

NBER WORKING PAPER SERIES

GLOBAL INNOVATION AND KNOWLEDGE DIFFUSION

Nelson Lind
Natalia Ramondo

Working Paper 29629
<http://www.nber.org/papers/w29629>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2022

We benefited from comments from Jess Benhabib, Simon Fuchs, In-Koo Cho, Hugo Hopenhayn, Jesse Perla, Ana Maria Santacreu, Chris Tonetti, and Mike Waugh, as well as seminar participants at various conferences and institutions. This research is supported by NSF grant #1919372. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2022 by Nelson Lind and Natalia Ramondo. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Global Innovation and Knowledge Diffusion
Nelson Lind and Natalia Ramondo
NBER Working Paper No. 29629
January 2022
JEL No. F1,O3

ABSTRACT

We develop a Ricardian model of trade in which countries innovate ideas that diffuse across the globe. In this model, the forces of innovation and diffusion combine to shape trade substitution patterns. Innovation makes a country technologically distinct, reducing their substitutability with other countries, while diffusion between countries generates technological similarity and increases head-to-head competition. In the special case of an innovation-only model where countries do not share ideas, productivities are independent across space, and the demand system is CES. As a consequence, departures from CES expenditure reveal diffusion patterns. Our theoretical results provide a mapping between the dynamics of observable trade flows and the dynamics of innovation and knowledge diffusion.

Nelson Lind
Department of Economics
Emory University
1602 Fishburne Drive
Rich Bldg., 3rd Floor
Atlanta, GA 30322-2240
nelslind@gmail.com

Natalia Ramondo
Boston University
Department of Economics
270 Bay State Rd # 445
Boston, MA 02215
and NBER
nramondo@bu.edu

1 Introduction

A central feature of growth is that new ideas improve production technologies. At the same time, as [Romer \(1990\)](#) emphasized, ideas are inherently non-rivalrous. In principle, sharing ideas should lead to similarity in production methods and productivity. However, not all people, firms, and countries are equally able to adopt new ideas in their original form, or to adopt them at all. Barriers to adoption can be large ([Parente and Prescott, 1994](#)), making ideas excludable, and hence, creating dissimilarity in production methods and productivity. The dynamics of innovation and knowledge diffusion can then shape similarities and differences in technology across countries. However, it has been challenging to disentangle directly from the data the dynamics of knowledge, as well as the effects coming from innovating new ideas versus adopting existing ideas that were created somewhere else.¹

In this paper, we formalize the idea that the footprints of knowledge dynamics are reflected in the dynamics of expenditure patterns, and more specifically, in the substitution patterns of trade shares between countries. From a Ricardian perspective, these observable outcomes are the consequence of the distribution of productivity across countries and goods at any point in time. Since heterogeneity in productivity determines patterns of comparative advantage and specialization across countries, trade flows should reflect the creation and spread of ideas that give rise to the different technologies used in the production of goods. In particular, countries with more similar technologies should specialize in the same or very similar goods and hence, they should have very elastic substitution patterns with each other.

We connect a Ricardian model of trade featuring a max-stable multivariate Fréchet distribution of productivity over space to a model of global innovation and knowledge diffusion. The creation and spread of ideas across countries over time changes the world-wide productivity distribution, and hence the observed patterns of trade, captured by the expenditure substitution elasticities.

The model starts by clearly separating innovation from diffusion. Ideas of different quality are discovered over time by individual countries according to a Poisson

¹For efforts to measure technology adoption and knowledge diffusion directly see [Comin and Hobijn \(2004\)](#), [Comin and Hobijn \(2010\)](#), and [Bloom, Hassan, Kalyani, Lerner, and Tahoun \(Bloom et al.\)](#). Also see [Comin and Mestieri \(2014\)](#) for a review.

process — innovation is exogenous. Conditional on an innovation’s discovery location and time, other countries learn about the idea through time. While any location can use an idea, leading to non-rivalry, locations may differ in the ability to apply the idea, leading to partial excludability. Over time, the applications of each idea across locations may change over time, which, together with the creation of new ideas, determines the dynamics of knowledge. The concept of the *applicability* of an idea in a particular location is key for distinguishing innovation from diffusion and for obtaining productivity distributed as a multivariate max-stable Fréchet with an arbitrary correlation over space. We establish this last result by applying the spectral representation theorem for max-stable processes (De Haan, 1984; Penrose, 1992; Stoev and Taqqu, 2005; Kabluchko, 2009; Wang and Stoev, 2010), which generates max-stable processes from Poisson processes.²

A max-stable Fréchet distribution of productivity allows for heterogenous patterns of correlation in productivity across production locations, and delivers closed-form solutions for expenditure shares. In particular, this distribution spans the entire class of Generalized Extreme Value (GEV) import demand system (McFadden, 1978). This class features rich substitution elasticities, departs from Independence of Irrelevant Alternatives (IIA), approximates any demand system satisfying gross substitutes (Fosgerau et al., 2013), and encompasses the large body of quantitative trade models inspired by Eaton and Kortum (2002) (see Lind and Ramondo, 2021, for details).

Loosely speaking, the lack of correlation in productivity across locations reflects knowledge acquired mainly through innovation, while high correlation in productivity reflects knowledge acquired mainly through diffusion. In turn, high correlation leads to high substitutability in expenditure, while low correlation in productivity leads to low elasticities of substitution. In the extreme, if ideas are never shared between countries, there can be no similarity in technology, productivity is independent across countries, expenditure has a Constant Elasticity of Substitution (CES), and IIA holds, as it is the case in Eaton and Kortum (2001).

Additionally, a key feature of Ricardian trade models is head-to-head competition: Producers of the same good compete for a market, with the lowest-cost source supplying the market. Because in our model diffusion occurs in a particular idea to produce a given good or variety, it is diffusion that tightens up head-to-head

²The spectral representation theorem for max-stable processes has previously been used by Dagsvik (1994) in the context of decision theory.

competition across suppliers in different countries, and hence, increases their substitutability in demand.

Overall, the link that we establish between GEV import demand and the underlying innovation and diffusion processes of ideas provides a framework to infer knowledge dynamics from estimates of trade elasticities over time.

Related literature. This paper builds on the literature that generates Fréchet productivity from Poisson processes. The basic idea was introduced in [Kortum \(1997\)](#) and used in the context of a trade model by [Eaton and Kortum \(2001\)](#). If the production technology is determined by the best “idea”, or blueprint, and if ideas become available according to a Poisson process, then after a sufficiently long period of time productivity can be approximated by an extreme value distribution. We depart from this approach by leveraging tools from the spectral representation of max-stable processes rather than from the extreme value theory. In this way, we are able to: (1) generalize from the case of independent Fréchet to the case of Fréchet with arbitrary dependence, while preserving the max-stability property key to the tractability of models of head-to-head competition; and (2) provide exact rather than asymptotic results.

The literature that followed [Kortum \(1997\)](#) and [Eaton and Kortum \(2001\)](#), such as [Buera and Oberfield \(2020\)](#), has been restricted to the case of independent Fréchet. As a consequence, the resulting models struggle to capture how the offsetting forces of innovation and diffusion shape correlation in productivity. Moreover, in those models, diffusion is modeled as sampling ideas across, rather than within, goods, so that sharing ideas does not increase head-to-head competition across suppliers.³ A key feature of our model is that, because ideas are specific to goods, diffusion occurs within a good across locations. In this way, this knowledge flow increases head-to-head competition among suppliers, leading to more elastic expenditure elasticities.

Our treatment of diffusion is closer to the model of international diffusion in [Eaton and Kortum \(1999\)](#). In their model, as in ours, ideas are specific to goods, have a common quality component across countries, and diffuse with a lag to other locations. Their model, however, does not include trade flows, and it is restricted to an independent Fréchet technological frontier. In contrast, we expand the model to

³This is also the case in [Cai et al. \(2020\)](#). The main difference is that their model of innovation and diffusion incorporates trade in many sectors. Additionally, while in [Buera and Oberfield \(2020\)](#), international trade is a vehicle for diffusion, in [Cai et al. \(2020\)](#), as in our model, it is not.

include Ricardian trade and use tools from the literature on max-stable processes to obtain a Fréchet technological frontier with arbitrary correlation across locations. By doing so, we are able to link the underlying knowledge dynamics to the properties of observed trade shares.

Papers such as [Perla et al. \(2021\)](#) and [Sampson \(2016b\)](#) introduce models where (endogenous) innovation, diffusion, and trade interact to generate growth.⁴ While our model is not as rich as theirs in terms of the micro foundations for innovation and diffusion, we depart from their modeling approach and use tools from the literature on max-stable process together with Ricardian trade as in [Eaton and Kortum \(2002\)](#). This allows us to generate import demand systems that are not restricted to CES. We do follow previous work by the same authors ([Perla and Tonetti, 2014](#); [Sampson, 2016a](#)) and model diffusion as a process in which not only the ideas currently used in production are evaluated for adoption, but any idea is.

Finally, our approach to model global innovation and diffusion has similar consequences for productivity to the ones in the model by [Benhabib et al. \(2021\)](#) where "[i]nnovation stretches the distribution [of productivity], while adoption compresses it." We go a step further and, in the context of an open economy, we link those distributions to Ricardian trade, using tools suitable to obtain closed-form solutions in models of head-to-head competition.

2 Set Up

Time is continuous and indexed by $t \in \mathbb{R}$. The world economy has a finite number of locations, $\ell \in \mathcal{L} \equiv \{1, \dots, L\}$. Each location, e.g. a country, represents a labor market with wage given by $W_\ell(t)$ that is populated by a measure $N_\ell(t)$ of individuals. These individuals are immobile across locations, inelastically supply their labor, consume a continuum of tradable goods, $v \in [0, 1]$, and have CES preferences with elasticity of substitution $\eta > 0$.

Production of each good v in location ℓ is done with an only-labor constant-return-to-scale technology,

$$Y_\ell(t, v) = Z_\ell(t, v)L_\ell(t, v),$$

where $Z_\ell(t, v)$ is good-location-time specific productivity. We focus on the case of

⁴Using a multi-sector model. the focus of [Sampson \(2016b\)](#) is, similarly to [Cai et al. \(2020\)](#)'s, on how innovation and diffusion affect comparative advantage across countries.

frictionless trade so that there is a common market for goods. This simplifies the notation to only one location subscript without changing any of our results.

We follow [Eaton and Kortum \(2002\)](#) (EK) and model productivity as a random draw across varieties and locations. Over time, productivity is a stochastic process, $\{Z_\ell(t, v)\}_{(\ell, t) \in \{1, \dots, L\} \times \mathbb{R}}$.

We proceed as follows. First, we develop a model of innovation and diffusion where productivity is the result of locations creating and adopting ideas, and where innovation tends to reduce technological similarity, while diffusion tends to increase it. Next, we show that this structure, combined with a Poisson assumption on innovation and an independence assumption on diffusion, is necessary and sufficient to characterize a distribution of productivity that exhibits correlation over space and the property of max stability.⁵ This property is key to ensure closed-form expressions for aggregate expenditure and prices in models of head-to-head competition. The model encompasses the special case of independent productivity across production locations, as in EK, but allows for richer patterns of technological similarity across countries, as in [Lind and Ramondo \(2021\)](#).

2.1 Ideas and productivity

For each good, there exists an infinite, but countable, set of ideas: $i = 1, 2, \dots$. Each idea represents a physical production technique (e.g., a blueprint) for a good and may be applied in different locations. The productivity of each idea has two components. First, the *quality* of the idea, $Q_i(v)$, represents the overall efficiency of the idea and is common to all locations. Second, the idea's productivity has a location and time specific component, $A_{i\ell}(t, v)$, which we refer to as the *applicability* of the idea in location ℓ at time t . This term captures the costs of technology adoption, which may differ across both locations and time, creating differences in the productivity of the idea (see [Parente and Prescott, 1994](#)). Applicability of an idea is the key concept in our model. It will allow us to cleanly distinguish between innovation and diffusion, introduce time-varying technological similarity (or lackof) across locations, and time-varying correlation (or lack of) in productivity over space.

⁵Productivity is a *max stable process* with Fréchet marginal distributions if for any integer J and $v_j \geq 0$, $\ell_j \in \{1, \dots, L\}$ and $t_j \in \mathbb{R}$ for $j = 1, \dots, J$ the distribution of $\max_{j=1, \dots, J} v_j Z_{\ell_j}(t_j, v)$ is Fréchet. See [Stoev and Taqqu \(2005\)](#).

Together, quality and applicability combine to shape the productivity of idea i at location ℓ at time t : $Q_i(v)A_{i\ell}(t, v)$. The first term generates heterogeneity in productivity across ideas, while the second term generates, for each idea, heterogeneity in productivity across locations.

Ideas can be used anywhere, which makes them non-rival. They are, however, partially excludable due to differences in applicability across locations: some ideas are more productive in some locations than in others. For example, a location is completely excluded from using an idea if its applicability at a point in time is zero. Among locations with knowledge of an idea—i.e. with positive applicability—technological similarity arises from similarity in applicability.

Finally, we assume that productivity for good v in location ℓ at time t is the result of choosing the best idea available to them. The following assumption summarizes the overall structure for productivity.

Assumption 1 (Technology Adoption). *For each $v \in [0, 1]$, there exists an infinite, but countable, set of ideas, $i = 1, 2, \dots$, with quality, $Q_i(v) > 0$, and applicabilities, $A_{i\ell}(t, v) \geq 0$ for each $\ell = 1, \dots, L$ and $t \in \mathbb{R}$, such that*

$$Z_\ell(t, v) = \max_{i=1,2,\dots} Q_i(v)A_{i\ell}(t, v). \quad (1)$$

If a new idea i to produce a good v becomes available to a location ℓ , this idea gets adopted only if its overall efficiency is higher than the efficiency of an already available idea. The overall efficiency of an idea can be high because its quality $Q_i(v)$ is high, or because its applicability $A_{i\ell}(t, v)$ in that location ℓ at time t is high. However, if that location does not have a very good application of a high-quality idea, some other location could easily overtake it via diffusion—i.e. by drawing a better application of the idea.

Changes in productivity over time reflect changes in the applicability of ideas, either due to the innovation of new ideas, or the diffusion of existing ideas, as described next.

2.2 Innovation and diffusion

The applicability of an idea at a location can change over time due to innovation and diffusion. In particular, an idea can be unknown to a location either because

the idea has not been innovated yet—and its applicability is zero in all locations— or because the location has not learned about it from another location.⁶ While innovation makes an idea available for the very first time somewhere—when applicability in some location goes from zero to some positive number, diffusion either makes the idea available to new locations, or changes the way an idea is applied in locations with prior knowledge of the idea.

Formally, the time when an idea first arrives in location ℓ is $t_{i\ell}(v) \equiv \inf\{t \in \mathbb{R} \mid A_{i\ell}(t, v) > 0\}$, while the time when an idea is innovated—its *discovery time*—is the first time when the idea becomes available to some location, $t_i^*(v) \equiv \min_{\ell} t_{i\ell}(v)$.

We next put stochastic structure on innovation.

Assumption 2 (Poisson Innovation). *The innovation history, $\{Q_i(v), t_i^*(v)\}_{i=1,2,\dots}$, consists of the points of a Poisson process. Ideas with quality q at time t get innovated with intensity $\theta q^{-\theta-1} dq d\Lambda(t)$ for some $\theta > 0$ and measure Λ .*

Assumption 2 implies that the expected number of ideas discovered up to time t with quality above \underline{q} is $\underline{q}^{-\theta} \Lambda(t)$. The variable $\Lambda(t)$ is the expected number of ideas with quality of at least 1 that have been discovered up to time t . Among ideas with quality of at least \underline{q} , quality is independently and identically distributed Pareto with shape θ and lower bound \underline{q} . A lower θ means a fatter tail so that innovation can create very good ideas. Furthermore, each idea's quality is independent of its discovery time.

An intuitive way of thinking about the distribution of ideas' quality is that qualities are drawn at the beginning of time and locations are just randomly discovering them. Assumption 2 means that there are many more low-quality ideas than high-quality ideas to be discovered, so that most likely a location will discover a low-quality idea. However, occasionally, they get a really good idea. Our structure departs from the previous literature (Kortum, 1997) in that locations do not draw ideas' quality (distributed Pareto), conditional on the idea's (Poisson) arrival. Rather, ideas of different qualities arrive at different rates as indicated by Assumption 2. What do we gain with this structure? Rather than using the extreme value theorem, which leads asymptotically to an independent Fréchet distribution for productivity, our structure for technology leads to exact results and to a max-stable

⁶While the infinite set of ideas for good v , $i = 1, 2, \dots$, represents all the physical production techniques that will ever exist for a good, before an idea is innovated, applicability is zero everywhere and all locations are excluded from using the idea.

multivariate Fréchet distribution for productivity with arbitrary correlation. As we show in Section 3, the Pareto tail of quality determines the shape of the Fréchet distribution for productivity; if this assumption did not hold, productivity would no longer be distributed Fréchet.

To distinguish between innovation and diffusion we group ideas according to their discovery time, and keep track of the distribution of applicability among those ideas. In this way, while innovation refers to the very first time applicability turns positive, diffusion refers to any change to the applicability of an idea across production locations after its discovery. We only require one restriction on the diffusion process to ensure max-stable productivity.

Assumption 3 (Independent Diffusion). *Applicability is a measurable stochastic process, which is independent of quality and independent and identically distributed across ideas conditional on discovery time with $\int_{-\infty}^t \mathbb{E}[A_{i\ell}(t, v)^\theta \mid t_i^*(v) = t^*] d\Lambda(t^*) < \infty$ for all $t \in \mathbb{R}$.*

Assumption 3 means that improvements in applicability are just as likely to occur for low quality as for high quality ideas—locations cannot select only the highest quality ideas to learn about. One consequence of Assumption 3 is that we only need to condition the distribution of applicability on the discovery time of an idea in order to keep track of how the idea diffuses. Formally, the state variable for the world economy at time t is given by the measure of ideas with quality of at least 1 and applicability below some vector (a_1, \dots, a_L) ,

$$M(a_1, \dots, a_L; t) \equiv \int_{-\infty}^t \mathbb{P}[A_{i\ell}(t, v) \leq a_\ell \quad \forall \ell \mid t_i^*(v) = t^*] d\Lambda(t^*). \quad (2)$$

Besides Assumption 3, we require no further restrictions on the dynamics of diffusion to ensure max-stability. However, to clearly separate the effects of innovation from the effects of diffusion and to ensure that M evolves continuously over time we impose a further restriction. We assume that each idea is innovated in a unique *discovery location*, $\ell_i^*(v)$,⁷ and at a unique time. This assumption implies that ideas arrive to locations other than the discovery location with a lag. As a consequence, no idea is initially shared and all shared knowledge arises from diffusion. The unique discovery time implies that discovery times are continuously distributed, and we denote the rate at which ideas with quality above one are discovered in ℓ^*

⁷Formally, for each idea, there exists a location, $\ell_i^*(v)$, such that for $\ell = \ell_i^*(v)$, $t_i^*(v) = t_{i\ell}(v)$ and $\forall \ell \neq \ell_i^*(v)$, $t_i^*(v) < t_{i\ell}(v)$.

by $\lambda_{\ell^*}(t) \equiv \mathbb{P}[\ell_i^*(v) = \ell^* \mid t_i^*(v) = t^*] \partial \Lambda(t) / \partial t$.⁸

Together these uniqueness assumptions imply that M in (2) becomes additively separable,

$$M(a_1, \dots, a_L; t) = \sum_{\ell^*=1}^L \int_{-\infty}^t F(a_1, \dots, a_L \mid t^*, \ell^*; t) \lambda_{\ell^*}(t^*) dt^*, \quad (3)$$

where $F(a_1, \dots, a_L \mid t^*, \ell^*; t) \equiv \mathbb{P}[A_{i\ell}(t, v) \leq a_\ell \ \forall \ell \mid t_i^*(v) = t^*, \ell_i^*(v) = \ell^*]$ denotes the conditional distribution of applicability across all locations. The dynamics of M can then be decomposed into an innovation effect and a diffusion effect as follows:

$$\frac{\partial}{\partial t} M(a_1, \dots, a_L; t) = \underbrace{\sum_{\ell^*=1}^L F^*(a_{\ell^*} \mid \ell^*, t; t) \lambda_{\ell^*}(t)}_{\text{Innovation Effect}} + \underbrace{\sum_{\ell^*=1}^L \int_{-\infty}^t \frac{\partial}{\partial t} F(a_1, \dots, a_L \mid \ell^*, t^*; t) \lambda_{\ell^*}(t^*) dt^*}_{\text{Diffusion Effect}}, \quad (4)$$

where $F^*(a_{\ell^*} \mid \ell^*, t; t) \equiv \mathbb{P}[A_{i\ell^*}(t, v) \leq a_{\ell^*} \mid \ell_i^*(v) = \ell^*, t_i^*(v) = t]$ denotes the conditional distribution of applicability in the discovery location. The first term in (4) captures how the innovation of new ideas at time t impacts the evolution of M . Applicability is concentrated in the discovery location, and the joint distribution of applicability among new ideas is simply the marginal distribution in the discovery location. Hence, the overall effect of innovation is additively separable across locations because ideas arrive as a Poisson process (Assumption 2) and have a unique discovery location. In contrast, the second term captures the diffusion of ideas that were innovated in the past: though separable across discovery locations (indexed by ℓ^*), this term is not separable across production locations (indexed by ℓ). Conditional on the discovery time and location of an idea, the effect of diffusion is captured by the evolution of the joint distribution of applicability.

So far, our model delivers a simple expression for the global evolution of knowledge, with a clear distinction between the contribution of innovation versus diffusion. Diffusion allows ideas, which are not shared initially, to be shared across production locations, while innovation reduces the prevalence of shared ideas. Although our model is parsimonious, it captures the non-rival trait of ideas—they can be shared across locations—while also allowing for excludability—some locations have limited access to ideas. In turn, the dynamics of innovation and diffusion

⁸Endogenous innovation would entail to endogenize λ_{ℓ^*} —e.g., proportional to the number of researchers in a country.

give rise to a max-stable Fréchet distribution of productivity over time and space, as we show next.

3 Productivity as Max-Stable Fréchet

We now provide a closed-form characterization for the distribution of productivity across locations and its evolution over time, given the dynamics of innovation and diffusion. The key result is that the proposed structure for innovation and diffusion generates productivity that is distributed max-stable multivariate Fréchet across locations. In turn, this distribution leads to closed-form results for expenditure shares.

Proposition 1 (Max-stable Fréchet Productivity). *Productivity is a measurable max-stable process with Fréchet marginal distributions if and only if Assumptions 1, 2, and 3 hold. In this case, the joint distribution of productivity across locations at time t is max-stable multivariate Fréchet,*

$$\mathbb{P}[Z_1(t, v) \leq z_1, \dots, Z_L(t, v) \leq z_L] = \exp[-G(T_1(t)z_1^{-\theta}, \dots, T_L(t)z_L^{-\theta}; t)]. \quad (5)$$

The shape parameter is $\theta > 0$, the scale parameter for each location is

$$T_\ell(t) \equiv \int a_\ell^\theta dM(a_1, \dots, a_L; t), \quad (6)$$

and the correlation function is

$$G(x_1, \dots, x_L; t) \equiv \int \max_\ell \frac{a_\ell^\theta}{T_\ell(t)} x_\ell dM(a_1, \dots, a_L; t). \quad (7)$$

Proof. See Appendix A. □

The marginal distribution of productivity in each country is Fréchet and characterized by the scale parameter $T_\ell(t)$ and shape parameter θ . The shape parameter regulates the heterogeneity of productivity over the continuum of goods, and it is the same Pareto-tail parameter that regulates the heterogeneity of the quality of ideas. When quality is more fat-tailed (lower θ), there is more dispersion in productivity across goods within each production location. The location's scale parameter $T_\ell(t)$ captures average applicability of ideas in location ℓ at time t .

Turning to the joint distribution of productivity, the correlation function G captures the dependence structure of productivity across locations and, therefore, patterns of technological similarity. Indeed, this function directly reflects the state of knowledge, M . The correlation has two key properties. First, it is homogenous of degree one, ensuring that the joint distribution of productivity is max-stable—i.e. the distribution of the maximum of productivity across any subset of production locations is also Fréchet with shape θ . Second, the function G has mixed partial derivatives that exist (almost everywhere) and are continuous up to order K , with the k 'th partial derivative of G with respect to k distinct arguments non-negative if k is odd and non-positive if k is even. This “sign-switching” property allows us to sign substitution elasticities, which, as we show in Section 4, reflect the curvature (or lackof) of this function.⁹

From (5), it is easy to see that the case of an additive correlation function corresponds to independent productivity across locations. Correlation in productivity arises when G has curvature. Previous models of innovation and diffusion that generate a max-stable Fréchet productivity distribution exclusively focus on the case of independent productivity (e.g. [Kortum, 1997](#); [Eaton and Kortum, 1999, 2001](#); [Buera and Oberfield, 2020](#)).

We next turn to concrete examples to illustrate the interactions between innovation and diffusion to generate correlation (or lackof) in productivity.

3.1 Examples

Suppose that ideas are never shared across locations after being innovated—ideas stay in their discovery location. In this case, the idea’s applicability is degenerate at zero in all locations except for the discovery location at all times. Hence, the measure of ideas becomes

$$M(a_1, \dots, a_L; t) = \sum_{\ell^*=1}^L \int_{-\infty}^t F^*(a_\ell \mid \ell^*, t^*; t) \lambda_{\ell^*}(t^*) dt^*. \quad (8)$$

Using Proposition 1, we get the productivity distribution,

$$\mathbb{P}[Z_1(t) \leq z_1, \dots, Z_L(t) \leq z_L] = \exp \left[- \sum_{\ell=1}^L T_\ell(t) z_\ell^{-\theta} \right], \quad (9)$$

⁹The function G is also unbounded: $G(x_1, \dots, x_K) \rightarrow \infty$ as $x_k \rightarrow \infty$ for any $k = 1, \dots, K$.

with scales given by average applicability among the stock of ideas previously discovered in the location, and an additive correlation function.¹⁰

The lack of shared ideas across locations, together with the Poisson assumption on the arrival of ideas, leads to independent productivity across locations. This is precisely what leads to independence of productivity in [Eaton and Kortum \(2001\)](#) — an innovation-only model based on Poisson arrival of ideas and where ideas never leave their discovery location. It is also the reason for independence of productivity in models of diffusion such as [Buera and Oberfield \(2020\)](#). In that model, locations sample productivities in other locations and update their productivity depending on the realization of a proportional adjustment, which is drawn from a Pareto distribution. In terms of our model, diffusion in [Buera and Oberfield \(2020\)](#) generates an applicability level that is related to the sampled productivity level, but it also generates a new quality level and therefore it is effectively a completely new idea specific to the new location. Effectively, diffusion generates new ideas that are never shared and this leads to independent productivity over space.

The case presented above establishes that a necessary condition for correlation in productivity is the sharing of ideas across locations. To make this concrete, consider the case of ideas that are instantly shared across all location once they diffuse. Further, assume that once an idea is known in a production location, its applicability is distributed as a unit Fréchet, independent across locations, and with shape $\sigma > \theta$. The measure of ideas at time t is given by a combination of non-shared and shared ideas,

$$M(a_1, \dots, a_L; t) = \sum_{\ell=1}^L \left[e^{-a_{\ell}^{-\sigma}} (1 - \delta_{\ell}(t)) + e^{-\sum_{\ell'=1}^L a_{\ell'}^{-\sigma}} \delta_{\ell}(t) \right] \Lambda_{\ell}(t), \quad (10)$$

¹⁰Derivation of (9):

$$\begin{aligned} -\ln \mathbb{P}[Z_1(t) \leq z_1, \dots, Z_L(t) \leq z_L] &= \int \max_{\ell} a_{\ell}^{\theta} z_{\ell}^{-\theta} d \sum_{\ell=1}^L \int_{-\infty}^t \mathbb{P}[A_{i\ell}(t, v) \leq a_{\ell} \mid \ell_i^*(v) = \ell, t_i^*(v) = s] \lambda_{\ell}(s) ds \\ &= \sum_{\ell=1}^L \int_{-\infty}^t \int a_{\ell}^{\theta} z_{\ell}^{-\theta} d \mathbb{P}[A_{i\ell}(t, v) \leq a_{\ell} \mid \ell_i^*(v) = \ell, t_i^*(v) = s] \lambda_{\ell}(s) ds = \sum_{\ell=1}^L T_{\ell}(t) z_{\ell}^{-\theta}. \end{aligned}$$

where $\Lambda_\ell(t) \equiv \int_{-\infty}^t \lambda_\ell(s) \mathbf{d}s$ and

$$\delta_\ell(t) \equiv \int_{-\infty}^t \mathbb{P}[\exists \ell' \neq \ell \text{ s.t. } A_{i\ell'}(t, v) > 0 \mid \ell_i^*(v) = \ell, t_i^*(v) = s] \frac{\lambda_\ell(s)}{\Lambda_\ell(t)} \mathbf{d}s$$

is the fraction of ideas innovated in ℓ that have diffused to the rest of the world. In this case, the productivity distribution is

$$\mathbb{P}[Z_1(t) \leq z_1, \dots, Z_L(t) \leq z_L] = \exp \left\{ -\Gamma(\rho) \sum_{\ell=1}^L \left[(1 - \delta_\ell(t)) z_\ell^{-\theta} + \delta_\ell(t) \left(\sum_{\ell'=1}^L z_{\ell'}^{-\frac{\theta}{1-\rho}} \right)^{1-\rho} \right] \Lambda_\ell(t) \right\}, \quad (11)$$

where $\rho \equiv 1 - \theta/\sigma$.¹¹ This case is a "convex" combination of the case of independence and symmetric correlation. Productivity may be correlated across production locations and the extent of correlation depends on the fraction of ideas innovated by each country that have diffused to the rest of the world. In one extreme, if there is no diffusion, then $\delta_\ell(t) = 0$ for all discovery locations and this case reduces to (9) where ideas are never shared from the discovery location and productivity is independent Fréchet. On the other extreme, if all ideas diffuse immediately after being innovated, then $\delta_\ell(t) \rightarrow 1$ for all locations and we get a multivariate θ -Fréchet distribution with correlation parameterized by $\rho \in [0, 1]$.¹² The amount of correlation in productivity reflects the heterogeneity in the applicability of an idea across locations, captured by σ . As $\sigma \rightarrow \theta$, applicability is as fat-tailed as the quality component of an idea, $\rho \rightarrow 0$, and productivity is independent across locations, despite the existence of diffusion. In contrast, as $\sigma \rightarrow \infty$, there is no dispersion in applicability across locations, and $\rho \rightarrow 1$. In this case, productivity is perfectly correlated across locations and diffusion equalizes productivity everywhere.

¹¹Derivation of (11):

$$\begin{aligned} G(T_1(t)x_1, \dots, T_L(t)x_L; t) &= \int \max_{\ell} a_{\ell}^{\theta} x_{\ell} \mathbf{d}M(a_1, \dots, a_L) \\ &= \sum_{\ell=1}^L \int a_{\ell}^{\theta} x_{\ell} \mathbf{d}e^{-a_{\ell}^{-\sigma}} (1 - \delta_{\ell}(t)) \Lambda_{\ell}(t) + \int \max_{\ell} a_{\ell}^{\theta} x_{\ell} \mathbf{d} \prod_{\ell'=1}^L e^{-a_{\ell'}^{-\sigma}} \sum_{\ell=1}^L \delta_{\ell}(t) \Lambda_{\ell}(t) \\ &= \sum_{\ell=1}^L \Gamma(1 - \theta/\sigma) (1 - \delta_{\ell}(t)) \Lambda_{\ell}(t) x_{\ell} + \Gamma(1 - \theta/\sigma) \left(\sum_{\ell} x_{\ell}^{\sigma/\theta} \right)^{\theta/\sigma} \sum_{\ell=1}^L \delta_{\ell}(t) \Lambda_{\ell}(t). \end{aligned}$$

¹²The case of instantaneous diffusion coincides with the case of multinational production in [Ramondo and Rodríguez-Clare \(2013\)](#): Technologies from ℓ are correlated across production locations ℓ' , with correlation given by ρ , while technologies are independent across source countries ℓ . Our model of innovation and diffusion lays out the knowledge primitives behind the Fréchet multinational production model.

This example illustrates that not only the extent of diffusion determines the degree of correlation in productivity, but also that different applications of the same idea across production locations need to be relatively similar for correlation to arise. That is, differences in applicability across production locations must be small relative to differences in quality across ideas. If applicability were as dispersed as idea quality, specific applications of an idea are virtually new ideas because they generate productivity differences that are just as large. Technological similarity only arises when differences in productivity due to idea quality are large relative to heterogeneity in productivity due to location-specific applications of ideas.

Summing up, departures from independent productivity require not only shared ideas, but also relatively similar applicability of those ideas across locations. Put differently, differences in productivity across locations using the same idea cannot be too large.

As a final example, consider the case of exponential diffusion as in [Eaton and Kortum \(1999\)](#). Suppose that for an idea innovated in ℓ at time t , the probability of diffusion to ℓ' by time t' is $1 - e^{-\varepsilon_{\ell\ell'}(t'-t)}$ and is independent across locations. Hence, the fraction of ideas innovated in ℓ that have diffused to the locations in any set $\mathcal{L} \subset \{1, \dots, L\}$ but not yet diffused to any other locations is

$$\delta_{\ell}(\mathcal{L}, t) \equiv \int_{-\infty}^t \left\{ \prod_{\ell' \in \mathcal{L}} [1 - e^{-\varepsilon_{\ell\ell'}(t-t^*)}] \prod_{\ell' \notin \mathcal{L}} e^{-\varepsilon_{\ell\ell'}(t-t^*)} \right\} \frac{\lambda_{\ell}(t^*)}{\Lambda_{\ell}(t)} dt^*$$

where $\varepsilon_{\ell\ell} = \infty$. As in the previous example, assume that once an idea is known in a location its applicability is distributed Fréchet with shape $\sigma > \theta$ and independent across locations. The measure of ideas at time t is

$$M(a_1, \dots, a_L; t) = \sum_{\ell=1}^L \sum_{\mathcal{L} \subset \{1, \dots, L\}} e^{-\sum_{\ell' \in \mathcal{L}} a_{\ell'}^{-\sigma}} \delta_{\ell}(\mathcal{L}, t) \Lambda_{\ell}(t),$$

reflecting the fact that each discovery location ℓ may have ideas that already diffused to some locations but not others. Using again [Proposition 1](#), the distribution of productivity is

$$\mathbb{P}[Z_1(t) \leq z_1, \dots, Z_L(t) \leq z_L] = \exp \left\{ -\Gamma(\rho) \sum_{\ell=1}^L \sum_{\mathcal{L} \subset \{1, \dots, L\}} \left(\sum_{\ell' \in \mathcal{L}} z_{\ell'}^{-\frac{\theta}{1-\rho}} \right)^{1-\rho} \delta_{\ell}(\mathcal{L}, t) \Lambda_{\ell}(t) \right\},$$

where $\rho \equiv 1 - \theta/\sigma$. Again, if $\rho = 0$, productivity is independent across locations.

As our examples demonstrate, Proposition 1 enables us to characterize the consequences of diffusion for technological similarity by providing a closed-form solution for the joint distribution of productivity in terms of the underlying measure of ideas. Correlation in productivity reflects the joint distribution of applicability across production locations, a result of the dynamics of innovation and diffusion. In turn, correlation in productivity determines the patterns of Ricardian trade: Locations with very similar technology face strong head-to-head competition with each other. These are very substitutable locations from the perspective of the importer. Additionally, only diffusion gets reflected in changes in substitution patterns across patterns over time, as we show in the next section.

4 Expenditure and Substitution Elasticities

We now proceed to characterize how innovation and diffusion shape trade flows and their substitution patterns through the evolution of correlation in productivity across space. Our results connect models of knowledge diffusion to the large empirical literature estimating trade elasticities. The main takeaway is that departures from CES expenditure imply that ideas are shared across space.

We present results for the case of no trade costs to simplify the analysis. The extension to an economy with trade frictions is straightforward and does not change the main intuition of the results.¹³

We assume that good markets are perfectly competitive. Given wages $W_\ell(t)$, head-to-head competition means that the lowest-cost location serves the destination market for good v at time t ,

$$P(t, v) = \min_{\ell} \frac{W_\ell(t)}{Z_\ell(t, v)}.$$

Thanks to max-stability (see [Lind and Ramondo, 2021](#)), and further assuming that

¹³With an iceberg-type trade cost to ship goods from production location ℓ to final destination n , $\tau_{\ell n}(t) > 1$, the price index would have the subscript n , $P_n(t)$ and trade shares would have bilateral subscript, $\pi_{\ell n}(t)$. In turn, the elasticities will reflect the substitution patterns between source countries ℓ and ℓ' to serve a particular final destination n , $\varepsilon_{\ell, \ell', n}(t)$.

$\theta > \eta - 1$, the share of expenditure allocated to production location ℓ is

$$\pi_\ell(t) \equiv \frac{X_\ell(t)}{X(t)} = \frac{T_\ell(t)W_\ell(t)^{-\theta} G_\ell(T_1(t)W_1(t)^{-\theta}, \dots, T_L(t)W_L(t)^{-\theta}; t)}{G(T_1(t)W_1(t)^{-\theta}, \dots, T_L(t)W_L(t)^{-\theta}; t)} \quad (12)$$

where $G_\ell(x_1, \dots, x_L; t) \equiv \frac{\partial G(x_1, \dots, x_L; t)}{\partial x_\ell}$, and the price index is

$$P(t) = \gamma G(T_1(t)W_1(t)^{-\theta}, \dots, T_L(t)W_L(t)^{-\theta}; t)^{-\frac{1}{\theta}}$$

where $\gamma \equiv \Gamma\left(\frac{\theta+1-\eta}{\theta}\right)^{\frac{1}{1-\eta}}$.

The import demand system in (12) has the same functional form as choice probabilities in generalized extreme value (GEV) discrete choice models (McFadden, 1978, 1981). Accordingly, we refer to it as a GEV import demand system. The shape of the import demand system comes entirely from the correlation function. For example, the case of independent productivity with an additive correlation function reduces the expression in (12) to the case of a CES import demand (Eaton and Kortum, 2002; Arkolakis et al., 2012), $\pi_\ell(t) = T_\ell W_\ell(t)^{-\theta} / \sum_{\ell'=1}^L T_{\ell'} W_{\ell'}(t)^{-\theta}$. Here, it is immediate to see that the case of no diffusion, which leads to an additive correlation function, implies CES import demand, and hence, very special substitution patterns, as we show next.

Using (12), we can calculate the expenditure elasticity of goods from ℓ to changes in the real wage in ℓ' as

$$\varepsilon_{\ell, \ell'}(t) \equiv \frac{\partial \ln \pi_\ell(t)}{\partial \ln \frac{W_{\ell'}(t)}{P(t)}} = -\theta \left[\mathbf{1}\{\ell = \ell'\} + \frac{T_{\ell'}(t)W_{\ell'}(t)^{-\theta} G_{\ell\ell'}(T_1(t)W_1(t)^{-\theta}, \dots, T_L(t)W_L(t)^{-\theta}; t)}{G_\ell(T_1(t)W_1(t)^{-\theta}, \dots, T_L(t)W_L(t)^{-\theta}; t)} \right], \quad (13)$$

where $G_{\ell\ell'}(x_1, \dots, x_L; t) \equiv \frac{\partial^2}{\partial x_\ell \partial x_{\ell'}} G(x_1, \dots, x_L; t)$. Because $G_{\ell\ell'}$ is negative, the elasticity of substitution for $\ell \neq \ell'$ is positive, as expected from a demand system in the gross-substitute class.

We can clearly see that the curvature of the correlation function, $G_{\ell\ell'}/G_\ell$, generates departures from CES expenditure. When the correlation function is additive—and productivity is independent—the second term inside the brackets in (13) drops out. In this case, the own-price elasticity is $-\theta$ and the cross-price elasticity is 0, reflecting Independence of Irrelevant Alternatives (IIA). Any departure from CES arises from some curvature in the correlation function—that is, from dependence of productivity across production locations.

In turn, correlation in productivity arises when production locations share ideas. How are expenditure elasticities connected to the underlying process of ideas? We can write expenditure shares and expenditure elasticities in terms of the measure of ideas at each time t . The substitution patterns in expenditure between location ℓ and ℓ' are related to how similar applicability of ideas is in those two locations — captured by the shape of the function M . Similar applicability translates into stronger head-to-head competition and higher substitutability.

Proposition 2 (Expenditure shares and substitution elasticities). *Under Assumptions 1, 2, and 3, if M is differentiable, then cross-price elasticities of substitution are*

$$\varepsilon_{\ell,\ell'}(t) = \left(\frac{W_{\ell}(t)}{P(t)} \right)^{-\theta} \frac{1}{\pi_{\ell}(t)} \int_0^{\infty} \frac{W_{\ell'}(t)}{W_{\ell}(t)} a_{\ell'}^{\theta+1} M_{\ell\ell'} \left(\frac{W_1(t)}{W_{\ell}(t)} a_{\ell}, \dots, a_{\ell}, \dots, \frac{W_L(t)}{W_{\ell}(t)} a_{\ell}; t \right) da_{\ell}, \quad (14)$$

for $\ell' \neq \ell$, where $M_{\ell\ell'}(a_1, \dots, a_L; t) \equiv \frac{\partial^2}{\partial a_{\ell} \partial a_{\ell'}} M(a_1, \dots, a_L; t)$, the own-price elasticity is $\varepsilon_{\ell,\ell}(t) = -\theta - \sum_{\ell' \neq \ell} \varepsilon_{\ell,\ell'}(t)$, and expenditure shares can be expressed as

$$\pi_{\ell}(t) = \left(\frac{W_{\ell}(t)}{P(t)} \right)^{-\theta} \int_0^{\infty} a_{\ell}^{\theta} M_{\ell} \left(\frac{W_1(t)}{W_{\ell}(t)} a_{\ell}, \dots, a_{\ell}, \dots, \frac{W_L(t)}{W_{\ell}(t)} a_{\ell} \right) da_{\ell}.$$

Proof. See Appendix B. □

Here, a_{ℓ} represents a level of applicability in ℓ . The corresponding cost that ℓ' needs to compete is $\frac{W_{\ell'}(t)}{W_{\ell}(t)} a_{\ell}$. The integrals in (14) are over a ray from the origin representing applicability levels at which each country is competitive with location ℓ . In turn, $M_{\ell\ell'}$ captures the joint density of applicability between ℓ and ℓ' , conditional on applicability in all other locations. The cross-price elasticity then reflects the likelihood that ℓ and ℓ' have higher applicability than all other locations and that they are head to head competitors. Hence, given wages and expenditure, increases in $\varepsilon_{\ell,\ell'}(t)$ should reflect an increase in the bilateral sharing of competitive ideas.

A related point refers to the evolution over time of the cross-price elasticity: Its dynamics is due to diffusion only. The expression in (4) shows that the contribution of innovation to changes in the measure of ideas M is additively separable across locations. When ideas are not initially shared, all similarity in applicability between locations comes from diffusion. Thus, holding fixed wages and trade shares, changes

in cross-price elasticities over time only reflect the evolution of shared ideas,

$$\frac{\partial}{\partial t} M_{\ell\ell'}(a_1, \dots, a_L; t) = \underbrace{\sum_{\ell^*=1}^L \int_{-\infty}^t \frac{\partial}{\partial t} \frac{\partial^2}{\partial a_{\ell} \partial a_{\ell'}} F(a_1, \dots, a_L \mid \ell^*, t^*; t) \lambda_{\ell^*}(t^*) dt^*}_{\text{Diffusion Effect}}.$$

Again, to make the results concrete, consider the case of ideas that are instantly shared across all location once they diffuse, and assume that once an idea is known in a production location, its applicability is distributed as a unit σ -Fréchet, independent across locations, with $\sigma > \theta$. The measure of ideas at any point in time is given by (10) and the productivity distribution is given by (11). The expenditure share allocated to location ℓ is

$$\pi_{\ell}(t) = \pi_{\ell}^{\text{ND}}(t) + \pi_{\ell}^{\text{D}}(t). \quad (15)$$

The first term is the share of expenditure in goods from ℓ produced with ideas that are unique to that location at time t — that is, those ideas that have not yet diffused elsewhere,

$$\pi_{\ell}^{\text{ND}}(t) = \Gamma(\rho)(1 - \delta_{\ell}(t))\Lambda_{\ell}(t) \frac{W_{\ell}(t)^{-\theta}}{P(t)^{-\theta}}.$$

The second term is the share of expenditure in goods from ℓ produced with ideas that have diffused from a different location at time t ,

$$\pi_{\ell}^{\text{D}}(t) = \Gamma(\rho) \frac{W_{\ell}(t)^{-\frac{\theta}{1-\rho}}}{\sum_{\ell'=1}^L W_{\ell'}(t)^{-\frac{\theta}{1-\rho}}} \sum_{\ell^*=1}^L \delta_{\ell^*}(t) \Lambda_{\ell^*}(t) \left[\sum_{\ell'=1}^L \left(\frac{W_{\ell'}(t)}{P(t)} \right)^{-\frac{\theta}{1-\rho}} \right]^{1-\rho},$$

where $\rho \equiv 1 - \theta/\sigma$. The elasticity of substitution for $\ell = \ell'$ is

$$\varepsilon_{\ell,\ell}(t) = -\theta \frac{\pi_{\ell}^{\text{ND}}(t)}{\pi_{\ell}(t)} - \frac{\theta}{1-\rho} \frac{\pi_{\ell}^{\text{D}}(t)}{\pi_{\ell}(t)},$$

with $\theta/(1-\rho) = \sigma$. For $\ell \neq \ell'$, the elasticity is

$$\varepsilon_{\ell,\ell'}(t) = \frac{\rho\theta}{1-\rho} \frac{\pi_{\ell}^{\text{D}}(t)}{\pi_{\ell}(t)},$$

with $\rho\theta/(1-\rho) = \sigma - \theta$. It is straightforward to see in this example that, holding fixed real wages and trade shares, only diffusion affects the dynamics of the cross elasticity, captured by the term $\pi_{\ell}^{\text{D}}(t)$. It is also immediate to see that if this cross

elasticity is positive, the import demand system is non-CES. Hence, a sufficient condition for the presence of shared ideas across locations is a non-CES import demand system.

The extreme cases that we present next are useful for intuition. With $\delta_\ell(t) = 0$, which means no diffusion, expenditure is CES; only the own elasticity is non-zero and equal to $-\theta$. With $\sigma \rightarrow \theta$, which means that applicability is relatively fat-tailed, productivity is independent, expenditure is also CES, and diffusion cannot be distinguish from innovation: Only the scale parameters of the productivity distribution reflect the process of innovation and diffusion of ideas, and there is no separate information about innovation and diffusion dynamics. With instant diffusion, $\delta_\ell(t) = 1$ and $\sigma > \theta$, only the second term in (15) survives. While the own elasticity is equal to $-\sigma < 0$, the cross elasticity is $\sigma - \theta > 0$. It is easy to see that this cross elasticity increases with σ : More similar applicability across locations increases head-to-head competition, and hence, substitutability. However, in this extreme case, diffusion does not change the elasticity.

Summing up, the substitution elasticities of trade flows at each point in time and over time contain information about the global dynamics of knowledge. These dynamics are both due to the presence of innovation and diffusion of ideas across countries. It is the cross-substitution elasticity alone, however, that reflects the presence of shared knowledge across production locations. Hence, these elasticities are key to separately identify innovation from diffusion of ideas.

5 Conclusion

The trade literature has produced extremely rich estimates of substitution elasticities for international trade flows (see [Broda et al., 2008](#); [Costinot and Rodriguez-Clare, 2014](#); [Feenstra et al., 2017](#); [Bas et al., 2017](#); [Adao et al., 2017](#), among others). In this paper, we show that there is more content to be read from those elasticities when trade flows are connected to technology primitives. To such end, we present a parsimonious model of Ricardian trade that links the dynamics of knowledge to the dynamics of the elasticities of substitution in expenditure. While innovation makes a country technologically distinct, reducing their substitutability with other countries, diffusion between countries generates technological similarity and increases head-to-head competition. In the special case of an innovation-only model where

countries do not share ideas, productivities are independent across space, and the import demand system is CES. As a consequence, non-CES expenditure indicates the presence of shared ideas across countries. Our theoretical result that establishes a mapping between max-stable Fréchet productivity distributions and the structure of innovation and diffusion allows us to directly connect the dynamics of observable trade flows with the dynamics of innovation and knowledge diffusion.

References

- Adao, R., A. Costinot, and D. Donaldson (2017). Nonparametric counterfactual predictions in neoclassical models of international trade. *The American Economic Review* 107(3), 633–689.
- Arkolakis, C., A. Costinot, and A. Rodríguez-clare (2012). New trade models, same old gains? *The American Economic Review*, 94–130.
- 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, 1–21.
- Benhabib, J., J. Perla, and C. Tonetti (2021). Reconciling models of diffusion and innovation: A theory of the productivity distribution and technology frontier. *Econometrica* (forthcoming).
- Bloom, N., T. A. Hassan, A. Kalyani, J. Lerner, and A. Tahoun. The diffusion of disruptive technologies. *NBER Working paper 28999*.
- Broda, C., N. Limao, and D. Weistein (2008). Optimal tariffs and market power: The evidence. *American Economic Review* 98(5), 2032–2065.
- Buera, F. J. and E. Oberfield (2020). The global diffusion of ideas. *Econometrica* Volume88(1).
- Cai, J., N. Li, and A. M. Santacreu (2020). Knowledge diffusion, trade and innovation across countries and sectors. *AEJ: Macroeconomics*, forthcoming.
- Comin, D. and B. Hobijn (2004). Cross-country technology adoption: making the theories face the facts. *Journal of Monetary Economics* 51, 39–83.
- Comin, D. and B. Hobijn (2010). An exploration of technology diffusion. *American Economic Review* 100, 2031–2059.
- Comin, D. and M. Mestieri (2014). Technology diffusion: Measurement, causes, and consequences. In *Handbook of economic growth*, Volume 2, pp. 565–622. Elsevier.

- Costinot, A. and A. Rodríguez-Clare (2014). Trade theory with numbers: Quantifying the consequences of globalization. Technical Report 4.
- Dagsvik, J. K. (1994). Discrete and continuous choice, max-stable processes, and independence from irrelevant attributes. *Econometrica*, 1179–1205.
- De Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability* 12(4), 1194–1204.
- Eaton, J. and S. Kortum (1999). International technology diffusion: Theory and measurement. *International Economic Review* 40(3), 537–570.
- Eaton, J. and S. Kortum (2001). Technology, trade, and growth: A unified framework. *European Economic Review* 45, 742–755.
- Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica*, 1741–1779.
- Feenstra, R. C., P. Luck, M. Obstfeld, and K. N. Russ (2017). In search of the armington elasticity. *The Review of Economics and Statistics*.
- Fosgerau, M., D. McFadden, and M. Bierlaire (2013). Choice probability generating functions. *Journal of Choice Modeling* 8, 1–18.
- Kabluchko, Z. (2009). Spectral representations of sum-and max-stable processes. *Extremes* 12(4), 401.
- Kingman, J. (1992). *Poisson Processes*. Oxford Studies in Probability. Clarendon Press.
- Klenke, A. (2013). *Probability theory: a comprehensive course*. Springer Science & Business Media.
- Kortum, S. (1997). Research, patenting, and technological change. *Econometrica* 65(6).
- Lind, N. and N. Ramondo (2021). Trade with correlation. *Mimeo*.
- McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record* (673).
- McFadden, D. (1981). *Econometric Models of Probabilistic Choice*. MIT Press: Cambridge.
- Parente, S. L. and E. C. Prescott (1994). Barriers to technology adoption and development. *Journal of political Economy* 102(2), 298–321.
- Penrose, M. D. (1992). Semi-min-stable processes. *The Annals of Probability* 20(3), 1450–1463.
- Perla, J. and C. Tonetti (2014). Equilibrium imitation and growth. *Journal of Political Economy* 122, 512–76.

- Perla, J., C. Tonetti, and M. Waugh (2021). Equilibrium technology diffusion, trade, and growth. *American Economic Review* 111(1), 73 – 128.
- Ramondo, N. and A. Rodríguez-Clare (2013). Trade, multinational production, and the gains from openness. *Journal of Political Economy* 121(2), 273–322.
- Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy* 98(5, Part 2), S71–S102.
- Sampson, T. (2016a). Dynamic selection: An idea flows theory of entry, trade and growth. *The Quarterly Journal of Economics* 131(1), 315–380.
- Sampson, T. (2016b). Technology gaps, trade and income. *Unpublished manuscript*.
- Stoev, S. A. and M. S. Taqqu (2005). Extremal stochastic integrals: a parallel between max-stable processes and α -stable processes. *Extremes* 8(4), 237–266.
- Wang, Y. and S. A. Stoev (2010). On the structure and representations of max-stable processes. *Advances in Applied Probability* 42(3), 855–877.

A Proof of Proposition 1

Proof. Sufficiency follows from the marking theorem for Poisson processes (see [Kingman, 1992](#)). In particular, under Assumption 2 and Assumption 3 we can take $\{Q_i(v), t_i^*(v)\}_{i=1,2,\dots}$ as a base Poisson process and take the stochastic process $\{A_{i\ell}(t, v)\}_{\ell=1,\dots,L, t \in \mathbb{R}}$ as a mark of the i 'th point. Then the collection $\{Q_i(v), t_i^*(v), \{A_{i\ell}(t, v)\}_{\ell=1,\dots,L, t \in \mathbb{R}}\}_{i=1,2,\dots}$ is itself a Poisson process. In particular, let J be an integer and fix some $\ell_j \in \{1, \dots, L\}$ and $t_j \in \mathbb{R}$ for each $j = 1, \dots, J$. Then, by the marking theorem for Poisson processes,

$$\begin{aligned}
& \mathbb{E} \sum_{i=1}^{\infty} \mathbf{1}\{Q_i(v) > \underline{q}, t_i^*(v) \leq t, A_{i\ell_j}(t_j, v) \leq a_j \quad \forall j = 1, \dots, J\} \\
&= \int_{\underline{q}}^{\infty} \int_{-\infty}^{\infty} \mathbb{P}[A_{i\ell_j}(t_j, v) \leq a_j \quad \forall j = 1, \dots, J \mid t_i^*(v) = t^*] \theta q^{-\theta-1} dq d\Lambda(t^*) \\
&= \underline{q}^{-\theta} \int_{-\infty}^{\infty} \mathbb{P}[A_{i\ell_j}(t_j, v) \leq a_j \quad \forall j = 1, \dots, J \mid t_i^*(v) = t^*] d\Lambda(t^*).
\end{aligned}$$

Then, under Assumption 1, the finite dimensional distributions of productivity satisfy

$$\begin{aligned}
\mathbb{P}[Z_{\ell_j}(t_j, v) \leq z_j, \forall j = 1, \dots, J] &= \mathbb{P}\left[\max_{i=1,2,\dots} Q_i(v) A_{i\ell_j}(t_j, v) \leq z_{\ell_j}, \forall j = 1, \dots, J\right] \\
&= \mathbb{P}[Q_i(v) A_{i\ell_j}(t_j, v) \leq z_{\ell_j}, \forall j = 1, \dots, J, \forall i = 1, 2, \dots] \\
&= \mathbb{P}\left[Q_i(v) \leq \min_{j=1,\dots,J} \frac{z_{\ell_j}}{A_{i\ell_j}(t_j, v)}, \forall i = 1, 2, \dots\right] \\
&= \mathbb{P}\left[Q_i(v) > \min_{j=1,\dots,J} \frac{z_{\ell_j}}{A_{i\ell_j}(t_j, v)}, \text{ for no } i = 1, 2, \dots\right].
\end{aligned}$$

This last expression is a void probability. Using the previous result for the mean measure,

$$\begin{aligned}
& \mathbb{P} \left[Q_i(v) > \min_{j=1, \dots, J} \frac{z_{\ell_j}}{A_{i\ell_j}(t_j, v)}, \text{ for no } i = 1, 2, \dots \right] \\
&= \exp \left[- \int_{-\infty}^{\infty} \int_{\mathbb{R}_+^J} \int_{\min_{j=1, \dots, J} \frac{z_j}{a_j}}^{\infty} \theta q^{-\theta-1} dq d\mathbb{P} [A_{i\ell_j}(t_j, v) \leq a_j \ \forall j = 1, \dots, J \mid t_i^*(v) = t^*] d\Lambda(t^*) \right] \\
&= \exp \left[- \int_{-\infty}^{\infty} \int_{\mathbb{R}_+^J} \max_{j=1, \dots, J} \left(\frac{a_j}{z_j} \right)^\theta d\mathbb{P} [A_{i\ell_j}(t_j, v) \leq a_j \ \forall j = 1, \dots, J \mid t_i^*(v) = t^*] d\Lambda(t^*) \right].
\end{aligned}$$

Now, let $v_j \geq 0$ for each $j = 1, \dots, J$. The distribution of $\max_{j=1, \dots, J} v_j Z_{\ell_j}(t_j, v)$ is

$$\begin{aligned}
& \mathbb{P} \left[\max_{j=1, \dots, J} v_j Z_{\ell_j}(t_j, v) \leq z \right] \\
& \mathbb{P} [Z_{\ell_j}(t_j, v) \leq z/v_j \ \forall j = 1, \dots, J] \\
&= \exp \left[- \int_{-\infty}^{\infty} \int_{\mathbb{R}_+^J} \max_{j=1, \dots, J} \left(\frac{v_j a_j}{z} \right)^\theta d\mathbb{P} [A_{i\ell_j}(t_j, v) \leq a_j \ \forall j = 1, \dots, J \mid t_i^*(v) = t^*] d\Lambda(t^*) \right] \\
&= \exp \left[- \int_{-\infty}^{\infty} \int_{\mathbb{R}_+^J} \max_{j=1, \dots, J} (v_j a_j)^\theta d\mathbb{P} [A_{i\ell_j}(t_j, v) \leq a_j \ \forall j = 1, \dots, J \mid t_i^*(v) = t^*] d\Lambda(t^*) z^{-\theta} \right].
\end{aligned}$$

Therefore, $\max_{j=1, \dots, J} v_j Z_{\ell_j}(t_j, v)$ is distributed Fréchet and productivity is a max-stable process. Moreover, if we take $J = L$, $\ell_j = j$ and $t_j = t$ for each $j = 1, \dots, L$, we have

$$\begin{aligned}
& \mathbb{P} [Z_\ell(t, v) \leq z_\ell, \forall \ell = 1, \dots, L] \\
&= \exp \left[- \int_{-\infty}^{\infty} \int_{\mathbb{R}_+^L} \max_{\ell=1, \dots, L} \left(\frac{a_\ell}{z_\ell} \right)^\theta d\mathbb{P} [A_{i\ell}(t, v) \leq a_\ell \ \forall \ell = 1, \dots, L \mid t_i^*(v) = t^*] d\Lambda(t^*) \right] \\
&= \exp \left[- \int_{-\infty}^t \int_{\mathbb{R}_+^L} \max_{\ell=1, \dots, L} \left(\frac{a_\ell}{z_\ell} \right)^\theta d\mathbb{P} [A_{i\ell}(t, v) \leq a_\ell \ \forall \ell = 1, \dots, L \mid t_i^*(v) = t^*] d\Lambda(t^*) \right] \\
&= \exp \left[- \int_{\mathbb{R}_+^L} \max_{\ell=1, \dots, L} \left(\frac{a_\ell}{z_\ell} \right)^\theta dM(a_1, \dots, a_L; t) \right],
\end{aligned}$$

where the second line uses the fact that applicability is zero at any time before an idea's discovery time, and the final line uses the definition of M . Therefore, at any moment in time t , the distribution of productivity across production locations is multivariate Fréchet with scale of $T_\ell(t) \equiv \int a_\ell^\theta dM(a_1, \dots, a_L; t)$ and correlation function $G(x_1, \dots, x_L; t) \equiv \int \max_{\ell=1, \dots, N} \frac{a_\ell^\theta}{T_\ell(t)} x_\ell dM(a_1, \dots, a_L; t)$.

It remains to show that productivity is a measurable stochastic process. From Assumption 1, productivity satisfies

$$Z_\ell(t, v) = \max_{i=1, 2, \dots} Q_i(v) A_{i\ell}(t, v).$$

Additionally, $\{A_{i\ell}(t, v) : \Omega \rightarrow \mathbb{R}\}_{(\ell, t) \in \{1, \dots, L\} \times \mathbb{R}}$ is measurable by Assumption 3. Since the maximum of a countable collection of measurable functions is measurable, productivity is a measurable stochastic process.

Necessity follows from Theorem 3.1 and Proposition 4.1 in Wang and Stoev (2010). The later ensures that productivity is separable in probability, which, combined with the former result,

ensures that a minimal spectral representation of productivity exists with respect to a standard Lebesgue space. Also see Theorem 2 in [Kabluchko \(2009\)](#), which states that any max-stable process has a spectral representation defined on a sufficiently rich background measure space.

Let $\{Z_\ell(t, v)\}_{\ell=1, \dots, L, t \in \mathbb{R}}$ be a max-stable process that is independent and identically distributed across $v \in [0, 1]$. Denote the background probability space by $(\Omega, \mathcal{F}, \mathbb{P})$. Further assume that productivity is measurable—for each fixed $\omega \in \Omega$ the map $(\ell, t) \rightarrow Z_\ell(t, v)$ is (Borel) measurable. Then by Theorem 3.1 and Proposition 4.1 in [Wang and Stoev \(2010\)](#) and the equivalence of extremal integral spectral representations and [De Haan \(1984\)](#) spectral representations (see [Stoev and Taqqu, 2005](#)), there exists a $\theta > 0$, a standard Lebesgue space $([0, 1], \mathcal{B}([0, 1]), \mu)$, measurable functions $\{A_\ell(t, s)\}_{\ell=1, \dots, L, t \in \mathbb{R}}$ with $\int_0^1 A_\ell(t, s)^\theta d\mu(s) < \infty$, and a Poisson process $\{Q_i(v), s_i(v)\}_{i=1, 2, \dots}$ for each v with intensity $\theta q^{-\theta-1} dq d\mu(s)$ such that $Z_\ell(t, v) = \max_{i=1, 2, \dots} Q_i(v) A_\ell(t, s_i(v))$. Moreover, the mapping $(\ell, t, s) \rightarrow A_\ell(t, s)$ can be taken to be jointly $\mathcal{B}(\{1, \dots, L\} \times \mathbb{R}) \otimes \mathcal{B}([0, 1])$ -measurable.

Since $s \rightarrow A_\ell(t, s)$ is measurable, we can define a stochastic process $\{A_{i\ell}(t, v)\}_{\ell=1, \dots, L, t \in \mathbb{R}}$ for each i and v such that $A_{i\ell}(t, v) = A_\ell(t, s_i(v))$ for all ℓ and t which is independent of $Q_i(v)$ and independent and identically distribution across i (since $\{Q_i(v), s_i(v)\}_{i=1, 2, \dots}$ is Poisson with intensity $\theta q^{-\theta-1} dq d\mu(s)$). The joint measurability of $(\ell, t, s) \rightarrow A_\ell(t, s)$ then implies that $A_{i\ell}(t, v) : \Omega \rightarrow \mathbb{R}$ is $\mathcal{B}(\{1, \dots, L\} \times \mathbb{R})$ -measurable for each $\omega \in \Omega$. In other words, $\{A_{i\ell}(t, v)\}_{\ell=1, \dots, L, t \in \mathbb{R}}$ is a measurable stochastic process for each $i = 1, 2, \dots$ and $v \in [0, 1]$.

Next, define

$$t_i^*(v) \equiv \min_{\ell=1, \dots, L} \inf\{t \in \mathbb{R} \mid A_{i\ell}(t, v) > 0\}$$

which is a hitting time. Since $\{A_{i\ell}(t, v)\}_{\ell=1, \dots, L, t \in \mathbb{R}}$ is measurable and adapted to its natural filtration, it has a progressively-measurable modification. Taking $\{A_{i\ell}(t, v)\}_{\ell=1, \dots, L, t \in \mathbb{R}}$ as this modification, by the debut theorem, $t_i^*(v)$ is then a stopping time and is therefore a well-defined random variable that is adapted to the natural filtration of $\{A_{i\ell}(t, v)\}_{\ell=1, \dots, L, t \in \mathbb{R}}$. As a result, the function $s \rightarrow \min_{\ell=1, \dots, L} \inf\{t \in \mathbb{R} \mid A_\ell(t, s) > 0\} \equiv \tau(s)$ is measurable. Then by the mapping theorem for Poisson processes (see [Klenke, 2013](#), Theorem 24.16), $\{Q_i(v), t_i^*(v)\}_{i=1, 2, \dots}$ is a Poisson process with intensity $\theta q^{-\theta-1} dq d\Lambda(t)$ where $\Lambda(B) \equiv \mu(\tau^{-1}(B))$ for each $B \in \mathcal{B}(\mathbb{R})$.

It remains to show that $\int_\infty^t \mathbb{E}[A_{i\ell}(t, v)^\theta \mid t_i^*(v) = t^*] d\Lambda(t^*)$ is finite. By Campbell's theorem (see [Kingman, 1992](#)), we have

$$\begin{aligned} \mathbb{E} \sum_{i=1}^{\infty} \mathbf{1}\{Q_i(v) > 1\} A_{i\ell}(t, v)^\theta &= \mathbb{E} \sum_{i=1}^{\infty} \mathbf{1}\{Q_i(v) > 1, t_i^*(v) \leq t\} A_{i\ell}(t, v)^\theta \\ &= \int_\infty^t \mathbb{E}[A_{i\ell}(t, v)^\theta \mid t_i^*(v) = t^*] d\Lambda(t^*). \end{aligned}$$

and we also have

$$\mathbb{E} \sum_{i=1}^{\infty} \mathbf{1}\{Q_i(v) > 1\} A_{i\ell}(t, v)^\theta = \mathbb{E} \sum_{i=1}^{\infty} \mathbf{1}\{Q_i(v) > 1\} A_\ell(t, s_i(v))^\theta = \int_0^1 A_\ell(t, s)^\theta d\mu(s) < \infty.$$

Together, these results imply that $\int_\infty^t \mathbb{E}[A_{i\ell}(t, v)^\theta \mid t_i^*(v) = t^*] d\Lambda(t^*) < \infty$.

□

B Proof of Proposition 2

Proof. Note that

$$\begin{aligned}
& T_\ell(t)W(t)^{-\theta}G_\ell(T_1(t)W_1(t)^{-\theta}, \dots, T_L(t)W_L(t)^{-\theta}; t) \\
&= \int \mathbf{1} \left\{ \frac{W_\ell(t)}{a_\ell} \leq \frac{W_l(t)}{a_l} \quad \forall l \neq \ell \right\} \left(\frac{W_\ell(t)}{a_\ell} \right)^{-\theta} dM(a_1, \dots, a_L; t) \\
&= \int \mathbf{1} \left\{ a_l \leq \frac{W_l(t)}{W_\ell(t)} a_\ell \quad \forall l \neq \ell \right\} \left(\frac{W_\ell(t)}{a_\ell} \right)^{-\theta} dM(a_1, \dots, a_L; t) \\
&= \int_0^\infty \left(\frac{W_\ell(t)}{a_\ell} \right)^{-\theta} M \left(\frac{W_1(t)}{W_\ell(t)} a_\ell, \dots, a_\ell, \dots, \frac{W_L(t)}{W_\ell(t)} a_\ell \right) da_\ell \\
&= \int_0^\infty \left(\frac{W_\ell(t)}{a_\ell} \right)^{-\theta} M_\ell \left(\frac{W_1(t)}{W_\ell(t)} a_\ell, \dots, a_\ell, \dots, \frac{W_L(t)}{W_\ell(t)} a_\ell \right) da_\ell.
\end{aligned}$$

Also,

$$\begin{aligned}
P(t)^{-\theta} &= G(T_1(t)W_1(t)^{-\theta}, \dots, T_L(t)W_L(t)^{-\theta}; t) \\
&= \sum_{\ell=1}^L T_\ell(t)W_\ell(t)^{-\theta} G_\ell(T_1(t)W_1(t)^{-\theta}, \dots, T_L(t)W_L(t)^{-\theta}; t) \\
&= \sum_{\ell=1}^L \int_0^\infty \left(\frac{W_\ell(t)}{a_\ell} \right)^{-\theta} M_\ell \left(\frac{W_1(t)}{W_\ell(t)} a_\ell, \dots, a_\ell, \dots, \frac{W_L(t)}{W_\ell(t)} a_\ell \right) da_\ell.
\end{aligned}$$

As a result we have

$$\pi_\ell(t) = \frac{\int_0^\infty \left(\frac{W_\ell(t)}{a_\ell} \right)^{-\theta} M_\ell \left(\frac{W_1(t)}{W_\ell(t)} a_\ell, \dots, a_\ell, \dots, \frac{W_L(t)}{W_\ell(t)} a_\ell \right) da_\ell}{\sum_{\ell'=1}^L \int_0^\infty \left(\frac{W_{\ell'}(t)}{a_{\ell'}} \right)^{-\theta} M_{\ell'} \left(\frac{W_1(t)}{W_{\ell'}(t)} a_{\ell'}, \dots, a_{\ell'}, \dots, \frac{W_L(t)}{W_{\ell'}(t)} a_{\ell'} \right) da_{\ell'}}.$$

Then, for $\ell = \ell'$,

$$\begin{aligned}
\frac{\partial \pi_\ell(t)}{\partial \ln \frac{W_\ell(t)}{P(t)}} &= -\theta \sum_{\ell=1}^L \int_0^\infty \left(\frac{W_\ell(t)/P(t)}{a_\ell} \right)^{-\theta} M_\ell \left(\frac{W_1(t)}{W_\ell(t)} a_\ell, \dots, a_\ell, \dots, \frac{W_L(t)}{W_\ell(t)} a_\ell \right) da_\ell \\
&\quad - \sum_{\ell' \neq \ell} \int_0^\infty \left(\frac{W_\ell(t)/P(t)}{a_\ell} \right)^{-\theta} \frac{W_{\ell'}(t)}{W_\ell(t)} a_\ell M_{\ell\ell'} \left(\frac{W_1(t)}{W_\ell(t)} a_\ell, \dots, a_\ell, \dots, \frac{W_L(t)}{W_\ell(t)} a_\ell \right) da_\ell \\
&= -\theta \pi_\ell(t) - \sum_{\ell' \neq \ell} \int_0^\infty \left(\frac{W_\ell(t)/P(t)}{a_\ell} \right)^{-\theta-1} \frac{W_{\ell'}(t)}{P(t)} M_{\ell\ell'} \left(\frac{W_1(t)}{W_\ell(t)} a_\ell, \dots, a_\ell, \dots, \frac{W_L(t)}{W_\ell(t)} a_\ell \right) da_\ell,
\end{aligned}$$

and for $\ell' \neq \ell$,

$$\begin{aligned} \frac{\partial \pi_\ell(t)}{\partial \ln \frac{W_{\ell'}(t)}{P(t)}} &= \int_0^\infty \left(\frac{W_\ell(t)/P(t)}{a_\ell} \right)^{-\theta} \frac{W_{\ell'}(t)}{W_\ell(t)} a_\ell M_{\ell\ell'} \left(\frac{W_1(t)}{W_\ell(t)} a_\ell, \dots, a_\ell, \dots, \frac{W_L(t)}{W_\ell(t)} a_\ell \right) da_\ell \\ &= \int_0^\infty \left(\frac{W_\ell(t)/P(t)}{a_\ell} \right)^{-\theta-1} \frac{W_{\ell'}(t)}{P(t)} M_{\ell\ell'} \left(\frac{W_1(t)}{W_\ell(t)} a_\ell, \dots, a_\ell, \dots, \frac{W_L(t)}{W_\ell(t)} a_\ell \right) da_\ell. \end{aligned}$$

These semi-elasticities can be re-expressed as elasticities by dividing by $\pi_\ell(t)$.

□