial \pi_{odt}}{\partial \ln \frac{P_{o'd}(t)}{P_d(t)}} = T_o(t)\left(\frac{P_{od}(t)}{P_d(t)}\right)^{-\theta} \frac{\partial}{\partial \ln \frac{P_{o'd}(t)}{P_d(t)}} H_o\left(\frac{P_{1d}(t)}{P_d(t)},\ldots,\frac{P_{Nd}(t)}{P_d(t)};t\right)$$

Note that

$$
\begin{aligned}
&T_o(t)z_o^{-\theta}\frac{\partial}{\partial \ln z_{o'}}H_o(z_1,\dots,z_N;t)\\
&=T_o(t)z_o^{-\theta}\sum_{s=0}^{S}\sum_{n=1}^{N}\int_0^\infty\int_0^\infty p_{o'ns}(a)e^{-\left(\frac{z_{o'}}{z_o}\right)^{-\sigma}x}\sigma\left(\frac{z_{o'}}{z_o}\right)^{-\sigma}x\\
&\qquad\times\left[\prod_{m\notin\{o,o'\}}\left(1-p_{mns}(a)+p_{mns}(a)e^{-\left(\frac{z_m}{z_o}\right)^{-\sigma}x}\right)\right]\\
&\qquad\times\frac{e^{-x}x^{\rho-1}}{\Gamma(\rho)}\mathrm{d}x\frac{p_{ons}(a)\Lambda'_{ns}(t-a)}{\sum_{s=0}^{S}\sum_{n=1}^{N}\int_0^\infty p_{ons}(a)\Lambda'_{ns}(t-a)\mathrm{d}a}\mathrm{d}a\\
&=\sigma z_o^{-\theta}\sum_{s=0}^{S}\sum_{n=1}^{N}\int_0^\infty\int_0^\infty p_{o'ns}(a)\left(\frac{z_{o'}}{z_o}\right)^{-\sigma}x^\rho\left[\prod_{m\notin\{o,o'\}}\left(1-p_{mns}(a)+p_{mns}(a)e^{-\left(\frac{z_m}{z_o}\right)^{-\sigma}x}\right)\right]\\
&\qquad\times e^{-\frac{z_o^{-\sigma}+z_{o'}^{-\sigma}}{z_o^{-\sigma}}x}\mathrm{d}x p_{ons}(a)\Lambda'_{ns}(t-a)\mathrm{d}a\\
&=\sigma\rho\frac{z_{o'}^{-\sigma}z_o^{-\sigma}}{(z_o^{-\sigma}+z_{o'}^{-\sigma})^2}\left(z_o^{-\sigma}+z_{o'}^{-\sigma}\right)^{1-\rho}\sum_{s=0}^{S}\sum_{n=1}^{N}\int_0^\infty\int_0^\infty\left[\prod_{m\notin\{o,o'\}}\left(1-p_{mns}(a)+p_{mns}(a)e^{-\frac{z_m^{-\sigma}}{z_o^{-\sigma}+z_{o'}^{-\sigma}}\tilde{x}}\right)\right]\frac{e^{-\tilde{x}}\tilde{x}^\rho}{\Gamma(1+\rho)}\mathrm{d}\tilde{x}\\
&\qquad\times p_{ons}(a)p_{o'ns}(a)\Gamma(\rho)\Lambda'_{ns}(t-a)\mathrm{d}a,
\end{aligned}
$$

with $\rho\equiv 1-\theta/\sigma$. Evaluating at $z_o=P_{od}(t)$, we get (20).

A.3 Innovation Stocks, Scales, and Sharing Probabilities

Since the innovation rate in country n at time t is

$$\lambda_n(t)=\Lambda'_n(t)=\gamma_n\Lambda_n(t)+\sum_{s=1}^{S}\lambda_{ns}\frac{1}{\nu_{ns}}\phi\left(\frac{t-\mu_{ns}}{\nu_{ns}}\right),$$

we have

$$\frac{\partial}{\partial t}\left[e^{-\gamma_n t}\Lambda_n(t)\right]=e^{-\gamma_n t}\Lambda'_n(t)-\gamma_n e^{-\gamma_n t}\Lambda_n(t)=\sum_{s=1}^{S}\lambda_{ns}e^{-\gamma_n t}\frac{1}{\nu_{ns}}\phi\left(\frac{t-\mu_{ns}}{\nu_{ns}}\right).$$

Integrating from 0 to t yields

$$e^{-\gamma_n t}\Lambda_n(t)-\Lambda_n(0)=\sum_{s=1}^{S}\lambda_{ns}\int_0^t e^{-\gamma_n t^*}\frac{1}{\nu_{ns}}\phi\left(\frac{t^*-\mu_{ns}}{\nu_{ns}}\right)\mathrm{d}t^*,$$

and hence,

$$\Lambda_n(t)=\Lambda_{n0}(t)+\sum_{s=1}^{S}\Lambda_{ns}(t), \tag{A.1}$$

where

$$\Lambda_{n0}(t) \equiv \lambda_{n0} e^{\gamma_n t} \quad \text{for} \quad \lambda_{n0} \equiv \Lambda_n(0) - \lambda_{ns} \int_{-\infty}^{0} e^{-\gamma_n t^*} \frac{1}{\nu_{ns}} \phi\left(\frac{t^* - \mu_{ns}}{\nu_{ns}}\right) \mathrm{d}t^*,$$

and

$$\Lambda_{ns}(t) \equiv \lambda_{ns} \int_{-\infty}^{t} e^{\gamma_n (t-t^*)} \frac{1}{\nu_{ns}} \phi\left(\frac{t^* - \mu_{ns}}{\nu_{ns}}\right) \mathrm{d}t^*.$$

We use the following lemma to solve for $\Lambda_{ns}(t)$ ($s \geq 1$) and for other results in the section.

Lemma A.1.

$$\int_{-\infty}^{t} e^{a+bx} \frac{1}{s} \phi\left(\frac{x-m}{s}\right) dx = e^{a+bm+b^2 s^2/2} \Phi\left(\frac{t-m-bs^2}{s}\right). \tag{A.2}$$

Proof.

$$\int_{-\infty}^{t} e^{a+bx} \frac{1}{s} \phi\left(\frac{x-m}{s}\right) \mathrm{d}x = \int_{-\infty}^{\frac{t-m}{s}} e^{a+b(m+sz)} \phi\left(z\right) \mathrm{d}x$$
$$= e^{a+bm+b^2 s^2/2} \Phi\left(\frac{t-m-bs^2}{s}\right).$$

□

Applying Lemma A.1 to (A.1) yields

$$\Lambda_{ns}(t) = \lambda_{ns} \int_{-\infty}^{t} e^{\gamma_n (t-t^*)} \frac{1}{\nu_{ns}} \phi\left(\frac{t^* - \mu_{ns}}{\nu_{ns}}\right) \mathrm{d}t^*$$
$$= \lambda_{ns} e^{\gamma_n (t-\mu_{ns}+\gamma_n \nu_{ns}^2/2)} \Phi\left(\frac{t - \mu_{ns} + \gamma_n \nu_{ns}^2}{\nu_{ns}}\right).$$

Next, we use the following lemma to calculate scale parameters and the probability that any two countries share an idea.

Lemma A.2. *Let* $\Lambda(t) = \lambda e^{\gamma(t-\mu+\gamma\nu^2/2)} \Phi\left(\frac{t-\mu+\gamma\nu^2}{\nu}\right)$*. Then,*

$$\int_{-\infty}^{t} (1 - e^{-\delta(t-t^*)}) \Lambda'(t^*) dt^* = \frac{\delta}{\delta+\gamma} \left[\Lambda(t) - \lambda e^{-\delta(t-\mu-\delta\nu^2/2)} \Phi\left(\frac{t-\mu-\delta\nu^2}{\nu}\right)\right]. \tag{A.3}$$

Proof. First, note that

$$\int_{-\infty}^{t} e^{\delta t^*}\gamma e^{\gamma t^*}\Phi\left(\frac{t^*-\mu+\gamma\nu^2}{\nu}\right)\mathrm{d}t^*$$
$$=\frac{\gamma}{\delta+\gamma}\left[e^{(\delta+\gamma)t}\Phi\left(\frac{t-\mu+\gamma\nu^2}{\nu}\right)-\int_{-\infty}^{t} e^{(\delta+\gamma)t^*}\frac{1}{\nu}\phi\left(\frac{t^*-\mu+\gamma\nu^2}{\nu}\right)\mathrm{d}t^*\right].$$

Then,

$$\int_{-\infty}^{t} e^{-\delta(t-t^*)}\Lambda'(t^*)\mathrm{d}t^*=\int_{-\infty}^{t} e^{-\delta(t-t^*)}\frac{\partial}{\partial t^*}\left[\lambda e^{\gamma(t^*-\mu+\gamma\nu^2/2)}\Phi\left(\frac{t^*-\mu+\gamma\nu^2}{\nu}\right)\right]\mathrm{d}t^*$$
$$=\lambda e^{-\delta t-\gamma\mu+\gamma^2\nu^2/2}\left[\frac{\gamma}{\delta+\gamma}e^{(\delta+\gamma)t}\Phi\left(\frac{t-\mu+\gamma\nu^2}{\nu}\right)+\frac{\delta}{\delta+\gamma}\int_{-\infty}^{t} e^{(\delta+\gamma)t^*}\frac{1}{\nu}\phi\left(\frac{t^*-\mu+\gamma\nu^2}{\nu}\right)\mathrm{d}t^*\right].$$

From Lemma A.1,

$$\int_{-\infty}^{t} e^{(\delta+\gamma)t^*}\frac{1}{\nu}\phi\left(\frac{t^*-\mu+\gamma\nu^2}{\nu}\right)\mathrm{d}t^*=e^{(\delta+\gamma)\mu+(\delta^2-\gamma^2)\nu^2/2}\Phi\left(\frac{t-\mu-\delta\nu^2}{\nu}\right).$$

Therefore,

$$\int_{-\infty}^{t}(1-e^{-\delta(t-t^*)})\Lambda'(t^*)\mathrm{d}t^*$$
$$=\Lambda(t)-\lambda e^{-\delta t-\gamma\mu+\gamma^2\nu^2/2}\left[\frac{\gamma}{\delta+\gamma}e^{(\delta+\gamma)t}\Phi\left(\frac{t-\mu+\gamma\nu^2}{\nu}\right)+\frac{\delta}{\delta+\gamma}e^{(\delta+\gamma)\mu+(\delta^2-\gamma^2)\nu^2/2}\Phi\left(\frac{t-\mu-\delta\nu^2}{\nu}\right)\right]$$
$$=\frac{\delta}{\delta+\gamma}\left[\Lambda(t)-\lambda e^{-\delta(t-\mu-\delta\nu^2/2)}\Phi\left(\frac{t-\mu-\delta\nu^2}{\nu}\right)\right].$$

□

We now use Lemma A.2 to calculate the scale in (12) normalized by $\Gamma(\rho)$:

$$\Gamma(\rho)^{-1}T_o(t)=\sum_{\mathcal{N}\subseteq\{1,\ldots,N\}}\mathbf{1}\{o\in\mathcal{N}\}\Lambda(\mathcal{N},t)$$
$$=\sum_{s=0}^{S}\sum_{n=1}^{N}\int_{-\infty}^{t}\mathbb{P}\left[o\in\mathcal{N}_i(t;v)\mid s_i(v)=s,n_i(v)=n,t_i^*(v)=t^*\right]\Lambda'_{ns}(t^*)\mathrm{d}t^*$$
$$=\sum_{s=0}^{S}\sum_{n=1}^{N}\int_{-\infty}^{t}(1-e^{-\delta_{ons}(t-t^*)})\Lambda'_{ns}(t^*)\mathrm{d}t^*$$
$$=\Lambda_o(t)+\sum_{s=0}^{S}\sum_{n\neq o}\frac{\delta_{ons}}{\delta_{ons}+\gamma_n}\left[\Lambda_{ns}(t)-\lambda_{ns}e^{-\delta_{ons}(t-\mu_{ns}-\delta_{ons}\nu_{ns}^2/2)}\Phi\left(\frac{t-\mu_{ns}-\delta_{ons}\nu_{ns}^2}{\nu_{ns}}\right)\right].$$

We can also use this result to calculate the probability of two countries sharing an idea. For $o \neq o'$,

$$
\begin{aligned}
&\sum_{s=0}^{S}\sum_{n=1}^{N}\int_{-\infty}^{t}\mathbb{E}\left[I_{oi}(t,v)I_{o'i}(t,v)\mid n_i(v)=n,t_i^*(v)=t^*\right]\Lambda'_{ns}(t^*)\mathrm{d}t^* \\
&=\sum_{s=0}^{S}\sum_{n=1}^{N}\int_{-\infty}^{t}(1-e^{-\delta_{ons}(t-t^*)})(1-e^{-\delta_{o'ns}(t-t^*)})\Lambda'_{ns}(t^*)\mathrm{d}t^* \\
&=\Gamma(\rho)^{-1}(T_o(t)+T_{o'}(t))-\sum_{s=0}^{S}\sum_{n=1}^{N}\int_{-\infty}^{t}(1-e^{-(\delta_{ons}+\delta_{o'ns})(t-t^*)})\Lambda'_{ns}(t^*)\mathrm{d}t^* \\
&=\Gamma(\rho)^{-1}(T_o(t)+T_{o'}(t))-\Lambda_o(t)-\Lambda_{o'}(t) \\
&\qquad-\sum_{s=0}^{S}\sum_{n\notin\{o,o'\}}\frac{\delta_{oo'ns}}{\delta_{oo'ns}+\gamma_n}\left[\Lambda_{ns}(t)-\lambda_{ns}e^{-\delta_{oo'ns}(t-\mu_{ns}-\delta_{oo'ns}\nu_{ns}^2/2)}\Phi\left(\frac{t-\mu_{ns}-\delta_{oo'ns}\nu_{ns}^2}{\nu_{ns}}\right)\right] \\
&=\sum_{s=0}^{S}\sum_{n\neq o}\frac{\delta_{ons}}{\delta_{ons}+\gamma_n}\left[\Lambda_{ns}(t)-\lambda_{ns}e^{-\delta_{ons}(t-\mu_{ns}-\delta_{ons}\nu_{ns}^2/2)}\Phi\left(\frac{t-\mu_{ns}-\delta_{ons}\nu_{ns}^2}{\nu_{ns}}\right)\right] \\
&\qquad+\sum_{s=0}^{S}\sum_{n\neq o'}\frac{\delta_{o'ns}}{\delta_{o'ns}+\gamma_n}\left[\Lambda_{ns}(t)-\lambda_{ns}e^{-\delta_{o'ns}(t-\mu_{ns}-\delta_{o'ns}\nu_{ns}^2/2)}\Phi\left(\frac{t-\mu_{ns}-\delta_{o'ns}\nu_{ns}^2}{\nu_{ns}}\right)\right] \\
&\qquad-\sum_{s=0}^{S}\sum_{n\notin\{o,o'\}}\frac{\delta_{oo'ns}}{\delta_{oo'ns}+\gamma_n}\left[\Lambda_{ns}(t)-\lambda_{ns}e^{-\delta_{oo'ns}(t-\mu_{ns}-\delta_{oo'ns}\nu_{ns}^2/2)}\Phi\left(\frac{t-\mu_{ns}-\delta_{oo'ns}\nu_{ns}^2}{\nu_{ns}}\right)\right],
\end{aligned}
$$

where $\delta_{oo'ns} \equiv \delta_{ons} + \delta_{o'ns}$ and the second to last equality follows from Lemma A.2. We get the probability by dividing this result by $\sum_{s=0}^{S}\sum_{n=1}^{N}\Lambda_{ns}(t)$.

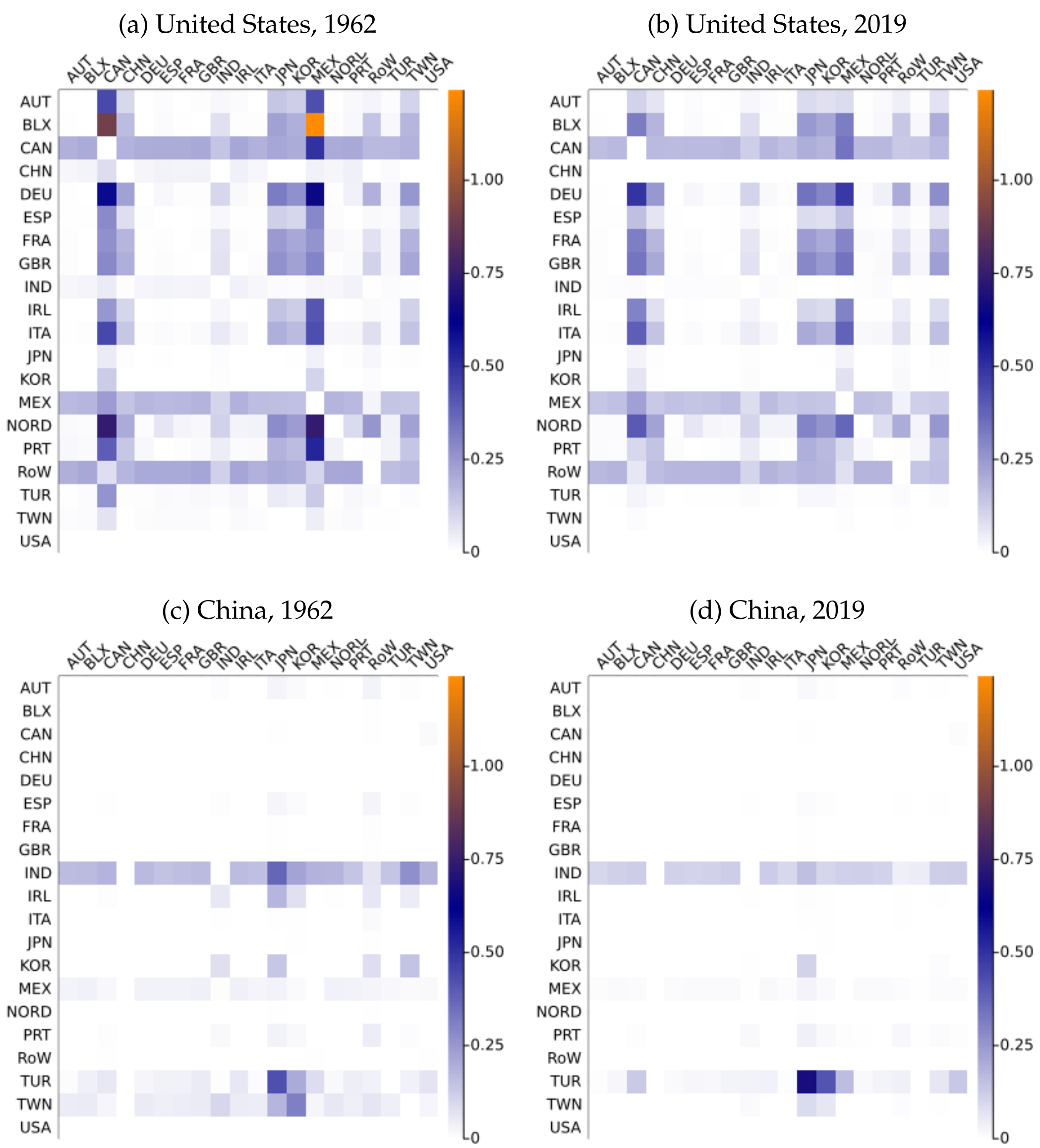
B Additional Results

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

Country code	Country name	γ_n	λ_{n0}	λ_{n1}	μ_{n1}	ν_{n1}
AUT	Austria	6.6e-5	0.011	0.003	1,987	29
BLX	Benelux	0.0	0.101	0.006	1,989	27
CAN	Canada	0.0	0.058	0.006	1,967	56
CHN	China	0.051	0.0001	0.049	1,999	6.2
DEU	Germany	0.0	0.026	0.005	1,983	37
ESP	Spain	0.016	0.001	0.021	1,985	16
FRA	France	0.0	0.018	0.002	1,980	38
GBR	United Kingdom	0.0	0.036	0.0	1,996	42
IND	India	0.0	0.007	0.0	2,001	49
IRL	Ireland	0.054	0.001	0.010	1,998	8.3
ITA	Italy	0.004	0.006	0.016	1,976	23
JPN	Japan	0.001	0.054	0.072	1,968	3.4
KOR	Republic of Korea	0.007	0.002	0.009	1,997	6.3
MEX	Mexico	0.0	0.111	0.001	1,990	36
NORD	Nordic Countries	3.0e-6	0.038	1.2e-5	1,988	46
PRT	Portugal	0.0	0.019	0.0	2,006	13
RoW	Rest of World	5.1e-5	0.216	0.048	1,985	39
TUR	Turkey	0.0	0.010	0.0	2,005	18
TWN	Taiwan	9.6e-5	0.033	0.012	1,984	11
USA	United States	0.0	0.248	7.0e-6	1,990	49

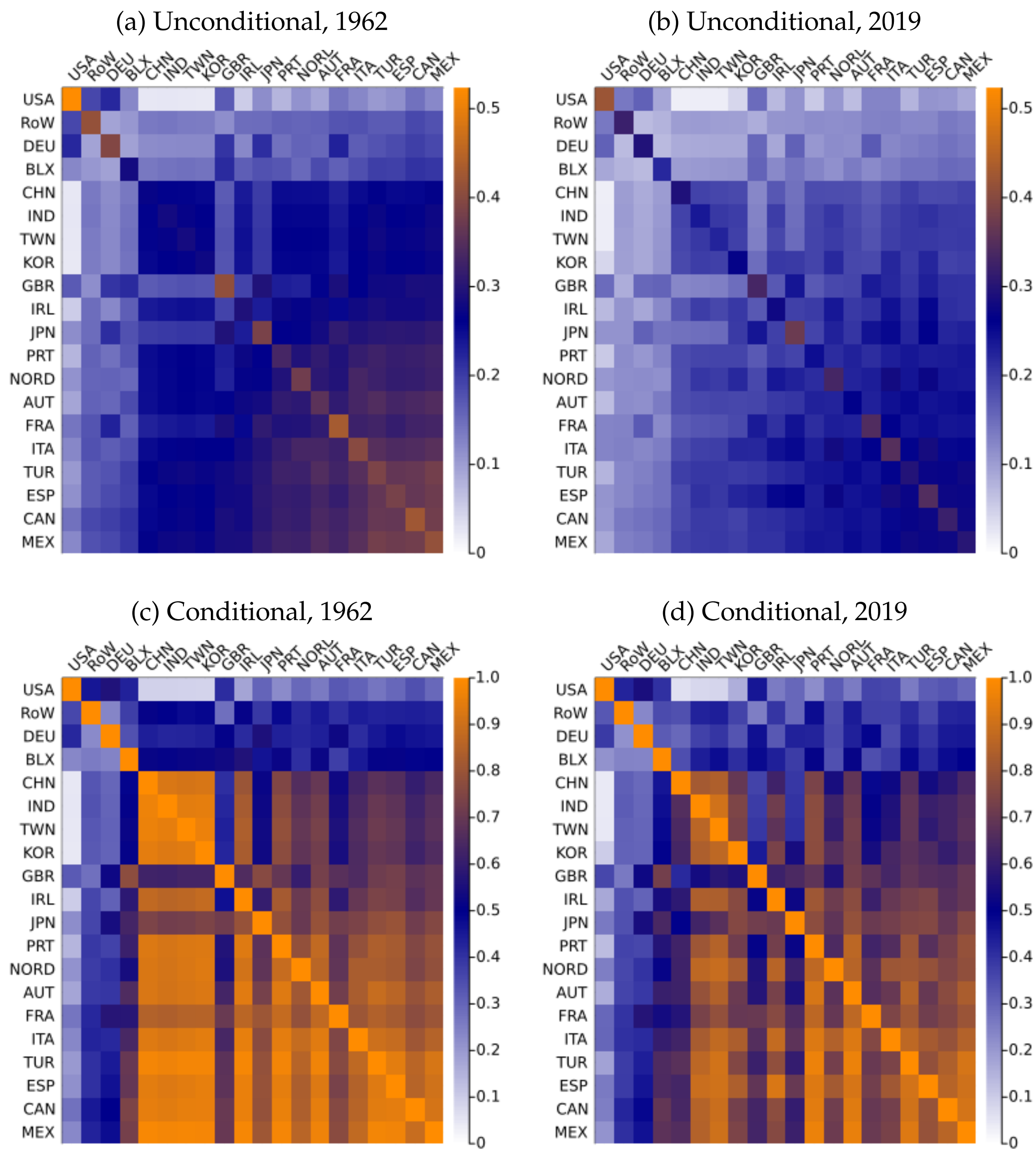
Notes: Benelux includes Belgium, Luxembourg, and Netherlands, while Nordic countries include to Denmark, Sweden, and Finland. Estimated parameters of the innovation process in (6) for the one-surge model. μ_{n1} is the (average) year of the innovation surge, $s = 1$, for each country.

Figure B.1: Estimation results: Elasticities.



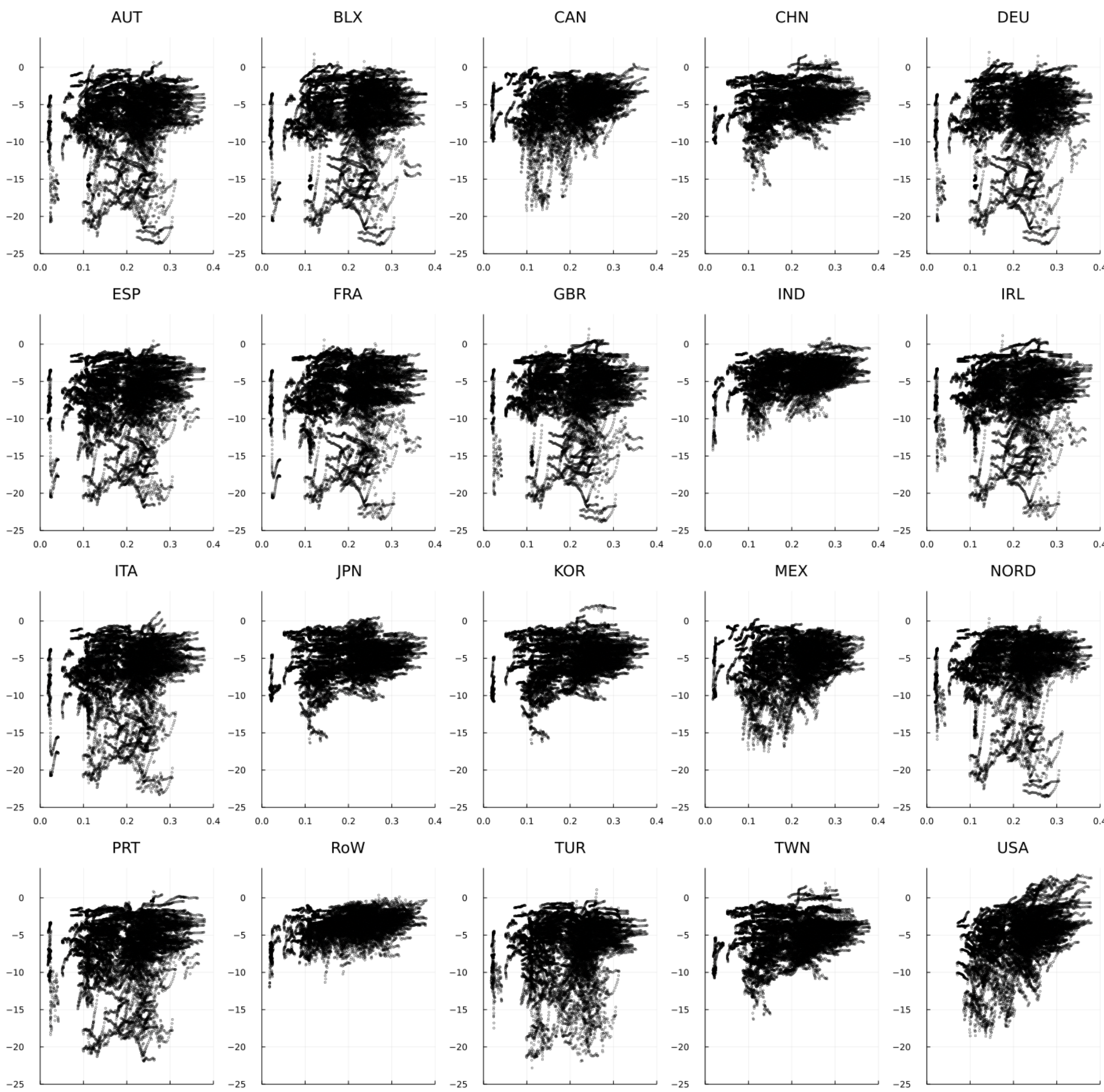
Notes: Estimates from the model with $S = 1$. $\varepsilon_{oo'd}(t)$ are calculated using (20).

Figure B.2: Estimation results: Probabilities of shared knowledge.



Notes: Estimates from the model with $S = 1$. Unconditional sharing probabilities are calculated as $K_{oo'}(t)$ in (21), relative to $T_o(t)$ in (12), while conditional probabilities are calculated as $K_{oo'}(t)/K_{oo}(t)$.

Figure B.3: Cross-price elasticities and shared knowledge, by destination country.



Notes: Estimates from the model with $S = 1$. $\varepsilon_{oo'd}(t)$ are calculated using (20), by destination country d, while sharing probabilities are calculated using (21).