Global Innovation and Knowledge Diffusion*

Nelson Lind†
Emory University

Natalia Ramondo‡
Boston University, NBER, and CEPR

December 16, 2022

Abstract

We develop a Ricardian model of trade where countries innovate ideas that diffuse globally. Our key result provides necessary and sufficient conditions for innovation and diffusion to generate max-stable Fréchet productivity, linking generalized extreme value expenditure to knowledge flows. Innovation makes a country technologically distinct, reducing their substitutability with other countries. In contrast, diffusion generates technological similarity, increasing head-to-head competition and substitutability. In an innovation-only model where countries do not share ideas, productivities are independent across countries and expenditure is CES. Consequently, departures from CES reveal diffusion patterns.

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

*We thank our discussant Chris Tonetti as well as Jess Benhabib, Simon Fuchs, In-Koo Cho, Hugo Hopenhayn, Jesse Perla, Thomas Sampson, Ana Maria Santacreu, Mike Waugh, and seminar participants at several conferences and institutions for their helpful comments. This research is supported by NSF grant #1919372. All errors are our own.
†E-mail: nlind@emory.edu.
‡E-mail: nramondo@bu.edu.

1 Introduction

A central feature of growth is that new ideas improve production technologies and are inherently non-rival, as Romer (1990) emphasized. While sharing ideas can lead to similarity in production methods, not all people, firms, or 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), creating dissimilarities in productivity. The dynamics of innovation and knowledge diffusion can then shape similarities and differences in technology across countries. However, estimating the dynamics of knowledge remains a challenge — in particular, disentangling the effects coming from innovating new ideas versus adopting existing ideas.[1]

We present a Ricardian model of trade and knowledge flows where productivity is the result of locations creating and adopting the best ideas available to them. Our main result shows that this structure for innovation and diffusion is *necessary and sufficient* to obtain a max-stable Fréchet productivity distribution with correlation over space and time.[2] The creation and spread of ideas changes the worldwide productivity distribution, shapes comparative advantages, and determines expenditure patterns. Specifically, countries with similar technologies are closer competitors, and hence, have more elastic expenditure with each other.

Our results formalize the idea that expenditure patterns reflect the footprints of knowledge. In particular, cross-price elasticities indicate shared knowledge. This link between expenditure and unobservable knowledge flows circumvents the need to use patent data, or direct measures of technology adoption, to estimate models of knowledge creation and diffusion.

We begin by assuming that ideas are specific to goods and their efficiency has a time-invariant global component, quality, and a time-varying location-specific component, applicability. Ideas of different quality are discovered over time by individual locations according to a Poisson process. Conditional on being discovered, other locations learn about the idea through time. While any location can use

[1] See Comin and Hobijn (2004), Comin and Hobijn (2010), and Bloom et al. (2021) for efforts to measure technology adoption and knowledge diffusion. For a review see Comin and Mestieri (2014).

[2] A distribution is max-stable when preserved by the max operator. This property ensures closed-form results in models of head-to-head competition — for example, quantitative trade models inspired by Eaton and Kortum (2002), and other models such as Bernard et al. (2003), Bryan and Morten (2019), and Hsieh et al. (2019).

an idea, locations may differ in their ability to apply the idea. Over time, the applicability of each idea across locations may change, which, together with the creation of new ideas, determines the dynamics of knowledge.

The concept of applicability is key for distinguishing between innovation and diffusion, introducing time-varying technological similarity (or lack-of) across locations, and accommodating many models of diffusion. We only need applicability to be independent of quality — with small differences across countries relative to differences in quality across ideas. Otherwise, applicability is unrestricted. This feature allows us to depart from the previous literature (e.g. Eaton and Kortum, 1999; Cai et al., 2020; Benhabib et al., 2021) by simultaneously abandoning a closed economy, independent Fréchet, and constant-elasticity-of-substitution (CES) expenditure functions.

This knowledge structure is necessary and sufficient for max-stable Fréchet productivity due to the spectral representation theorem for max-stable processes, which connects max-stable processes to Poisson processes.[3] Max-stable distributions lead to closed-form solutions for expenditure shares, and generate the class of generalized extreme value (GEV) demand systems (McFadden, 1978). This class features rich substitution elasticities, departs from independence of irrelevant alternatives (IIA), and approximates any demand system satisfying gross substitutes (Fosgerau et al., 2013). The necessity of our structure means that there always exists a model of innovation and diffusion — that is, some specification for applicability over space and time — that rationalizes GEV expenditure.

Intuitively, innovation makes a country technologically distinct, while diffusion generates technological similarity, reflected in high correlation in productivity. In turn, high correlation in productivity increases head-to-head competition and leads to high substitutability in expenditure. In the special case of an innovation-only model where countries do not share ideas, there can be no similarity in technology, productivity is independent across countries, expenditure is CES, and IIA holds. It follows that departures from CES expenditure reveal diffusion patterns.

Overall, our result linking max-stable Fréchet productivity distributions to the underlying structure of innovation and diffusion of ideas provides not only a framework to infer knowledge from data on expenditure, but also provides a tool to develop

[3] De Haan (1984) introduces the spectral representation — using a Poisson process to generate continuous max-stable processes. Stoev and Taqqu (2005) extend to separable max-stable processes, Kabluchko (2009) extends to arbitrary index sets, and Wang and Stoev (2010) provide classification results. Dagsvik (1994) applies the spectral representation theorem to decision theory.

new models where head-to-head competition across locations, workers, or firms, is key.

Related literature. This paper builds on the literature that generates Fréchet productivity from Poisson processes. Kortum (1997) introduces the basic idea and Eaton and Kortum (2001) uses it in the context of a trade model. If the production technology is determined by the best idea, and if ideas become available according to a Poisson process, after a sufficiently long period of time, productivity can be approximated by an extreme value distribution. We depart from this approach by leveraging the spectral representation of max-stable processes. In this way, we are able to: introduce both innovation and diffusion of ideas; generalize from the case of independent Fréchet, while preserving the max-stability property key to tractability in models of head-to-head competition; and provide exact rather than asymptotic results.

The literature following Kortum (1997) and Eaton and Kortum (2001), such as Buera and Oberfield (2020) and Cai et al. (2020), has been restricted to the case of independent Fréchet, and hence, to CES expenditure. As a consequence, the resulting models struggle to capture how the offsetting forces of innovation and diffusion shape correlation in productivity. In those models, diffusion is modeled as (independently) sampling ideas across — rather than within — goods, so that diffusion does not increase head-to-head competition among suppliers. A central 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 non-zero cross-price elasticities and more substitution in expenditure.

Our treatment of diffusion is closer to Eaton and Kortum (1999). In fact, their innovation and diffusion structure fits into our framework because ideas are specific to goods, have a common quality component across countries, and diffuse with a lag to other countries. Their model, however, does not include trade flows, and hence, the pattern of comparative advantage across countries is not relevant. Using our results, one can introduce Ricardian trade in their model and bring back comparative advantage forces.

Papers such as Sampson (2016b) and Perla et al. (2021) introduce models where (endogenous) innovation, diffusion, and trade interact to generate growth.[4] While

[4]Using 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.

our model features exogenous innovation and diffusion, we depart from their modeling approach and use tools from the literature on max-stable processes together with Ricardian trade. This allows us to generate expenditure functions that are not restricted to CES. Similar to previous work by the same authors (Perla and Tonetti, 2014; Sampson, 2016a), we model diffusion as a process where not only the ideas currently used in production are evaluated for adoption, but any idea is. In this way, countries can develop better applications of an idea created elsewhere but not yet used anywhere.

Finally, our approach modeling of innovation and diffusion has similar consequences for productivity as Benhabib et al. (2021) where, loosely speaking, "[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 these distributions to Ricardian trade.

2 Set Up

Time is continuous and indexed by $t \in \mathbb{R}$. The world economy has a finite number of locations, $\ell \in \mathbb{L} \equiv \{1, \ldots, L\}$. Each location is populated by a measure of individuals, who supply labor, consume a continuum of tradable goods, $v \in [0, 1]$, and have CES preferences with elasticity of substitution $\eta > 0$.

Locations produce goods using an only-labor constant-returns-to-scale technology,

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

where $Z_\ell(t, v)$ is location-time-good specific productivity. We focus on the case of frictionless trade, simplifying to a common market for goods without changing any of our results.

Following Eaton and Kortum (2002) (EK), we model productivity as a random draw across locations and goods. Over time, productivity is a stochastic process, $\{Z_\ell(t, v)\}_{(\ell,t) \in \mathbb{L} \times \mathbb{R}}$.

In what follows, we first develop a model of innovation and diffusion where productivity is the result of locations creating ideas and adopting the best ones available to them. Next, we show that this structure, combined with a Poisson assumption on innovation and an independence assumption on diffusion, is necessary and sufficient for max-stable Fréchet productivity.

2.1 Ideas and productivity

For each good, there exists an infinite but countable set of ideas, $i = 1, 2, \ldots$, containing all ideas that will be discovered. Each idea represents a physical production technique (e.g., a blueprint) and may be applied in different locations. The productivity of idea i at location ℓ at time t is $Q_i(v)A_{i\ell}(t,v)$. The *quality* of the idea, $Q_i(v)$, represents the idea's overall efficiency and is common to all locations. This first term generates heterogeneity in productivity across ideas. The *applicability* of idea i in location ℓ at time t, $A_{i\ell}(t,v)$, is a location and time specific component, which generates heterogeneity in productivity across locations for each idea. This component captures costs of technology adoption, which may differ across both locations and time, and adjustments needed to implement an idea in a different location from the one that innovated it.

The applicability of an idea is the key concept in our model. It allows us to cleanly distinguish between innovation and diffusion, and to introduce time-varying technological similarity (or lack-of) and time-varying correlation (or lack of) in productivity across locations.

Ideas can be used anywhere, but 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 is zero, $A_{i\ell}(t,v) = 0$. Among locations with knowledge of an idea, $A_{i\ell}(t,v) > 0$, technological similarity arises from similarity in applicability.

Finally, each location adopts the best idea available to them for each good.

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

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

If a new idea becomes available to location ℓ at time t (i.e. applicability goes from zero to a positive number), it gets adopted only if improves on the most efficient already available idea. The overall efficiency of an idea can be high because its quality is high, or because applicability in that location 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 by drawing a better application.

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

Innovation occurs when an idea first becomes known to some location. Formally, the *discovery time* time of idea i is $t_i^*(v) \equiv \inf\{t \in \mathbb{R} \mid \exists \ell \text{ s.t. } A_{i\ell}(t,v) > 0\}$.

We next put stochastic structure on innovation.

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

While $\Lambda(\mathrm{d}t^*)$ controls the overall arrival rate of ideas, θ controls the arrival rate of low versus high quality ideas. Assumption 2 implies that the expected number of ideas with $Q_i(v) > \underline{q}$ discovered up to time t is $\underline{q}^{-\theta}\Lambda(t)$ where $\Lambda(t) \equiv \int_{-\infty}^{t} \Lambda(\mathrm{d}t^*)$.[5] Among these ideas, quality is independent of time and distributed Pareto with shape θ and scale $\underline{q}$.[6] A lower θ means a fatter tail so that locations sometimes innovate very good ideas.

As we show in Theorem 1, this particular functional form for the intensity is *necessary* to obtain max-stable Fréchet productivity. Specifically, we need the time-varying arrival rate to be multiplicatively separable from the time-invariant quality component of the intensity — ensuring that quality is independent and identically distributed over time — and we need quality to have a Pareto tail — ensuring that the marginal distributions of productivity are Fréchet. Without either of these restrictions, we break Fréchet. All the growth models that generate Fréchet productivity from Poisson processes, in the tradition of Kortum (1997), satisfy this assumption.[7]

[5] $\int_{\underline{q}}^{\infty} \int_{-\infty}^{t} \theta q^{-\theta-1} \Lambda(\mathrm{d}t^*) \mathrm{d}q = \underline{q}^{-\theta}\Lambda(t)$.

[6] $\mathbb{P}[Q_i(v) \leq q \mid Q_i(v) > \underline{q}, t_i^*(v) \leq t] = \frac{\underline{q}^{-\theta}\Lambda(t) - q^{-\theta}\Lambda(t)}{\underline{q}^{-\theta}\Lambda(t)} = 1 - \left(q/\underline{q}\right)^{-\theta}$.

[7] Specifically, Kortum (1997) is the case where $L = 1$ and $A_{i\ell}(t,v) = 1$. A large literature uses non-Poisson stochastic processes to model innovation and diffusion in closed economies (e.g. the early work by Jovanovic and Rob, 1989; Jovanovic and Nyarko, 1996).

An intuitive way of thinking about the quality distribution is that qualities are drawn at the beginning of time and lie "dormant" until discovery. Assumption 2 means that there are many more low-quality ideas waiting to be discovered, so that, most likely, a location will discover a low-quality idea. However, occasionally they get a really good idea.

Conditional on an idea's discovery, diffusion is any subsequent change in applicability in locations other than the discovery location.[8] Hence, to characterize diffusion, we keep track of an idea's applicability over time and across locations. We only require the following assumption.

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

Apart from regularity conditions — that ideas are i.i.d. and applicabilities are measurable and on-average finite — Assumption 3 means that improvements in applicability are just as likely to occur for low as for high quality ideas. Although this restriction means that locations cannot target high quality ideas for diffusion, it does not limit how the applicability of an idea in one location at one moment in time relates to the applicability of the same idea in a different location at another moment in time — that is, the structure of applicability over space and time.

In line with the separability of time-varying components in Assumption 2, this independence assumption is *necessary* to obtain max-stable Fréchet productivity in Theorem 1. However, it is less restrictive than it appears. For instance, different types of ideas (e.g. young vs old, high vs low costs of adoption) can have different diffusion dynamics, but diffusion cannot be higher among ideas of a particular quality within each idea type. That is, applicability can be targeted, but not quality.[9] Ultimately, which ideas diffuse is random within each idea type. This randomness maintains the quality Pareto-tail and ensures that marginal productivity distributions are Fréchet. In this way, applicability captures different diffusion models, as we show below.

An important implication of the independence between quality and applicability is

[8] One can interpret changes in applicability within a location as better applications of the idea, in the spirit of Lucas and Moll (2014).

[9] For example, search-based endogenous diffusion provides a possible micro-foundation if countries can only direct search towards types of ideas, and cannot directly target high quality ideas.

that we only need to condition the applicability distribution on the idea's discovery time in order to keep track of how the idea diffuses. As a consequence, the state variable for the world economy at time t is given by the measure of ideas with $Q_i(v) > 1$ discovered up to t:

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

For the rest of the paper, we characterize outcomes in terms of this variable. Specific cases of innovation and diffusion processes place further restrictions on M, as in the examples below.

Our structure departs from the previous literature (Kortum, 1997) in that locations do not draw the idea's (Pareto) quality, conditional on the idea's (Poisson) arrival. Rather, ideas of different qualities arrive at different rates as indicated by Assumption 2. Then, variation in the applicability of an idea controls similarities and differences in productivity across locations and times. Rather than using extreme value theory, which leads asymptotically to independent Fréchet distributions, our structure for technology leads to exact results and to max-stable Fréchet productivity with arbitrary correlation, as we formalize in Theorem 1.

We have left out the underpinnings of innovation. The intensity of innovation, $\Lambda(dt)$, is key in growth models, and may depend, for instance, on the number of researchers in a country, or the strength of the patent system. The goal of those models is to explain the determinants of growth and, in open economies, how trade liberalization affects economic growth.

We have also put minimal structure on diffusion, but further assumptions on applicability would give rise to different models of diffusion. Again, the goal of many diffusion models is to study the determinants of technology gaps across countries, the speed of convergence to the technology frontier, and the role of trade in closing those gaps. Our goal here is somewhat different: to provide tools that allow for a tractable characterization of the global evolution of knowledge and its consequences for the world production possibility frontier.

3 Productivity as Max-Stable Fréchet

We now provide a closed-form characterization for the global distribution of productivity and show that the structure for innovation and diffusion in Section 2 generates a max-stable Fréchet productivity distribution.[10]

Theorem 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,*

$$\mathbb{P}\left[Z_1(t,v) \leq z_1, \ldots, Z_L(t,v) \leq z_L\right] = \exp\left[-\int \max_{\ell \in \mathbb{L}} a_\ell^\theta z_\ell^{-\theta} dM(a_1, \ldots, a_L; t)\right]. \quad (3)$$

Proof. Necessity follows from the spectral representation of max-stable processes in Wang and Stoev (2010), which ensures that a Poisson process with the properties in Assumptions 2 and 3 exists such that productivity arises from maximizing over the points of the process (Assumption 1). The proof of sufficiency is constructive and uses properties of Poisson processes. We sketch the proof for a point in time, t. First, because each location adopts the best idea available to them (Assumption 1), the joint distribution of productivity is

$$\mathbb{P}\left[Z_1(v) \leq z_1, \ldots, Z_L(v) \leq z_L\right] = \mathbb{P}\left[\max_{i=1,2,\ldots \text{ s.t. } t_i^*(v) \leq t} Q_i(v) A_{i\ell}(t,v) \leq z_\ell \;\; \forall \ell \in \mathbb{L}\right],$$

which can be expressed as a void probability,

$$\mathbb{P}\left[Z_1(v) \leq z_1, \ldots, Z_L(v) \leq z_L\right] = \mathbb{P}\left[Q_i(v) > \min_{\ell \in \mathbb{L}} \frac{z_\ell}{A_{i\ell}(t,v)} \text{ for no } i \text{ s.t. } t_i^*(v) \leq t\right]. \quad (4)$$

We exploit properties of Poisson processes to calculate (4). Because $\{Q_i(v), t_i^*(v)\}_{i=1,2,\ldots}$ forms a Poisson process (Assumption 2) and applicabilities are conditionally independent (Assumption 3), $\{Q_i(v), t_i^*(v), \{A_{i\ell}(t,v)\}_{(\ell,t) \in \mathbb{L} \times \mathbb{R}}\}_{i=1,2,\ldots}$ is a marked Poisson process where $\{A_{i\ell}(t,v)\}_{(\ell,t) \in \mathbb{L} \times \mathbb{R}}$ is the mark of the i'th point. In turn, the subset of known ideas is a (thinned) Poisson process. We use its mean measure, $\underline{q}^{-\theta} M(a_1, \ldots, a_L; t)$, to solve for (4) by calculating the expected number of known ideas with $Q_i(v)$ above $\min_{\ell \in \mathbb{L}} \frac{z_\ell}{A_{i\ell}(t,v)}$ to get (3). See Appendix A for details.

[10] Productivity is a *max stable process* on $\mathbb{L} \times \mathbb{R}$ with Fréchet marginals if for any $J \in \mathbb{N}$ and $v_j \geq 0$, $(\ell_j, t_j) \in \mathbb{L} \times \mathbb{R}$ for $j = 1, \ldots, J$, $\max_{j=1,\ldots,J} v_j Z_{\ell_j}(t_j, v)$ is Fréchet (see Stoev and Taqqu, 2005).

$\square$

Changes in the distribution of productivity over time arise from changes in the measure of ideas: the Poisson arrival of ideas captures innovation, while the evolution of an idea's applicability conditional on its discovery captures diffusion. As the theorem formalizes, the resulting productivity distribution at each point in time is max-stable Fréchet, and can be re-written as

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

where the shape is θ, scales are $T_\ell(t) \equiv \int a_\ell^\theta \mathrm{d}M(a_1, \ldots, a_L; t)$, and the correlation function, as in Lind and Ramondo (2023), is

$$G(x_1, \ldots, x_L; t) \equiv \int \max_{\ell \in \mathbb{L}} \frac{a_\ell^\theta}{T_\ell(t)} x_\ell \mathrm{d}M(a_1, \ldots, a_L; t). \quad (6)$$

Theorem 1 makes clear that: the shape parameter θ, which regulates heterogeneity in productivity over the continuum of goods, is the same Pareto-tail parameter that regulates heterogeneity in ideas' quality — when quality is more fat-tailed (lower θ), productivity is more disperse across goods within each ℓ; the scale parameter reflects the average applicability of ideas in location ℓ; and the correlation function, G, which measures dependence in productivity across locations, reflects cross-location similarities in applicability.

We next turn to examples to illustrate the interactions between innovation and diffusion that generate correlation (or lack-of) in productivity.

3.1 From knowledge to productivity: examples

We develop three examples to illustrate the close link between correlation in productivity and diffusion. The first example shows that ideas must be shared across countries for correlation to arise. The second example shows that applicabilities must be relatively similar across countries to obtain correlation in productivity, and that the extent of diffusion directly determines the strength of this correlation. The third example shows how to create correlation in productivity by modeling diffusion in a way that nests Eaton and Kortum (1999) as a limiting case. To construct each example, we condition applicability on additional idea-level variables with clear economic interpretation, and then apply Theorem 1.

In all three examples, we make the standard assumption that ideas start in one place and then spread. In our model, this means that each idea has a unique discovery location, $\ell_i^*(v)$, and arrives to other locations with a lag. As a consequence, no idea is initially shared and all shared knowledge comes from diffusion. To add this assumption, we condition the distribution of applicability on an idea's discovery location, which makes (2) additively separable.

$$M(a_1, \ldots, a_L; t) = \sum_{\ell^*=1}^{L} \int_{-\infty}^{t} \mathbb{P}[A_{i\ell}(t, v) \leq a_\ell \;\; \forall \ell \mid \ell_i^*(v) = \ell^*, t_i^*(v) = t^*] \lambda_{\ell^*}(t^*) \mathrm{d}t^* \quad (7)$$

Here, $\lambda_{\ell^*}(t^*) \equiv \mathbb{P}[\ell_i^*(v) = \ell^* \mid t_i^*(v) = t^*] \partial \Lambda(t^*)/\partial t^*$ is the rate at which ideas with quality above one are discovered in ℓ^*. Conditioning on an idea-level variable, such as the idea's discovery location, introduces additive separability to the measure of ideas — a feature that will have consequences for expenditure patterns.

Our first example assumes that locations never share the ideas they innovate. The measure of ideas in (7) becomes

$$M(a_1, \ldots, a_L; t) = \sum_{\ell^*=1}^{L} \int_{-\infty}^{t} \mathbb{P}[A_{i\ell^*}(t, v) \leq a_{\ell^*} \mid \ell_i^*(v) = \ell^*, t_i^*(v) = t^*] \lambda_{\ell^*}(t^*) \mathrm{d}t^*. \quad (8)$$

In this case, each ideas' applicability is degenerate at zero in all locations except for the discovery location at all times. Using Theorem 1, productivity is distributed

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

with scales given by the average applicability of ideas previously discovered in the location, and an additive correlation function (see Appendix C for derivations). The lack of shared ideas across locations, together with the Poisson arrival of ideas, leads to independent productivity across locations.

This result clarifies that independent productivity arises whenever ideas are unique to their discovery location. It is the reason for independence in Eaton and Kortum (2001) — an innovation-only model where ideas arrive Poisson and never leave their discovery location — and in Buera and Oberfield (2020) — where diffusion generates new ideas.[11] Correlated productivity requires shared ideas.

[11] In their model, ideas arrive Poisson as in Assumption 2. Given idea quality, q, producers sample from the local productivity distribution (across goods). If adopted, productivity is qz^β

Our second example assumes that ideas are either non-diffused (exclusive to the discovery location) or diffused (known everywhere). This all-or-northing form of diffusion introduces an additional idea-level variable, $D_i(t, v)$, indicating if an idea has diffused by time t. Once an idea is known in a location, applicability is independent across locations and unit-Fréchet with shape $\sigma > \theta$. The measure of ideas reflects non-diffused and diffused ideas,

$$M(a_1, \ldots, a_L; t) = \sum_{\ell^*=1}^{L} \left[(1 - \delta_{\ell^*}(t)) e^{-a_{\ell^*}^{-\sigma}} + \delta_{\ell^*}(t) e^{-\sum_{\ell=1}^{L} a_\ell^{-\sigma}} \right] \Lambda_{\ell^*}(t), \tag{10}$$

where $\Lambda_{\ell^*}(t) \equiv \int_{-\infty}^{t} \lambda_{\ell^*}(t^*) \mathrm{d}t^*$ and $\delta_{\ell^*}(t) \equiv \int_{-\infty}^{t} \mathbb{P}[D_i(t, v) = 1 \mid \ell_i^*(v) = \ell^*, t_i^*(v) = t^*] \frac{\lambda_{\ell^*}(t^*)}{\Lambda_{\ell^*}(t)} \mathrm{d}t^*$ is the fraction of ideas from ℓ^* that have diffused to the rest of the world. Using Theorem 1, the productivity distribution is

$$\mathbb{P}[Z_1(t) \leq z_1, \ldots, Z_L(t) \leq z_L] = \exp\left[-\sum_{\ell=1}^{L} T_\ell^{ND}(t) z_\ell^{-\theta} + T^D(t) \left(\sum_{\ell=1}^{L} z_\ell^{-\frac{\theta}{1-\rho}} \right)^{1-\rho} \right], \tag{11}$$

where $\rho \equiv 1 - \theta/\sigma \in [0, 1)$ is the correlation parameter, $T_\ell^{ND}(t) \equiv \Gamma(1 - \theta/\sigma)(1 - \delta_\ell(t))\Lambda_\ell(t)$ is the stock of non-diffused ideas from ℓ, and $T^D(t) \equiv \Gamma(1-\theta/\sigma) \sum_{\ell=1}^{L} \delta_\ell(t)\Lambda_\ell(t)$ is the global stock of diffused ideas (see Appendix C for derivations).

This case combines an additive correlation function, which implies independence, with a CES correlation function, which implies symmetric correlation. The extent of correlation depends on the fraction of ideas that have diffused. If there is no diffusion, $\delta_\ell(t) = 0$, (11) reduces to (9) where ideas are never shared, and productivity is independent Fréchet. If all ideas diffuse immediately after being innovated, $\delta_\ell(t) \to 1$, $T_\ell^{ND}(t) = 0$, and the productivity distribution exhibits correlation parameterized by ρ.[12] The amount of correlation reflects heterogeneity in applicability across locations (σ) relative to the heterogeneity in quality across ideas (θ). As $\sigma \to \theta$, applicability is as fat-tailed as idea quality, $\rho \to 0$ and productivity is independent

where z is the sampled productivity and $0 \leq \beta < 1$. From the perspective of our model, applicability in the location creating the idea is z^β, and since this idea is specific to the location, its applicability is zero everywhere else. Effectively, this form of diffusion generates ideas that are never shared.

[12] This case coincides with multinational production in Ramondo and Rodríguez-Clare (2013): Technologies are correlated across production locations (ρ), but are independent across discovery locations. Cai and Xiang (2022) obtain this distribution by extending Buera and Oberfield (2020) so that the original efficiency of an idea is a random vector correlated across countries. In our framework, this setup is equivalent to assuming a particular distribution for applicability.

across locations, despite the existence of diffusion. In contrast, as $\sigma \to \infty$, there is no dispersion in applicability across locations, $\rho \to 1$ and diffusion equalizes productivity everywhere.

This example illustrates that not only the extent of diffusion determines the degree of correlation in productivity, but also that differences in the applicability of the same idea across locations need to be small relative to differences in quality across ideas. If applicability were as dispersed as 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 across locations due applicability.

As our final example, we apply the result in Theorem 1 to a structure for diffusion that includes Eaton and Kortum (1999) as a limiting case. Let $\mathcal{L}_i(t,v) \equiv \{\ell \in \mathbb{L} \mid A_{il}(t,v) > 0\}$, and assume that across locations with knowledge of an idea applicability is independent and distributed unit-Fréchet with shape $\sigma > \theta$. The measure of ideas is

$$M(a_1,\ldots,a_L;t) = \sum_{\ell^*=1}^{L} \sum_{\mathcal{L}\subset\mathbb{L}} e^{-\sum_{\ell\in\mathcal{L}} a_\ell^{-\sigma}} \delta_{\ell^*}(\mathcal{L},t)\Lambda_{\ell^*}(t),$$

where $\delta_{\ell^*}(\mathcal{L},t) \equiv \int_{-\infty}^{t} \mathbb{P}[\mathcal{L}_i(t,v) = \mathcal{L} \mid \ell_i^*(v) = \ell^*, t_i^*(v) = t^*] \frac{\lambda_{\ell^*}(t^*)}{\Lambda_{\ell^*}(t)} \mathrm{d}t^*$ is the fraction of ideas discovered by ℓ^* commonly known to locations in $\mathcal{L}$ at time t. In contrast to the previous example, this expression reflects the possibility that each discovery location ℓ^* may have ideas that already diffused to some locations but not others. This example nests the all-or-nothing diffusion model as the special case where $\delta_{\ell^*}(\mathcal{L},t)$ is non-zero only if $\mathcal{L} = \{\ell^*\}$ or $\mathcal{L} = \mathbb{L}$. In the limiting case with $\sigma \to \infty$, so that applicabilities are identical and degenerate at one, and under the additional restriction that ideas diffuse exponentially over time, we get the model in Eaton and Kortum (1999). With $\sigma < \infty$, applicability introduces idiosyncratic differences in how countries can use the same idea.

As before, we apply Theorem 1 to get the productivity distribution,

$$\mathbb{P}[Z_1(t) \le z_1,\ldots,Z_L(t) \le z_L] = \exp\left[-\sum_{\ell^*=1}^{L}\sum_{\mathcal{L}\subset\mathbb{L}}\left(\sum_{\ell\in\mathcal{L}} z_\ell^{-\frac{\theta}{1-\rho}}\right)^{1-\rho} T_{\ell^*}(\mathcal{L},t)\right],$$

where $\rho \equiv 1 - \theta/\sigma \in [0,1)$ is the correlation parameter, and the scale is $T_{\ell^*}(\mathcal{L},t) \equiv$

$\Gamma(1-\theta/\sigma)\delta_{\ell^*}(\mathcal{L},t)\Lambda_{\ell^*}(t)$. Again, if $\sigma \to \theta$, $\rho \to 0$ and productivity is independent across locations.

As our examples demonstrate, Theorem 1 provides a closed-form solution for the joint distribution of productivity given the underlying measure of ideas. In turn, this distribution determines the patterns of Ricardian trade, as we show next.

4 Linking Knowledge and Expenditure Patterns

Using the result in Theorem 1, we now characterize how innovation and diffusion shape expenditure shares and their substitution patterns. Our results establish that departures from CES expenditure imply that ideas are shared across space, connecting models of knowledge creation and diffusion to the large empirical literature that estimates import demand systems.

We assume that good markets are perfectly competitive, and that international trade is frictionless. Due to head-to-head competition, the lowest-cost location serves the market for good v, and its price equals

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

where $W_\ell(t)$ denotes the nominal wage in ℓ. Thanks to max-stability, and assuming that $\theta > \eta - 1$, the share of expenditure allocated to each production location equals the probability that it is the lowest-cost producer,

$$\pi_\ell(t) = \frac{T_\ell(t)W_\ell(t)^{-\theta}G_\ell(T_1(t)W_1(t)^{-\theta},\ldots,T_L(t)W_L(t)^{-\theta};t)}{G(T_1(t)W_1(t)^{-\theta},\ldots,T_L(t)W_L(t)^{-\theta};t)}, \tag{12}$$

where $G_\ell \equiv \partial G/\partial x_\ell$. The price index is given by the correlation function, $P_\ell(t) = \gamma G(T_1(t)W_1(t)^{-\theta},\ldots,T_L(t)W_L(t)^{-\theta};t)^{-1/\theta}$ with γ a positive constant. Given frictionless trade, we normalize $P_\ell(t) = 1$ for all ℓ and t.

The expenditure function in (12) corresponds to the generalized extreme value (GEV) class, and is shaped by the correlation function. We can see this by calculating the following elasticity for $\ell \neq \ell'$:[13]

$$\varepsilon_{\ell,\ell'}(t) \equiv \frac{\partial \ln \pi_\ell(t)}{\partial \ln W_{\ell'}(t)} = -\theta \frac{T_{\ell'}(t) W_{\ell'}(t)^{-\theta} G_{\ell\ell'}(T_1(t)W_1(t)^{-\theta}, \ldots, T_L(t)W_L(t)^{-\theta}; t)}{G_\ell(T_1(t)W_1(t)^{-\theta}, \ldots, T_L(t)W_L(t)^{-\theta}; t)}, \tag{13}$$

where $G_{\ell\ell'} \equiv \frac{\partial^2 G}{\partial x_\ell \partial x_{\ell'}} \leq 0$. When the correlation function is additive, expenditure is CES, and this elasticity is zero. Given wages, any departure from CES arises from some curvature in the correlation function, $G_{\ell\ell'}/G_\ell < 0$.

The following proposition uses the result in Theorem 1 to connect expenditure and elasticities to the measure of ideas.

Proposition 1 (Expenditure shares and substitution elasticities). *Under Assumptions 1, 2, and 3, if M is differentiable, expenditure shares are*

$$\pi_\ell(t) = \frac{\int_0^\infty \frac{W_\ell(t)}{q} M_\ell\left(\frac{W_1(t)}{q}, \ldots, \frac{W_L(t)}{q}\right) \theta q^{-\theta-1} dq}{\sum_{\ell'=1}^L \int_0^\infty \frac{W_{\ell'}(t)}{q} M_{\ell'}\left(\frac{W_1(t)}{q}, \ldots, \frac{W_L(t)}{q}\right) \theta q^{-\theta-1} dq}, \tag{14}$$

where $M_\ell \equiv \frac{\partial M}{\partial a_\ell}$, and elasticities for $\ell' \neq \ell$ are

$$\varepsilon_{\ell,\ell'}(t) \equiv \frac{\partial \ln \pi_\ell(t)}{\partial \ln W_{\ell'}(t)} = \frac{\int_0^\infty \frac{W_\ell(t)}{q} \frac{W_{\ell'}(t)}{q} M_{\ell\ell'}\left(\frac{W_1(t)}{q}, \ldots, \frac{W_L(t)}{q}\right) \theta q^{-\theta-1} dq}{\int_0^\infty \frac{W_\ell(t)}{q} M_\ell\left(\frac{W_1(t)}{q}, \ldots, \frac{W_L(t)}{q}\right) \theta q^{-\theta-1} dq}, \tag{15}$$

where $M_{\ell\ell'} \equiv \frac{\partial^2 M}{\partial a_\ell \partial a_{\ell'}}$. For $\ell' = \ell$, the elasticity is implied by $\sum_{\ell'=1}^L \varepsilon_{\ell,\ell'}(t) = -\theta$.

Proof. We use the definition in (6) to replace in (12) and (13). See Appendix B. □

In (14) and (15), $\frac{W_\ell(t)}{q}$ represents a level of applicability a_ℓ. The corresponding unit cost that ℓ' needs to compete with ℓ is $\frac{W_{\ell'}(t)}{q a_\ell}$. The integrals in these equations are over applicability levels at which all locations have the same unit cost for a given q. Consequently, the expenditure share for ℓ captures the share of ideas that the location uses to produce — i.e., the share of ideas for which ℓ is the lowest-cost producer — relative to all other locations. In turn, the cross-price elasticity reflects the amount of ideas with applicability levels at which all locations have marginal cost equal to one given q, relative to the ideas for which ℓ is the lowest-cost producer. Higher $\varepsilon_{\ell,\ell'}(t)$ reflects a larger share of head-to-head competitive

[13]The elasticity is defined with respect to the real wage in location ℓ', $W_{\ell'}(t)/P_{\ell'}(t)$. With frictionless trade, $P_{\ell'}(t) = 1$, which effectively eliminates the denominator in Equation (12).

ideas with ℓ'. That is, substitution is high when the density of ideas has more mass along the ray of identical marginal costs, relative to the mass of ideas for which ℓ is the lowest-cost producer.

To illustrate Proposition 1, we provide a numerical example. Assume that the world has two locations of identical size normalized to one. To ease notation, we suppress time subscripts. We assume a functional form for the measure of ideas, M, and explore three cases: (1) diffusion is symmetric, but does not generate similar applicability levels; (2) diffusion is symmetric but with similar applicability levels; and (3) diffusion generates similar applicability when ideas come from location 1 but dissimilar applicability when ideas come from location 2.[14]

The upper panels of Figure 1 plot the density of ideas $\frac{\partial^2 M}{\partial a_1 \partial a_2}$. Additionally, these panels show three rays from the origin indicating different levels of relative wages, $\ln W_2/W_1$, with wages equalize on the 45-degree line. The lower panels show cross-price elasticities for different levels of relative wages.

In the top left panel, the distribution of ideas has two peaks. For each of these peaks, one country has a high average level of applicability and the other country has a low level. These two groups correspond to ideas discovered in each location, when it is difficult to adopt ideas discovered elsewhere. The bottom left panel shows the consequence for elasticities. Starting from equal wages, there are few ideas for which locations are head-to-head competitors, and the elasticity is low. As the relative wage increases, the locations become head-to-head competitors for more ideas and elasticities increase. In this symmetric case, equilibrium wages are equal, and the resulting cross-price elasticities are close to zero — there is little mass of ideas around the 45-degree line.

The center panels illustrate a case of (symmetric) high similarity in applicability between locations. The density of ideas is concentrated around the 45-degree line. For equal wages, cross-price elasticities are large and symmetric.

Finally, the right panels portray a case where location 2 has high applicability for all ideas, but location 1 has low applicability for some ideas. We can interpret this case as one where ideas discovered by location 1 easily diffuse to location 2, but ideas discovered by location 2 do not easily diffuse to location 1. The responses

[14] $M(a_1, a_2) = \int_0^1 \exp\left[-a_1^{-\sigma} - (a_2/\phi)^{-\sigma}\right] f(\phi; \alpha_1, \beta_1) d\phi + \int_0^1 \exp\left[-(a_1/\phi)^{-\sigma} - a_2^{-\sigma}\right] f(\phi; \alpha_2, \beta_2) d\phi$, where $f(\phi; \alpha, \beta)$ is the density of a beta random variable. Our examples correspond to: 1) $\alpha_1 = \alpha_2 = 2$ and $\beta_1 = \beta_2 = 4$; 2) $\alpha_1 = \alpha_2 = 5$ and $\beta_1 = \beta_2 = 1$; and 3) $\alpha_1 = 5$, $\alpha_2 = 2$, and $\beta_1 = 1$ and $\beta_2 = 4$.

of expenditure shares to changes in relative wages are not symmetric. Even for equal wages, expenditure on location 1 is more sensitive to relative wages. When wages in location 2 fall relative to location 1, location 2 gains a cost advantage, and because they have high applicability for all of the ideas known to location 1, expenditure on location 1 falls rapidly. Essentially, location 2 is a fierce head-to-head competitor for many of location 1's ideas, which easily diffuse, while location 1 is rarely a fierce competitor for location 2's ideas, which do not easily diffuse.

Analytical example. We now illustrate the link between the measure of ideas and expenditure using the all-or-nothing diffusion model in Section 3.1. The measure of ideas and productivity distribution are given by (10) and (11). Expenditure on ℓ is

$$\pi_\ell(t) = T_\ell^{ND}(t) W_\ell(t)^{-\theta} + T^D(t)\mathbb{W}^{-\theta}(t)\left[\frac{W_\ell(t)}{\mathbb{W}(t)}\right]^{-\sigma}, \tag{16}$$

where $\mathbb{W}(t)^{-\sigma} \equiv \sum_{\ell'=1}^{L} W_{\ell'}(t)^{-\sigma}$. The first term is the share of expenditure in goods from ℓ produced with ideas that are unique to that location (non-diffused ideas). The second term is the share of expenditure in goods from ℓ produced with diffused ideas.

The elasticity of substitution for $\ell \neq \ell'$ is

$$\varepsilon_{\ell,\ell'}(t) = (\sigma - \theta)\,\frac{T^D(t)\mathbb{W}^{-\theta}(t)}{\pi_\ell(t)}\left(\frac{W_\ell(t)}{\mathbb{W}(t)}\frac{W_{\ell'}(t)}{\mathbb{W}(t)}\right)^{-\sigma}. \tag{17}$$

This cross-price elasticity is positive when locations share ideas ($T^D(t) > 0$) and $\sigma > \theta$. With no diffusion, $T^D(t) = 0$, $\varepsilon_{\ell,\ell'}(t) = 0$, and expenditure is CES. With $\sigma \to \theta$, which means that applicability is as fat-tailed as quality, expenditure is also CES and equal to $[T_\ell^{ND}(t) + T^D(t)]W_\ell(t)^{-\theta}$. Cross-price elasticities are also zero and they contain no information about diffusion versus innovation: Only the scale parameters of the productivity distribution reflect the state of knowledge.

With instant diffusion (and $\sigma > \theta$), expenditure is $\pi_\ell(t) = T^D(t)\mathbb{W}^{-\theta}(t)\left[\frac{W_\ell(t)}{\mathbb{W}(t)}\right]^{-\sigma}$, and $\varepsilon_{\ell,\ell'}(t) = (\sigma - \theta)\left(W_{\ell'}(t)/\mathbb{W}(t)\right)^{-\sigma}$. The amount of diffused ideas only affects the elasticity through its effect on equilibrium wages, with locations with relative low wages having higher elasticities — they are fiercer head-to-head competitors.

Finally, in the limiting case of $\sigma \to \infty$, applicability is the same in all locations. Hence, diffused ideas are only used by lower-wage locations, while higher-wage locations produce with their non-diffused ideas. Because productivity is not independent across locations, cross-price elasticities convey information about the diffusion process — they reflect the shape of the distribution of applicability across ideas.

Summing up, this example demonstrates that expenditure elasticities (over and at each point in time) contain information about the global dynamics of knowledge. Cross-price elasticities reflect the presence of shared knowledge across production locations, and are therefore key to identify diffusion and to distinguish it from innovation.[15]

5 Conclusion

The trade literature has produced extremely rich estimates of substitution elasticities for international expenditure patterns (e.g. Broda et al., 2008; Costinot and Rodrìguez-Clare, 2014; Feenstra et al., 2018; Bas et al., 2017; Adao et al., 2017). In this paper, we show that, through the lens of a Ricardian model, there is more content to be read from those elasticities when trade flows are connected to technology primitives. Thanks to our result that links a structure for innovation and diffusion to max-stable Fréchet productivity, we can directly connect the dynamics of knowledge to GEV expenditure patterns. While innovation makes a country technologically distinct, reducing substitutability in expenditure, diffusion generates technological similarity, increasing head-to-head competition and substitutability. In an innovation-only model where countries do not share ideas, productivities are independent across space, and expenditure is CES. Consequently, non-CES expenditure indicates the presence of shared ideas across countries.

This result suggests that cross-price expenditure elasticities contain information about flows of ideas due to diffusion. Hence, one can use expenditure data across countries and cost shifters (such as tariffs) to estimate the parameters governing innovation and diffusion without relying on patent creation and citation data, or direct measures of technology adoption.[16] More generally, our results can prove useful for many models featuring head-to-head competition across economic units.

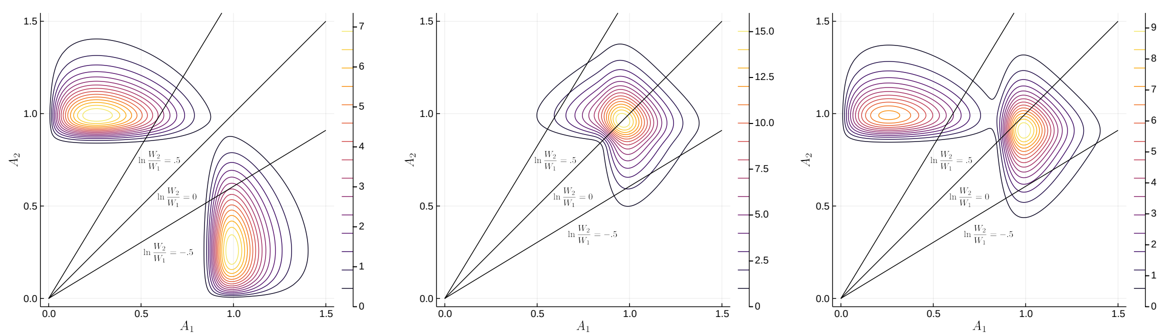
[15]In the working paper version of this paper (Lind and Ramondo, 2022a), we use United States regional income data to estimate the all-or-nothing diffusion model and uncover cross-regional innovation and diffusion patterns.

[16]In Lind and Ramondo (2022b), we introduce trade into Eaton and Kortum (1999). Using trade and factor cost data, we estimate that many high income countries primarily produce using diffused ideas, and several countries have emerged as major innovators over time.

Figure 1: Diffusion, relative wages, and cross-price elasticities.

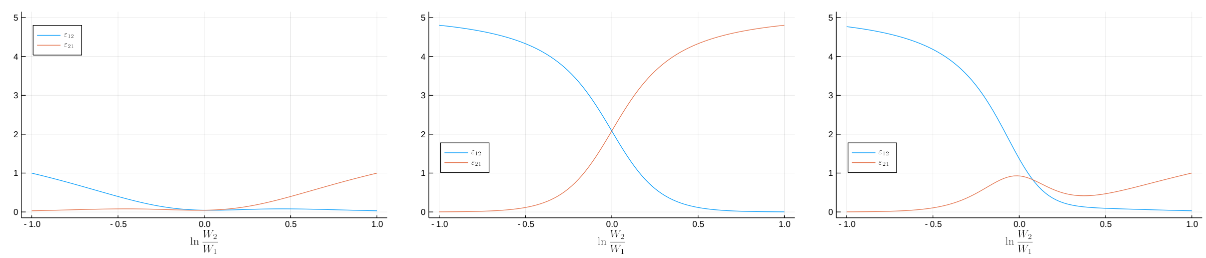
A. Joint density of applicability

Sym. diffusion, low similarity | Sym. diffusion, high similarity | Asymmetric diffusion



B. Cross-price elasticities

Sym. diffusion, low similarity | Sym. diffusion, high similarity | Asymmetric diffusion



Notes: Two sourcing locations. Left panel: symmetric diffusion, low similarity in ideas' applicability.; Center panel: symmetric diffusion, high similarity in ideas' applicability. Right panel: asymmetric diffusion between locations. Upper panels: Surface plots of $M_{12}(a_1, a_2)$. Lower panels: Cross-price elasticities agains log changes in relative wages.

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.

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.

Bass, R. (2010). The measurability of hitting times. *Electronic communications in Probability 15*, 99–105.

Bass, R. (2011). Correction to" the measurability of hitting times". *Electronic Communications in Probability 16*, 189–191.

Benhabib, J., J. Perla, and C. Tonetti (2021). Reconciling models of diffusion and innovation: A theory of the productivity distribution and technology frontier. *Econometrica 89*(5), 2261–2301.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum (2003). Plants and productivity in international trade. *The American Economic Review 93*(4), 1268–1290.

Bloom, N., T. A. Hassan, A. Kalyani, J. Lerner, and A. Tahoun (2021). 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.

Bryan, G. and M. Morten (2019). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy 127*(5).

Buera, F. J. and E. Oberfield (2020). The global diffusion of ideas. *Econometrica Volume88*(1), 83–114.

Cai, J., N. Li, and A. M. Santacreu (2020). Knowledge diffusion, trade and innovation across countries and sectors. *AEJ: Macroeconomics, forthcoming*.

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

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. Rodrìguez-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 (2018). In search of the armington elasticity. *The Review of Economics and Statistics 100*(1), 135–150.

Fosgerau, M., D. McFadden, and M. Bierlaire (2013). Choice probability generating functions. *Journal of Choice Modeling 8*, 1–18.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019). The allocation of talent and u.s. economic growth. *Econometrica 87*(5), 1439–1474.

Jovanovic, B. and Y. Nyarko (1996). Learning by doing and the choice of technology. *Econometrica 64*(6), 1299–1310.

Jovanovic, B. and R. Rob (1989). The growth and diffusion of knowledge. *Review of Economic Studies 56*, 569–582.

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), 1389–1419.

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

Lind, N. and N. Ramondo (2022b). Global knowledge and trade flows: Theory and measurement. *NBER Working Paper 30590*.

Lind, N. and N. Ramondo (2023). Trade with correlation. *American Economic Review (forthcoming) 113*(2), 1–38.

Lucas, R. E. and B. Moll (2014). Knowledge growth and the allocation of time. *Journal of Political Economy 122*(1), 1–51.

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

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

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 Theorem 1

Proof. We first prove sufficiency. Let J be an integer and fix some $\ell_j \in \{1, \dots, L\}$ and $t_j \in \mathbb{R}$ for each $j = 1, \dots, J$. Under Assumption 1, the distribution of productivity satisfies

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

where we take $1/0 = \infty$. This last expression is a void probability. We can use the marking theorem for Poisson processes (see Kingman, 1992) to calculate this void probability. 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, by the marking theorem, 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 and

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

Using this result for the mean measure,

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

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

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

Therefore, $\max_{j=1,\dots,J} \upsilon_j Z_{\ell_j}(t_j, v)$ 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

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

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 max-stable multivariate Fréchet with scale $T_\ell(t) \equiv \int a_\ell^\theta \mathrm{d}M(a_1,\dots,a_L;t)$ and correlation function $G(x_1,\dots,x_L;t) \equiv \int \max_{\ell=1,\dots,N} \frac{a_\ell^\theta}{T_\ell(t)} x_\ell \mathrm{d}M(a_1,\dots,a_L;t)$.

It remains to show that productivity is a measurable stochastic process. From Assumption 1, productivity satisfies $Z_\ell(t,v) = \max_{i=1,2,\dots} Q_i(v)A_{i\ell}(t,v)$, and $t \mapsto A_{i\ell}(t,v)$ 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 second result ensures that productivity is separable in probability, which, combined with first result, implies that a minimal spectral representation exists with respect to a standard Lebesgue space.

Let $\{Z_\ell(t,v)\}_{(\ell,t)\in\mathbb{L}\times\mathbb{R}}$ 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 $(\ell,t) \to Z_\ell(t,v)$ 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 to 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 $s \mapsto A_\ell(t,s)$ for each $(\ell,t) \in \mathbb{L}\times\mathbb{R}$ with $\int_0^1 A_\ell(t,s)^\theta \mathrm{d}\mu(s) < \infty$, and a Poisson process $\{Q_i(v), s_i(v)\}_{i=1,2,\dots}$ for each v with intensity $\theta q^{-\theta-1}\mathrm{d}q\mathrm{d}\mu(s)$ such that $Z_\ell(t,v) = \max_{i=1,2,\dots} Q_i(v)A_\ell(t,s_i(v))$. Moreover, the mapping $(\ell,t,s) \to A_\ell(t,s)$ can be taken to be jointly $\mathcal{B}(\mathbb{L}\times\mathbb{R}) \otimes \mathcal{B}([0,1])$-measurable.

Since $s \to A_\ell(t,s)$ is measurable, we can define a stochastic process $\{A_{i\ell}(t,v)\}_{(\ell,t)\in\mathbb{L}\times\mathbb{R}}$ 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 $\theta q^{-\theta-1}\mathrm{d}q\mathrm{d}\mu(s)$). The joint measurability of $(\ell,t,s) \to A_\ell(t,s)$ then implies that $A_{i\ell}(t,v) : \Omega \to \mathbb{R}$ is $\mathcal{B}(\mathbb{L}\times\mathbb{R})$-measurable for each $\omega \in \Omega$. In other words, $\{A_{i\ell}(t,v)\}_{(\ell,t)\in\mathbb{L}\times\mathbb{R}}$ is a measurable stochastic process for each $i = 1,2,\dots$ and $v \in [0,1]$.

Next, define $t_i^*(v) \equiv \min_{\ell\in\mathbb{L}} \inf\{t \in \mathbb{R} \mid A_{i\ell}(t,v) > 0\}$, which is a hitting time. Since $\{A_{i\ell}(t,v)\}_{(\ell,t)\in\mathbb{L}\times\mathbb{R}}$ is measurable and adapted to its natural filtration, it has a progressively-measurable modification. Taking $\{A_{i\ell}(t,v)\}_{(\ell,t)\in\mathbb{L}\times\mathbb{R}}$ as this modification, by the debut theorem (Bass, 2010, 2011), $t_i^*(v)$ is then a stopping time and is therefore a well-defined random variable that is adapted to the natural filtration of $\{A_{i\ell}(t,v)\}_{(\ell,t)\in\mathbb{L}\times\mathbb{R}}$. As a result, the function $s \to \min_{\ell\in\mathbb{L}} \inf\{t \in \mathbb{R} \mid A_\ell(t,s) > 0\} \equiv \tau(s)$ is measurable. Then by the mapping theorem for Poisson processes (see Klenke, 2013, Theorem

24.16), $\{Q_i(v), t_i^*(v)\}_{i=1,2,\ldots}$ is a Poisson process with intensity $\theta q^{-\theta-1}\mathrm{d}q\Lambda(\mathrm{d}t)$ where $\Lambda(B) \equiv \mu(\tau^{-1}(B))$ for each $B \in \mathcal{B}(\mathbb{R})$.

Finally, we get finite moments by applying Campbell's theorem (see Kingman, 1992):

$$\int_{\infty}^{t} \mathbb{E}\left[A_{i\ell}(t,v)^{\theta} \mid t_i^*(v) = t^*\right] \Lambda(\mathrm{d}t^*) = \mathbb{E}\sum_{i=1}^{\infty} \mathbf{1}\{Q_i(v) > 1, t_i^*(v) \leq t\} A_{i\ell}(t,v)^{\theta}$$
$$= \mathbb{E}\sum_{i=1}^{\infty} \mathbf{1}\{Q_i(v) > 1\} A_{i\ell}(t,v)^{\theta} = \mathbb{E}\sum_{i=1}^{\infty} \mathbf{1}\{Q_i(v) > 1\} A_{\ell}(t, s_i(v))^{\theta} = \int_0^1 A_\ell(t,s)^{\theta} \mathrm{d}\mu(s) < \infty.$$

□

B Proof of Proposition 1

Proof. Using the definition of the correlation function G in (6), we calculate

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

with

$$G(T_1(t)W_1(t)^{-\theta}, \ldots, T_L(t)W_L(t)^{-\theta}; t) = \sum_{\ell=1}^{L} T_\ell(t)W_\ell(t)^{-\theta} G_\ell(T_1(t)W_1(t)^{-\theta}, \ldots, T_L(t)W_L(t)^{-\theta}; t).$$

Using (12), we have

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

Then, for $\ell' \neq \ell$,

$$\frac{\partial \pi_\ell(t)}{\partial \ln W_{\ell'}(t)} = \int_0^\infty \left(\frac{W_\ell}{a_\ell}\right)^{-\theta} \frac{W_{\ell'}(t)}{W_\ell(t)} a_\ell M_{\ell\ell'}\left(\frac{W_1(t)}{W_\ell(t)} a_\ell, \ldots, a_\ell, \ldots, \frac{W_L(t)}{W_\ell(t)} a_\ell\right) \mathrm{d}a_\ell.$$

These semi-elasticities can be re-expressed as elasticities by dividing by $\pi_\ell(t)$. We then do a change of variables from a_ℓ to $q = W_\ell/a_\ell$.

□

C Independent Max-Stable Fréchet Applicability

To operationalize the closed form for the productivity distribution in Theorem 1, we focus on the class of models where (conditional) applicability is distributed independent max-stable Fréchet with shape σ. In this case, the measure of ideas can be written as

$$
\begin{aligned}
M(a_1,\ldots,a_L;t) &= \mathbb{P}\left[A_{i1}(t,v)\le a_L,\ldots,A_{iL}(t,v)\le a_L \mid t_i^*(v)\le t\right]\Lambda(t) \\
&= \int \exp\left[-\sum_{\ell=1}^{L}\left(\frac{a_\ell}{\phi_\ell}\right)^{-\sigma}\right]\mathrm{d}\mathcal{F}(\sigma,\phi_1,\ldots,\phi_L;t)\Lambda(t),
\end{aligned}
\tag{18}
$$

where $\mathcal{F}$ is a distribution function for each t, and ϕ_ℓ^σ is the scale of Fréchet applicability. Due to max stability, the conditional distribution of $\max_{\ell\in\mathbb{L}} A_{i\ell}(t,v)^\theta z_\ell^{-\theta}(t,v)$ is also max-stable Fréchet with shape σ/θ. As a consequence, we can smooth over the max operator in (3) to get

$$
\mathbb{P}\left[Z_1(t,v)\le z_L,\ldots,Z_L(t,v)\le z_L\right] = \exp\left[-\Gamma(1-\theta/\sigma)\int\left[\sum_{\ell=1}^{L}\left(\frac{a_\ell}{\phi_\ell}\right)^{-\sigma}\right]^{\frac{\theta}{\sigma}}\mathrm{d}\mathcal{F}(\sigma,\phi_1,\ldots,\phi_L;t)\Lambda(t)\right].
$$

Note that the sum in this expression converges to a max as $\sigma\to\infty$, undoing the smoothing.

This smoothed version of (3) is convenient because it implies the following closed form for expenditure.

$$
\pi_\ell(t) = \int \frac{(W_\ell(t)/\phi)^{-\sigma}}{\sum_{\ell'=1}^{L}(W_{\ell'}(t)/\phi)^{-\sigma}}\left[\sum_{\ell'=1}^{L}(W_{\ell'}(t)/\phi)^{-\sigma}\right]^{\frac{\theta}{\sigma}}\mathrm{d}\mathcal{F}(\sigma,\phi_1,\ldots,\phi_L;t).
$$

This demand system is a generalization of the mixed-CES demand system used in Adao et al. (2017), which arises as the limiting case as $\theta\to 0$.

The examples we use throughout the paper imply functional forms for the measure of ideas as in (18). For example, the productivity distribution implied by the case of ideas that are shared across all locations once they diffuse (all-or-nothing diffusion) corresponds to the case of

$$
\begin{aligned}
\mathcal{F}(\tilde\sigma,\phi_1,\ldots,\phi_L;t) = \sum_{\ell^*=1}^{L}\mathbf{1}\{\tilde\sigma\le\sigma,\phi_{\ell^*}\le 1,\phi_\ell\le 0\ \ \forall\ell\ne\ell^*\}&\frac{T_{\ell^*}^{ND}(t)}{\Gamma(1-\theta/\sigma)\Lambda(t)} \\
+\mathbf{1}\{\tilde\sigma\le\sigma,\phi_\ell\le 1\ \ \forall\ell\in\mathbb{L}\}&\frac{T^{D}(t)}{\Gamma(1-\theta/\sigma)\Lambda(t)}.
\end{aligned}
$$

Using these results, we can derive (9),

$$
\begin{aligned}
&-\ln\mathbb{P}\left[Z_1(t)\le z_1,\ldots,Z_L(t)\le z_L\right] = \int\max_\ell a_\ell^\theta z_\ell^{-\theta}\mathrm{d}\sum_{\ell=1}^{L}\int_{-\infty}^{t}\mathbb{P}[A_{i\ell}(t,v)\le a_\ell\mid \ell_i^*(v)=\ell,t_i^*(v)=s]\lambda_\ell(s)\mathrm{d}s \\
&=\sum_{\ell=1}^{L}\left[\int a_\ell^\theta\int_{-\infty}^{t}\mathrm{d}F^*(a_\ell\mid\ell^*,s;t)\lambda_\ell(s)\mathrm{d}s\right]z_\ell^{-\theta} = \sum_{\ell=1}^{L}\left[\int a_\ell^\theta\mathrm{d}M(a_1,\ldots,a_L;t)\right]z_\ell^{-\theta}\equiv\sum_{\ell=1}^{L}T_\ell(t)z_\ell^{-\theta},
\end{aligned}
$$

and (11),

$$
\begin{aligned}
&-\ln \mathbb{P}\left[Z_1(t) \leq z_1, \ldots, Z_L(t) \leq z_L\right] = \int \max_{\ell} a_\ell^\theta z_\ell^{-\theta} \mathrm{d}M(a_1, \ldots, a_L) \\
&= \sum_{\ell=1}^{L} \int a_\ell^\theta z_\ell^{-\theta} \mathrm{d}\left[e^{-a_\ell^{-\sigma}}(1-\delta_\ell(t))\Lambda_\ell(t)\right] + \int \max_{\ell} a_\ell^\theta z^{-\theta} \mathrm{d}\left[\prod_{\ell'=1}^{L} e^{-a_{\ell'}^{-\sigma}} \sum_{\ell=1}^{L} \delta_\ell(t)\Lambda_\ell(t)\right] \\
&= \sum_{\ell=1}^{L} \Gamma(\rho)(1-\delta_\ell(t))\Lambda_\ell(t) z^{-\theta} + \left(\sum_{\ell} z_\ell^{-\frac{\theta}{1-\rho}}\right)^{1-\rho} \Gamma(\rho) \sum_{\ell=1}^{L} \delta_\ell(t)\Lambda_\ell(t).
\end{aligned}
$$