DISCUSSION PAPER SERIES

DP17385

Global Innovation and Knowledge Diffusion

Natalia Ramondo and Nelson Lind

INTERNATIONAL TRADE AND REGIONAL ECONOMICS

MACROECONOMICS AND GROWTH

ISSN 0265-8003

Global Innovation and Knowledge Diffusion

Natalia Ramondo and Nelson Lind

Discussion Paper DP17385
Published 13 June 2022
Submitted 07 June 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Global Innovation and Knowledge Diffusion

Abstract

We develop a Ricardian model of trade in which countries innovate ideas that diffuse globally. The forces of innovation and diffusion combine to shape expenditure substitution patterns. Innovation makes a country technologically distinct, reducing their substitutability with other countries, while diffusion 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 expenditure is CES. Consequently, departures from CES expenditure reveal diffusion patterns. Our theoretical results provide a mapping between the dynamics of observable expenditure and the dynamics of innovation and knowledge diffusion.

JEL Classification: F1

Keywords: Innovation, Diffusion, Poisson processes, Frechet distribution, generalized extreme value, international trade

Natalia Ramondo - nramondo@bu.edu
Boston University and CEPR

Nelson Lind - nelson.rolland.lind@emory.edu
Emory University

Global Innovation and Knowledge Diffusion*

Nelson Lind — Emory University

Natalia Ramondo† — BU, NBER, and CEPR

March 29, 2022

Abstract

We develop a Ricardian model of trade in which countries innovate ideas that diffuse globally. The forces of innovation and diffusion combine to shape expenditure substitution patterns. Innovation makes a country technologically distinct, reducing their substitutability with other countries, while diffusion 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 expenditure is CES. Consequently, departures from CES expenditure reveal diffusion patterns. Our theoretical results provide a mapping between the dynamics of observable expenditure and the dynamics of innovation and knowledge diffusion.

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

*We benefited from comments from Jess Benhabib, Simon Fuchs, In-Koo Cho, Hugo Hopenhayn, Jesse Perla, Chris Tonetti, Thomas Sampson, Ana Maria Santacreu, and Mike Waugh, as well as seminar participants at several conferences and institutions. This research is supported by NSF grant #1919372. All errors are our own.

†E-mail: nlind@emory.edu; 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 dissimilarities 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 — in particular, the effects coming from innovating new ideas versus adopting existing ideas that were created elsewhere.[1]

In this paper, we formalize how the footprints of knowledge dynamics are reflected in the dynamics of expenditure patterns, and more specifically, in the substitution patterns of expenditure shares between countries. From a Ricardian perspective, these observable outcomes are the consequence of the distribution of productivity across production locations and goods at any point in time. Since heterogeneity in productivity determines patterns of comparative advantage and specialization, expenditure should reflect the creation and spread of ideas. Specifically, countries with more similar technologies should specialize in the same or very similar goods, and hence, they should have more elastic expenditure substitution patterns with each other.[2]

To capture this concept, 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 locations and time changes the world-wide productivity distribution, which,

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

[2]As an example, take the case of U.S. imports of washing machines, documented in Flaaen et al. (2020). When antidumping duties were imposed to imports from Korea and Mexico in 2012, and from China in 2016, imports from the affected sources sharply decreased but quickly reallocated to other countries — to China in 2012 and to Vietnam and Thailand in 2016. These large substitution effects suggest that those countries were close competitors, and produced very similar goods with the same technologies — in fact, the same firms (LG and Samsung) operated affiliate plants in all those countries.

in turn, determines observed patterns of expenditure and substitution elasticities.

In our model, 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 expenditure has a Constant-Elasticity-of-Substitution (CES). It follows that departures from CES expenditure reveal diffusion patterns. These departures from CES are precisely what the empirical literature estimates regarding trade substitution patterns (see, among others, Broda et al., 2008; Feenstra et al., 2018; Bas et al., 2017; Adao et al., 2017; Lind and Ramondo, 2021).

The model starts by clearly separating innovation from diffusion. Ideas of different quality are discovered over time by individual locations according to a Poisson process. Conditional on an innovation's discovery location and time, other locations learn about the idea through time. While any location can use an idea, leading to non-rivalry, locations may differ in their ability to apply the idea, leading to partial excludability. Over time, the applications of each idea across locations may change, 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 max-stable multivariate Fréchet productivity with an arbitrary correlation over space. We establish this result by applying the spectral representation theorem for max-stable processes, which generates max-stable processes from Poisson processes.[3]

A max-stable Fréchet distribution of productivity delivers closed-form solutions for expenditure shares. In particular, this distribution spans the entire class of Generalized Extreme Value (GEV) demand systems (McFadden, 1978). This class features rich substitution elasticities, departs from Independence of Irrelevant Alternatives (IIA), approximates any demand system satisfying gross substitutes (Fos-

[3] De Haan (1984) establishes the existence of a spectral representation for max-stable processes that are continuous in probability in terms of an underlying Poisson process. Penrose (1992) extends the result to max-stable processes that are right-continuous in probability. Stoev and Taqqu (2005) further extend the representation to max-stable processes that are separable in probability, while Wang and Stoev (2010) provides classification results for this representation. Finally, Kabluchko (2009) establishes the existence of spectral representations on arbitrary index sets. The spectral representation theorem for max-stable processes has previously been used by Dagsvik (1994) in the context of decision theory.

gerau et al., 2013), and encompasses the large body of quantitative trade models inspired by Eaton and Kortum (2002).

Loosely speaking, a lack of correlation in productivity across locations reflects knowledge acquired mainly through innovation, while high correlation in productivity reflects knowledge acquired mainly through diffusion.[4] In turn, high correlation in productivity leads to high substitutability in expenditure, while low correlation 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 is CES, and IIA holds — as is the case in Eaton and Kortum (2001).

Relatedly, at the heart of Ricardian trade models lies 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, it is diffusion that tightens up head-to-head competition across suppliers in different locations, and hence, leads to increases in substitutability.

Overall, the link that we establish between GEV demand and the underlying innovation and diffusion of ideas provides a framework to infer knowledge dynamics from data on expenditure and costs shifters 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, 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: (1) generalize from the case of independent Fréchet 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.

[4]Our model can also capture diffusion that occurs through the activity of multinational firms. These firms bring their home technologies to the countries where they operate, making them available for production in that location.

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 sampling ideas across goods, rather than within, so that sharing ideas does not increase head-to-head competition among suppliers.[5] 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 substitution in expenditure and higher elasticities.

Our treatment of diffusion is closer to the model 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 countries. Their model, however, does not include trade flows, and hence, the pattern of comparative advantage across countries is not relevant. We expand the model to include Ricardian trade so that comparative advantage across countries matters.

Papers such as Perla et al. (2021) and Sampson (2016b) introduce models where (endogenous) innovation, diffusion, and trade interact to generate growth.[6] While 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 as in Eaton and Kortum (2002) and Lind and Ramondo (2021). This allows us to generate expenditure functions that are not restricted to CES. Similarly to previous work by the same authors (Perla and Tonetti, 2014; Sampson, 2016a), we model diffusion as a process in which 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 used anywhere.

Finally, our approach to the modeling of global innovation and diffusion has similar consequences for productivity to the ones in the model by 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.

[5] This 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.

[6] 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 across sectors and countries.

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 represents a labor market with wage $W_\ell(t)$, populated by a unit measure 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 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.

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

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 global distribution of productivity that exhibits correlation over space and time and has the property of max stability.[7] 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 locations, as in Lind and Ramondo (2021).

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

2.1 Ideas and productivity

For each good, there exists an infinite, but countable, set of ideas: $i = 1, 2, \dots$. This set contains all the ideas that will ever exists for a good. Each idea represents a physical production technique (e.g., a blueprint) 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 idea i applied to good v 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), and also adjustments needed to implement an idea in a different location from the one that innovated it, or at a different time.

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. The 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 lack-of) across locations, and time-varying correlation (or lack of) in productivity over space.

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, 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 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 $\ell \in \mathbb{L}$ and $t \in \mathbb{R}$, such that productivity is given by*

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

If a new idea i to produce a good v becomes available to a location ℓ at time t (i.e. applicability goes from zero to a positive number), 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 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 in a location can change over time due to innovation and diffusion. We define innovation and diffusion in terms of whether or not an idea is known. An idea is unknown if all locations are currently excluded from using the idea, meaning that its applicability is zero everywhere. In contrast, if some location has positive applicability, then the idea is known.

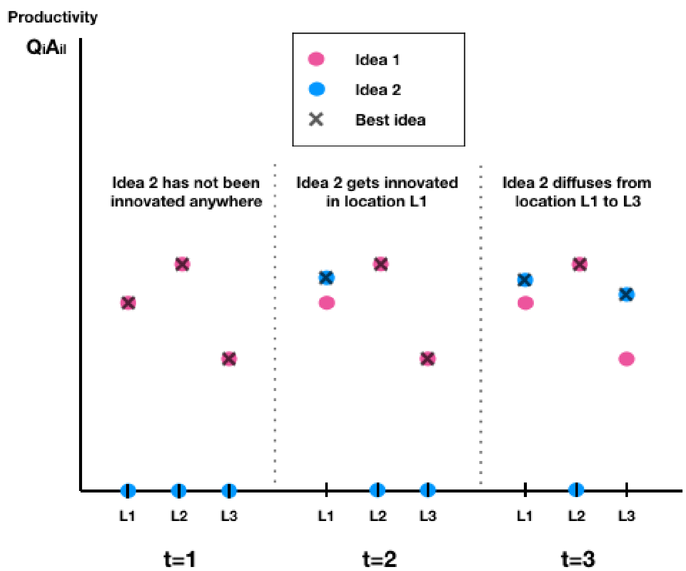
Innovation occurs when an idea first becomes known to some location. Formally, the time when an idea first arrives in location ℓ is $t_{i\ell}(v) \equiv \inf\{t \in \mathbb{R} \mid A_{i\ell}(t, v) > 0\}$, while the time when an idea is innovated — its *discovery time* — is the first time when the idea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

Given an idea's discovery, diffusion is defined as any subsequent change in applicability. While innovation makes an idea available for the very first time somewhere — applicability goes from zero to some positive number for the very first time — 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.

Hence, distinguishing between innovation and diffusion entails grouping ideas according to their discovery time. Keeping track of innovation entails measuring the accumulation of known ideas, while keeping track of diffusion means measuring the joint distribution of applicability across locations conditional on discovery times.

The diagram in Figure 1 illustrates how our structure captures the difference be-

Figure 1: Ideas, innovation, and diffusion.



Notes: The figure represents two ideas with productivity $Q_i(v)A_{1\ell}(t,v)$ and $Q_2(v)A_{2\ell}(t,v)$ (y-axis) for a given good, three locations $\ell = L1, L2, L3$, and three time periods $t = 1, 2, 3$.

tween innovation and diffusion. There are three locations and we depict two ideas applied to a given good v. The y-axis shows the productivity of each idea in each location, $Q_i(v)A_{i\ell}(t,v)$. These productivities differ across locations and times only due to differences and changes in the idea's applicability $A_{i\ell}(t,v)$; the idea's quality $Q_i(v)$ does not change over time and is common across locations. At time $t = 1$, idea 1 is known everywhere so that $A_{1\ell}(1,v) > 0$ for all ℓ. No location has innovated idea 2 so that $A_{2\ell}(1,v) = 0$ for all ℓ. Initially, all locations adopt idea 1 since this idea is the best available in each location. At time $t = 2$, idea 2 gets innovated in $L1$ — indicated by the blue point going from zero in $t = 1$ to a positive number in $t = 2$ only for $L1$. Either because this is a high-quality idea or $L1$ has high applicability, this idea gets adopted by $L1$ as it is now the best idea available to them. At $t = 3$, idea 2 diffuses from $L1$ to $L3$ — indicated by the blue point for $L3$ going from zero at $t = 2$ to positive at $t = 3$, with $L1$ already having a positive blue dot at $t = 2$. The applicability of idea 2 in $L3$ is high enough that this idea gets adopted, but not high enough to overcome the high productivity in $L1$. This lower applicability of idea 2 in $L3$ after diffusion can represent the costs to $L3$ of adapting the idea from its original form in the innovator location $L1$.

We next put stochastic structure to both the quality and applicability of ideas. We first put structure on the process of 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.*

The expected number of ideas discovered up to time t with quality above $\underline{q}$ is $\underline{q}^{-\theta}\Lambda(t)$ where $\Lambda(t) \equiv \int_{-\infty}^{t} \Lambda(\mathrm{d}t^*)$ denotes the expected number of known ideas with quality of at least 1.[8] Here, $\Lambda(\mathrm{d}t^*)$ controls the overall arrival rate of ideas, while θ controls the arrival rate of low versus high quality ideas. Among ideas known at time t with minimum quality of $\underline{q}$, quality is independent of time and distributed Pareto with shape θ and lower bound $\underline{q}$.[9] A lower θ means a fatter tail so that locations sometimes innovate very good ideas.

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.

Our structure departs from the previous literature (Kortum, 1997) in that locations do not draw the idea's 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 extreme value theory, which leads asymptotically to independent Fréchet distributions for productivity, our structure for technology leads to exact results and to max-stable Fréchet productivity with arbitrary correlation. As we show in Section 3, the Pareto tail of quality will determine the shape of the Fréchet distribution.

To characterize diffusion, we keep track of an idea's applicability over time. We only require the following assumption on diffusion.

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 $t \in \mathbb{R}$.*

Apart from regularity conditions — that ideas are independent and applicabilities are measurable and on-average finite — Assumption 3 means that improvements in applicability are just as likely to occur for low quality as for high quality ideas.

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

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

Although this restriction means that locations cannot only select the highest quality ideas to learn about, it does not limit how the applicability of an idea in one location relates to the applicability of that idea in a different location. In this way, many models of diffusion are consistent with Assumption 3.

An important implication of independence between quality and applicability 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. As a consequence, 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 $(a_1, \dots, a_L)$ 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 state variable, and specific examples of innovation and diffusion processes place further structure on M over time.

Besides Assumption 3, we require no further structure on the dynamics of diffusion to ensure max-stability of the productivity distribution. However, to clearly separate the effects of innovation from the effects of diffusion and to ensure that M evolves continuously over time, it is useful to impose one additional restriction.

We assume that each idea is innovated in a *unique discovery location*, $\ell_i^*(v)$, and at a unique time. Formally, 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 $\forall \ell \neq \ell_i^*(v)$, $t_i^*(v) < t_{i\ell}(v)$. 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^* by $\lambda_{\ell^*}(t) \equiv \mathbb{P}[\ell_i^*(v) = \ell^* \mid t_i^*(v) = t^*]\partial\Lambda(t)/\partial t$. The count of ideas with quality above $\underline{q}$ discovered up until time t in location ℓ^* is itself a Poisson process with intensity $\theta q^{-\theta-1}\mathrm{d}q\lambda_{\ell^*}(t)dt$. Hence, among ideas discovered at time t, the probability of discovery in location ℓ is $\mathbb{P}\left[\ell^*(v) = \ell \mid t_i^*(v) = t\right] = \lambda_\ell(t)/\sum_{\ell'} \lambda_{\ell'}(t)$.

Together, the assumptions on a unique location and time of discovery for an idea imply that the function in (2) becomes additively separable across discovery locations,

$$M(a_1,\ldots,a_L;t)=\sum_{\ell^*=1}^{L}\int_{-\infty}^{t}F(a_1,\ldots,a_L\mid t^*,\ell^*;t)\lambda_{\ell^*}(t^*)\mathrm{d}t^*, \tag{3}$$

where $F(a_1,\ldots,a_L \mid t^*,\ell^*;t)\equiv\mathbb{P}[A_{i\ell}(t,v)\le a_\ell\ \forall\ell \mid t_i^*(v)=t^*,\ell_i^*(v)=\ell^*]$ denotes the conditional distribution of applicability across all locations.

The dynamics of the joint distribution of applicability can then be decomposed into an innovation effect and a diffusion effect as follows:

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

where $F^*(a_{\ell^*}\mid\ell^*,t;t)\equiv\mathbb{P}[A_{i\ell^*}(t,v)\le a_{\ell^*}\mid\ell_i^*(v)=\ell^*,t_i^*(v)=t]$ 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 contributions of innovation and 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 give rise to a max-stable Fréchet distribution of productivity over time and space, as we show next.

We have left out the underpinnings of innovation, which we treat as exogenous. A micro-founded model of innovation can be introduced by endogeneizing the variable $\lambda_{\ell^*}(t)$. This variable is key in growth models, and ends up depending, for instance, on the number of researchers in a country, or the strength of the patent system. The goal of these papers is to study the determinants of growth and, in open economies, how trade liberalization affects economic growth. We also have put minimal structure on diffusion, but further assumptions on applicability would give rise to different models of diffusion. Again, the goal of many of those papers 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 different: We are less interested in proposing a new model of endogenous innovation and diffusion in an open economy, but rather in developing general tools that allow for a tractable characterization of the global evolution of knowledge, which clearly distinguishes between innovation and diffusion, as well as its consequences for the world production possibility frontier.

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, which arises from the dynamics of innovation and diffusion. The key result is that the structure for innovation and diffusion in Section 2 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,*

$$\mathbb{P}\left[Z_1(t,v) \leq z_1, \ldots, Z_L(t,v) \leq z_L\right] = \exp\left[-\int \max_{\ell \in \mathbb{L}} a_\ell^\theta z_\ell^{-\theta} dM(a_1, \ldots, a_L; t)\right]. \quad (5)$$

Proof. See Appendix A. □

Necessity follows from applying the spectral representation of max-stable processes in Wang and Stoev (2010). Essentially, this representation ensures that there always exists a Poisson process with the properties in Assumptions 2 and 3 for which productivity arises from maximizing over the points of the process (Assumption 1). The proof of sufficiency is constructive and uses properties of Poisson processes. To illustrate how Assumptions 1, 2, and 3 imply that productivity is max-stable Fréchet and generates the closed form in (5), it is useful to sketch the proof for a point in time t.

First, because each location adopts the best idea available to them under Assumption 1, the distribution of productivity across locations is given by

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

which can be expressed as a void probability,

$$\mathbb{P}\left[Z_1(v) \leq z_1, \ldots, Z_L(v) \leq z_L\right] = \mathbb{P}\left[Q_i(v) > \min_{\ell \in \mathbb{L}} \frac{z_\ell}{A_{i\ell}(t;v)} \text{ for no } i \text{ s.t. } t_i^*(v) \leq t\right]. \tag{6}$$

This expression says that no known idea can have quality above minimum productivity adjusted for the idea's applicability in each location.

We exploit properties of Poisson processes to solve for the expression in (6). Because $\{Q_i(v), t_i^*(v)\}_{i=1,2,\ldots}$ forms a Poisson process under Assumption 2 and applicabilities are conditionally independent under Assumption 3, the collection of qualities, discovery times, and applicabilities, $\{Q_i(v), t_i^*(v), \{A_{i\ell}(t,v)\}_{(\ell,t)\in\mathbb{L}\times\mathbb{R}}\}_{i=1,2,\ldots}$, contains the points of a marked Poisson process where the stochastic process for applicability, $\{A_{i\ell}(t,v)\}_{(\ell,t)\in\mathbb{L}\times\mathbb{R}}$, is the mark of the i'th point. Here, quality and discovery times capture the process of innovation, while the evolution of the marks conditional on quality and discovery captures diffusion. In turn, the subset of known ideas at time t is itself a (thinned) Poisson process with mean measure given by $\underline{q}^{-\theta} M(a_1, \ldots, a_L; t)$.[10] This mean measure can be used to solve for (6), which can be calculated using the expected number of known ideas with quality above $\min_{\ell\in\mathbb{L}} \frac{z_\ell}{a_\ell}$ and integrated over applicability levels to get (5). Changes in the distribution of productivity over time arise from changes in the measure of ideas due to innovation and diffusion: the Poisson arrival of ideas captures innovation, while the evolution of an idea's applicability conditional on its discovery (i.e. its mark) captures diffusion.

[10] $\mathbb{E} \sum_{i=1,2,\cdots|t_i^*(v)\leq t} \mathbf{1}\{Q_i(v) \geq \underline{q}, A_{i\ell}(t,v) \leq a_\ell \;\; \forall \ell \in \mathbb{L}\} = \underline{q}^{-\theta} M(a_1, \ldots, a_L; t)$, a result that uses the Pareto tail of quality and the independence between quality and discovery times in Assumption 2, as well as the independence between quality and applicability in Assumption 3.

The resulting world-wide productivity distribution in Proposition 1 at each point in time is max-stable multivariate Fréchet, as described in Lind and Ramondo (2021), and can be written as

$$\mathbb{P}\left[Z_1(t,v) \leq z_1, \ldots, Z_L(t,v) \leq z_L\right] = \exp\left[-G(T_1(t)z_1^{-\theta}, \ldots, T_L(t)z_L^{-\theta}; t)\right], \quad (7)$$

where the shape is given by θ, scales are defined as

$$T_\ell(t) \equiv \int a_\ell^\theta \mathrm{d}M(a_1, \ldots, a_L; t), \quad (8)$$

and the correlation function is defined as

$$G(x_1, \ldots, x_L; t) \equiv \int \max_{\ell \in \mathbb{L}} \frac{a_\ell^\theta}{T_\ell(t)} x_\ell \mathrm{d}M(a_1, \ldots, a_L; t). \quad (9)$$

The link with the underlying knowledge process in Proposition 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 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 scale parameter captures the average applicability of ideas in location ℓ at time t; and the correlation function, G, which captures the dependence structure of productivity across locations, reflects the patterns of similarity in applicability of M.[11]

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

3.1 From knowledge to productivity: examples

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

[11]The correlation function has several important properties. First, it is homogenous of degree one, ensuring that the joint distribution of productivity is max-stable. Second, the function G has mixed partial derivatives that exist (almost everywhere) and are continuous up to order R, with the r'th partial derivative of G with respect to r distinct arguments non-negative if r is odd and non-positive if r is even. Finally, the function G is unbounded ($G(x_1, \ldots, x_L) \to \infty$ as $x_\ell \to \infty$ for any $\ell = 1, \ldots, L$). See Lind and Ramondo (2021).

measure of ideas in (3) becomes

$$M(a_1,\ldots,a_L;t)=\sum_{\ell^*=1}^{L}\int_{-\infty}^{t}F^*(a_\ell\mid\ell^*,t^*;t)\lambda_{\ell^*}(t^*)\mathrm{d}t^*. \tag{10}$$

Using Proposition 1, the joint productivity distribution is

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

with scales given by the average applicability of ideas previously discovered in the location, and an additive correlation function (see Appendix C).

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 the Poisson arrival of ideas that 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 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 in their framework.

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 extreme opposite case of ideas that are shared across *all* locations 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,

$$M(a_1,\ldots,a_L;t)=\sum_{\ell=1}^{L}\left[(1-\xi_\ell(t))e^{-a_\ell^{-\sigma}}+\xi_\ell(t)e^{-\sum_{\ell'=1}^{L}a_{\ell'}^{-\sigma}}\right]\Lambda_\ell(t), \tag{12}$$

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

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

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

$$\mathbb{P}[Z_1(t) \leq z_1, \ldots, Z_L(t) \leq z_L] = \exp\left\{-\sum_{\ell=1}^{L} K_\ell^{ND}(t) z_\ell^{-\theta} + K^D(t)\left(\sum_{\ell=1}^{L} z_\ell^{-\frac{\theta}{1-\rho}}\right)^{1-\rho}\right\}, \tag{13}$$

where $\rho \equiv 1 - \theta/\sigma$, $K_\ell^{ND}(t) \equiv \Gamma(\rho)(1 - \xi_\ell(t))\Lambda_\ell(t)$ is the stock of non-diffused ideas from ℓ, and $K^D(t) \equiv \Gamma(\rho)\sum_{\ell=1}^{L} \xi_\ell(t)\Lambda_\ell(t)$ is the global stock of diffused ideas (see Appendix C for derivations). 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 location that have diffused to the rest of the world. In one extreme, if there is no diffusion, then $\xi_\ell(t) = 0$ for all discovery locations and this case reduces to (11) where ideas are never shared and productivity is independent Fréchet. On the other extreme, if all ideas diffuse immediately after being innovated, then $\xi_\ell(t) \to 1$ for all locations and we get a multivariate θ-Fréchet distribution with correlation parameterized by a common $\rho \in [0, 1)$.[12] The amount of correlation in productivity reflects the heterogeneity in the applicability of an idea across locations, captured by σ. As $\sigma \to \theta$, applicability is as fat-tailed as the quality component of an idea, $\rho \to 0$, and productivity is independent across locations, despite the existence of diffusion. In contrast, as $\sigma \to \infty$, there is no dispersion in applicability across locations, and $\rho \to 1$. In this case, productivity is perfectly correlated across locations 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 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 (controlled by σ)

[12] The 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 discovery locations ℓ^*. Our model of innovation and diffusion lays out the knowledge primitives behind the Fréchet multinational production model.

must be small relative to differences in quality across ideas (controlled by θ). 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 our examples demonstrate, Proposition 1 enables us to characterize the consequences of the state of knowledge 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 over time, as we show in the next section.

4 Expenditure Shares and Substitution Elasticities

We now proceed to characterize how innovation and diffusion shape expenditure shares and their substitution patterns through the evolution of correlation in productivity across space. Our results connect models of knowledge creation and 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. In this case, we normalize $P(t) = 1$. The extension to an economy with trade frictions is straightforward and does not change the main intuition of the results.[13]

[13] With an iceberg-type trade cost to ship goods from production location ℓ to final destination n, $\tau_{\ell n}(t) \geq 1$, the price index would have the subscript n, $P_n(t)$, expenditure shares would have ex-

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, and further assuming that $\theta > \eta - 1$, the share of expenditure allocated to production location ℓ equals the probability that location ℓ is the lowest-cost producer,

$$\pi_\ell(t) = \frac{T_\ell(t)W_\ell(t)^{-\theta}G_\ell(T_1(t)W_1(t)^{-\theta},\dots,T_L(t)W_L(t)^{-\theta};t)}{G(T_1(t)W_1(t)^{-\theta},\dots,T_L(t)W_L(t)^{-\theta};t)}, \tag{14}$$

where $G_\ell \equiv \frac{\partial G}{\partial x_\ell}$ (see Lind and Ramondo, 2021).[14] The demand system in (14) has the same functional form as the choice probabilities in generalized extreme value (GEV) discrete choice models (McFadden, 1978). The shape of these expenditure functions comes entirely from the correlation function. For example, the case of no diffusion, which leads to an additive correlation function — and independent productivity — entails CES expenditure (Arkolakis et al., 2012), $\pi_\ell(t) = T_\ell(t)W_\ell(t)^{-\theta}/\sum_{\ell'=1}^{L} T_{\ell'}(t)W_{\ell'}(t)^{-\theta}$.

Using (14), we can calculate the elasticity of expenditure in goods from ℓ to changes in the wage in ℓ', for $\ell \neq \ell'$,

$$\varepsilon_{\ell,\ell'}(t) \equiv \frac{\partial \ln \pi_\ell(t)}{\partial \ln W_{\ell'}(t)} = -\theta\frac{T_{\ell'}(t)W_{\ell'}(t)^{-\theta}G_{\ell\ell'}(T_1(t)W_1(t)^{-\theta},\dots,T_L(t)W_L(t)^{-\theta};t)}{G_\ell(T_1(t)W_1(t)^{-\theta},\dots,T_L(t)W_L(t)^{-\theta};t)}, \tag{15}$$

where $G_{\ell\ell'} \equiv \frac{\partial^2 G}{\partial x_\ell \partial x_{\ell'}}$. Because $G_{\ell\ell'}$ is negative, the cross-price elasticity of substitution is positive — (14) belongs to the gross substitutes class. When the correlation function is additive and expenditure is CES, the cross-price elasticity is zero and we get Independence of Irrelevant Alternatives (IIA). Any departure from CES arises from some curvature in the correlation function, $G_{\ell\ell'}/G_\ell$, and hence from production locations sharing ideas. While no diffusion leads to the CES case, a non-CES expenditure function, in the context of a Ricardian model, indicates the

porter and importer subscripts, $\pi_{\ell n}(t)$, and elasticities will reflect the substitution patterns between production locations ℓ and ℓ' to serve a particular final destination n, $\varepsilon_{\ell,\ell',n}(t)$.

[14]The expression in (14) arises from the property of max-stability implying that the conditional and unconditional distributions of the maximum coincide. This is the same property that leads to closed-form solutions in Eaton and Kortum (2002).

presence of correlation in productivity across exporters, and hence, ideas' sharing across production locations.

4.1 Linking GEV expenditure to knowledge patterns

Using the result in Proposition 1 we can directly connect expenditure and elasticities to 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 across locations, captured by 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, expenditure shares are*

$$\pi_\ell(t) = \frac{\int_0^\infty \frac{W_\ell(t)}{q} M_\ell\left(\frac{W_1(t)}{q}, \ldots, \frac{W_L(t)}{q}\right) \theta q^{-\theta-1} dq}{\sum_{\ell'=1}^{L} \int_0^\infty \frac{W_{\ell'}(t)}{q} M_{\ell'}\left(\frac{W_1(t)}{q}, \ldots, \frac{W_L(t)}{q}\right) \theta q^{-\theta-1} dq} \tag{16}$$

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

$$\varepsilon_{\ell\ell'}(t) \equiv \frac{\partial \ln \pi_\ell(t)}{\partial \ln W_{\ell'}(t)} = \frac{\int_0^\infty \frac{W_\ell(t)}{q} \frac{W_{\ell'}(t)}{q} M_{\ell\ell'}\left(\frac{W_1(t)}{q}, \ldots, \frac{W_L(t)}{q}\right) \theta q^{-\theta-1} dq}{\int_0^\infty \frac{W_\ell(t)}{q} M_\ell\left(\frac{W_1(t)}{q}, \ldots, \frac{W_L(t)}{q}\right) \theta q^{-\theta-1} dq}, \tag{17}$$

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

Proof. See Appendix B. □

In (16) and (17), the quantity $a_\ell = \frac{W_\ell(t)}{q}$ represents a level of applicability in ℓ. The corresponding cost that ℓ' needs to compete with ℓ is $\frac{W_{\ell'}(t)}{q a_\ell}$. The integrals in (16) and (17) are over applicability levels at which all locations have the same unit cost for a given quality of the idea. Consequently, the expenditure share for ℓ at time t captures the share of ideas that the location uses to produce, relative to all other locations — that is, the share of ideas for which location ℓ is the lowest-cost producer. In turn, the cross-price elasticity reflects the amount of marginal ideas (i.e. the ones with applicability levels at which all locations have marginal cost equal to one given the quality of the idea), relative to the ideas for which location ℓ is the lowest-cost producer. Higher $\varepsilon_{\ell,\ell'}(t)$ reflects a larger share of head-to-head competitive 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 better illustrate the workings of Proposition 2, we provide a numerical example. We assume that the world has two locations of identical size (normalized to one). To ease the notation, we suppress time subscripts. We assume a functional form for the joint distribution of applicability across sourcing locations $M(a_1, a_2)$ and explore three cases.[15] In the first case, diffusion is symmetric between the two locations, but does not generates similar applicability levels between locations. In the second case, diffusion is still symmetric but now leads to similar applicability levels. Finally, in the third case, diffusion occurs asymmetrically: it generates similar applicability when ideas come from location 1 but dissimilar applicability when ideas come from location 2.

Figure 2 plots the three cases. The upper panels plot the density of ideas $\frac{\partial^2 M(a_1,a_2)}{\partial a_1 \partial a_2}$ using surface heat maps to indicate the areas with more density. Additionally, these panels show three rays from the origin indicating different levels of relative wages, $\ln W_2/W_1$, with the 45-degree line indicating wage equalization. The lower panels show cross-price elasticities for different levels of relative wages across locations (in addition to the three levels plotted in the upper panels).

In the top left panel, the distribution of ideas has two peaks. For each of these peaks, there is one country with a high average level of applicability while the other country has a low level. We interpret these two groups as corresponding to ideas discovered in each location, where it is difficult to adopt ideas discovered elsewhere. Turning to the bottom left panel, we see the consequence for elasticities. Starting from equal wages, there are few ideas where the two locations are head-to-head competitors, and the elasticity is low. As the relative wage increases, the mass of ideas where the locations are head-to-head competitors increases and elasticities increase. Of course, in this symmetric case, equilibrium wages are equal

[15] We assume the measure of ideas takes the following form:

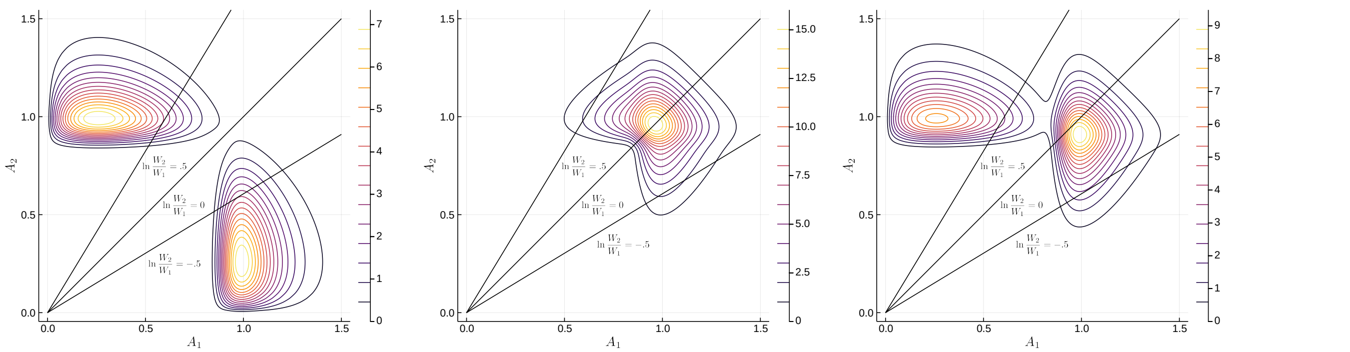
$$M(a_1, a_2) = \int_0^1 \exp\left[-a_1^{-\sigma} - (a_2/\phi)^{-\sigma}\right] f(\phi; \alpha_1, \beta_1) d\phi + \int_0^1 \exp\left[-(a_1/\phi)^{-\sigma} - a_2^{-\sigma}\right] f(\phi; \alpha_2, \beta_2) d\phi,$$

where $f(\phi; \alpha, \beta)$ is the density of a beta random variable. Our three examples vary the parameters α and β such that: 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 $\beta_2 = 4$.

Figure 2: Diffusion, relative wages, and cross-price elasticities.

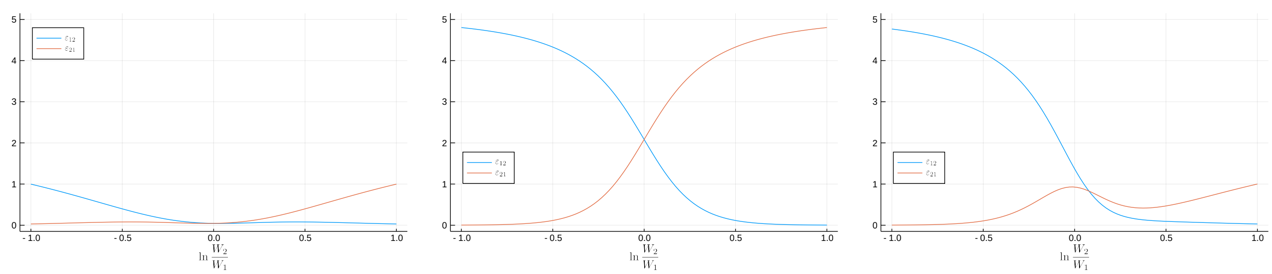
A. Joint density of applicability

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.

($\ln W_2/W_1 = 0$), 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. In this case, 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 in which location 1 has high applicability for all ideas, but location 2 has low applicability for some ideas. In this case, the responses of log expenditure shares to changes in log relative wages are not symmetric. Even for equal wages, location 1 has a higher cross-price elasticity: since their ideas have higher similarity, once they are available in location 2, this location becomes a fiercer head-to-head competitor — i.e. the share of competitive ideas is higher for location 1 than for 2. The equilibrium relative wage depends on the amount of knowledge diffused versus non-diffused as well as the parameters governing similarity in applicability.

4.2 All-or-nothing diffusion: expenditure, elasticities, and wages

The model's general equilibrium simply entails to equate income to expenditure in each location. Under the assumptions of frictionless trade and equal-sized locations, the good market equilibrium condition is

$$W_\ell(t) = \pi_\ell(t) \sum_{\ell'} W_{\ell'}(t), \tag{18}$$

for each ℓ and t. At any point in time, given the global state of knowledge, summarized by the function M, the equilibrium exists and is unique due to expenditure shares being in the gross substitute class.

To provide a solution for equilibrium expenditure shares, elasticities, and wages, we consider the case of ideas that are shared across all location once they diffuse. As in previous sections, we assume that once an idea is known in a production location, its applicability is distributed as a unit max-stable Fréchet with shape σ, independent across locations, and $\sigma > \theta$. The measure of ideas at any point in time is given by (12) and the productivity distribution is given by (13).

The expenditure share allocated to location ℓ at time t is

$$\pi_\ell(t) = \pi_\ell^{\mathrm{ND}}(t) + \pi_\ell^{\mathrm{D}}(t). \tag{19}$$

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,

$$\pi_\ell^{\mathrm{ND}}(t) = K_\ell^{ND}(t) W_\ell(t)^{-\theta}. \tag{20}$$

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

$$\pi_\ell^{\mathrm{D}}(t) = K^D(t) \left[\frac{W_\ell(t)}{\mathbb{W}(t)^\rho}\right]^{-\frac{\theta}{1-\rho}}, \tag{21}$$

where $\rho \equiv 1 - \theta/\sigma$, and $\mathbb{W}(t)^{-\frac{\theta}{1-\rho}} \equiv \sum_{\ell'=1}^{L} W_{\ell'}(t)^{-\frac{\theta}{1-\rho}}$.

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

$$\varepsilon_{\ell,\ell'}(t) = \frac{\rho\theta}{1-\rho} \frac{\pi_\ell^{\mathrm{D}}(t)}{\pi_\ell(t)} \left(\frac{W_{\ell'}(t)}{\mathbb{W}(t)}\right)^{-\frac{\theta}{1-\rho}}, \tag{22}$$

with $\rho\theta/(1-\rho) = \sigma - \theta$. This cross-price elasticity is positive when there exist shared ideas across locations and $\sigma > \theta$. It is clear that a sufficient condition for the presence of shared ideas across locations is a non-CES demand system.

With no diffusion, $K^D(t) = 0$ and expenditure is CES: cross-price elasticities are zero and the own elasticity is equal to $-\theta$. With $\sigma \to \theta$, which means that applicability is relatively fat-tailed and $\rho = 0$, expenditure is also CES, and collapses to $\pi_\ell(t) = W_\ell^{-\theta}(t) \left[K_\ell^{ND}(t) + K^D(t)\right]$. In turn, income follows the standard formula for CES expenditure as in Arkolakis et al. (2012), $W_\ell(t) = \left[\pi_\ell(t)/(K_\ell^{ND}(t) + K_\ell^D(t))\right]^{-\frac{1}{\theta}}$. In this case, cross-elasticities are zero and, therefore, contain no information about diffusion versus innovation: Only the scale parameters of the productivity distribution reflect the state of knowledge.

With instant diffusion (and $\sigma > \theta$), new ideas become immediately available to other locations so that expenditure collapses to $\pi_\ell(t) = \pi_\ell^D(t)$ and the cross elasticity is equal to $(\sigma - \theta)\left(W_{\ell'}(t)/\mathbb{W}(t)\right)^{-\sigma}$. In this case, 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 $\rho \to 1$, applicability is the same in all locations. Hence, diffused ideas will only be used by lower-wage locations, while higher-wage locations will produce with their non-diffused ideas. In this case, 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, similarity in applicability (σ) relative to quality (θ) regulates substitution patterns between locations. Given wages, innovation — captured by the prevalence of non-diffused ideas — decreases cross-price substitutability, while diffusion — captured by the prevalence of shared ideas — increases it.

To solve for the equilibrium wages, we further assume that a subset of locations innovates while the remaining ones do not. Formally, we assume that $K_\ell^{ND}(t) > 0$ for $\ell = 1, \ldots, M$ and $K_\ell^{ND}(t) = 0$ for $\ell = M+1, \ldots, N$. Let x^* denote variable x for the non-innovator locations. Specializing the equilibrium condition in (18) reveals that the wage for innovator ℓ relative to a non-innovator is equal to their relative expenditure share, $\mathcal{W}_\ell(t) \equiv W_\ell(t)/W^*(t) = \pi_\ell(t)/\pi^*(t)$. We can then characterize the equilibrium in terms of relative wages.

Specializing the expressions in (20) and (21) to the case of $N - M$ identical non-innovators, combining them as in (19), and taking the ratio of expenditures between an innovator ℓ and a non innovator, yields the following condition characterizing the relative wage of each innovator to non innovators,

$$\mathcal{W}_\ell(t) = k_\ell(t)\mathcal{W}_\ell(t)^{-\theta}\tilde{\mathbb{W}}(t)^{-\frac{\theta\rho}{1-\rho}} + \mathcal{W}_\ell(t)^{-\frac{\theta}{1-\rho}} \equiv F(\mathcal{W}_\ell(t), t),$$

with $\tilde{\mathbb{W}}(t)^{-\frac{\theta}{1-\rho}} \equiv \sum_{\ell=1}^{M} \mathcal{W}_\ell(t)^{-\frac{\theta}{1-\rho}} + N - M$ and $k_\ell(t) \equiv K_\ell^{ND}(t)/K^D(t)$. Since the function F is strictly decreasing and $F(1,t) > 1$, the relative wage for any innovator to non-innovators is unique and satisfies $\mathcal{W}(t) > 1$. This is just saying that equilibrium wages in innovator locations are strictly higher than in non-innovator locations: $W_\ell(t) > W^*(t)$, for all $\ell = 1, \ldots, M$. Moreover, as $k_\ell(t) \to 0$ so that innovator ℓ's knowledge is all diffused, their wage equalizes with the wage in non-innovator locations. Wages across the two types of locations diverge the larger the share of non-diffused knowledge, since the innovator has more sources of potentially better ideas to compete in the world market. Additionally, wages diverge more the larger the number of non-innovator locations. Because all the adopting locations have equal access to diffused ideas simultaneously, the head-to-head competition among those locations is fiercer the larger the number of competitors.

We can also calculate the relative wages between any two innovators ℓ and ℓ' at time t,

$$\frac{W_\ell(t)}{W_{\ell'}(t)} = \left[\frac{k_\ell(t) + \left(\mathcal{W}_\ell(t)/\tilde{\mathbb{W}}(t)\right)^{-\frac{\theta\rho}{1-\rho}}}{k_{\ell'}(t) + \left(\mathcal{W}_{\ell'}(t)/\tilde{\mathbb{W}}(t)\right)^{-\frac{\theta\rho}{1-\rho}}}\right]^{\frac{1}{1+\theta}}.$$

Relative to another innovator, the wage of innovator ℓ increases the higher its share of non-diffused knowledge, but the effect is partially offset by the increase in its relative wage with respect to non-innovator locations.

Inspecting (22) reveals that a lower share of diffused knowledge decreases the cross-price elasticities of innovators directly through higher $k_\ell(t)$ and through higher relative wages. For non-innovators, a higher share of diffused knowledge makes them more substitutable with an innovator since their relative wages are lower. Non innovators are always more substitutable with other non-innovator locations since wages across this set of locations are always equalized, making head-to-head competition among them fiercer.[16]

Turning to the special case of no correlation in productivity, $\rho = 0$, the wage gap between an innovator ℓ and a non-innovator location simply reflects the share of non-diffused to diffused knowledge, $\mathcal{W}_\ell(t) = (k_\ell(t) + 1)^{\frac{1}{1+\theta}}$, and similarly for the relative wage between two innovators, $W_\ell(t)/W_{\ell'}(t) = (k_\ell(t)/k_{\ell'}(t))^{\frac{1}{1+\theta}}$. In turn, expenditure shares,

$$\pi_\ell(t) = \frac{(k_\ell(t)+1)^{\frac{1}{1+\theta}}}{\sum_{\ell'}(k_{\ell'}(t)+1)^{\frac{1}{1+\theta}} + (N-M)} \quad \text{and} \quad \pi^*(t) = \frac{1}{\sum_{\ell}(k_{\ell}(t)+1)^{\frac{1}{1+\theta}} + (N-M)},$$

reflect the share of world knowledge used in production in each location — diffused knowledge in non-innovator locations, and a mix of diffused and non-diffused knowledge for innovators.

In the limiting case of $\rho \to 1$, the wage gap between innovator ℓ and any non-innovator location is $\mathcal{W}_\ell(t) = \max\left[1, (k_\ell(t)(N-M))^{\frac{1}{1+\theta}}\right]$, and expenditure shares collapse to

$$\pi_\ell(t) = \frac{(k_\ell(t)(N-M))^{\frac{1}{1+\theta}}}{\sum_{\ell'=1}^{M}(k_{\ell'}(t)(N-M))^{\frac{1}{1+\theta}} + (N-M)} \quad \text{and} \quad \pi^*(t) = \frac{1}{\sum_{\ell=1}^{M}(k_{\ell}(t)(N-M))^{\frac{1}{1+\theta}} + (N-M)},$$

when $k_\ell(t) > (N-M)^{-1}$. These are analogous to the expressions for $\rho = 0$, the difference being the type of knowledge used for production in each location. The innovators only use non-diffused knowledge; as soon as their knowledge diffuses, the adopting locations are the ones producing the good as they can do it cheaply. For this reason, cross-price elasticities collapse to infinity for the non-innovators and to zero for the innovators.

Summing up, this case demonstrates that expenditure elasticities at each point in time and over time contain information about the global dynamics of knowledge, which arises both from innovation and diffusion of ideas across locations. It is

[16]Using (22), the cross-price elasticities for each non-innovator location and an innovator ℓ with respect to location ℓ' are, respectively, $\varepsilon^*_{\ell'}(t) = \frac{\rho\theta}{1-\rho}\left(\mathcal{W}_{\ell'}(t)/\tilde{\mathbb{W}}(t)\right)^{-\frac{\theta}{1-\rho}}$, and $\varepsilon_{\ell,\ell'}(t) = \varepsilon^*_{\ell'}(t)\left[k_\ell(t)\left(\mathcal{W}_\ell(t)/\tilde{\mathbb{W}}(t)\right)^{\frac{\theta\rho}{1-\rho}} + 1\right]^{-1}$.

the cross-substitution elasticity alone, however, that reflects the presence of shared knowledge across production locations. These elasticities are therefore key to separately identify innovation from diffusion of ideas.

4.3 All-or-nothing diffusion: from expenditure to knowledge

Using the case of ideas that are shared across all location once they diffuse, and some additional assumptions on the evolution of diffused and non-diffused knowledge, we now perform a simple estimation exercise. The goal is to illustrate in a transparent way how one can estimate the evolution of knowledge from data on expenditure and some cost shifters.

We assume that time is discrete and make two additional assumptions on the diffusion and innovation processes. First, between any time t and $t+1$, a constant fraction of non-diffused ideas diffuses so that the law of motion for diffused knowledge is

$$K^D(t+1) = K^D(t) + \sum_{\ell=1}^{L} \delta_\ell K_\ell^{ND}(t). \tag{23}$$

Given the diffusion rate δ_ℓ and an initial condition on diffused knowledge, $K^D(1)$, we can infer the level of diffused knowledge for any t.

Second, we assume that new ideas arrive in proportion to a location's existing stock of total knowledge, which consists of their non-diffused knowledge and the global stock of diffused knowledge, $K_\ell^{ND}(t) + K^D(t)$. We assume that the ratio of new ideas ideas in ℓ at time $t+1$ relative to their knowledge at time t, $\alpha_\ell(t+1)$, is random with mean λ_ℓ. To get a moment condition for estimation, we assume shocks to innovation, $u_\ell(t+1)$, are unforecastable:

$$\alpha_\ell(t+1) = \lambda_\ell + u_\ell(t+1) \quad \text{where} \quad \mathbb{E}_t u_\ell(t+1) = 0. \tag{24}$$

The realized value of non-diffused knowledge in ℓ at time $t+1$ is then

$$K_\ell^{ND}(t+1) = (1-\delta_\ell) K_\ell^{ND}(t) + \alpha_\ell(t+1) \left[K^D(t) + K_\ell^{ND}(t)\right]. \tag{25}$$

Table 1: Innovation and diffusion rates: estimates. US regions.

	Innovation rates λ_ℓ	Diffusion rates δ_ℓ
Far West	0.111	0.002
Mid East	0.103	0.002
New England	0.099	0.003
South East	0.095	0.012
Great Lakes	0.084	0.002
South West	0.082	0.013
Plains	0.073	0.011
Rocky Mountains	0.045	0.017

Notes: Period: 1948-2020. See Appendix D for estimation details.

From (19), (20), and (21), expenditure shares are linear in knowledge stocks,

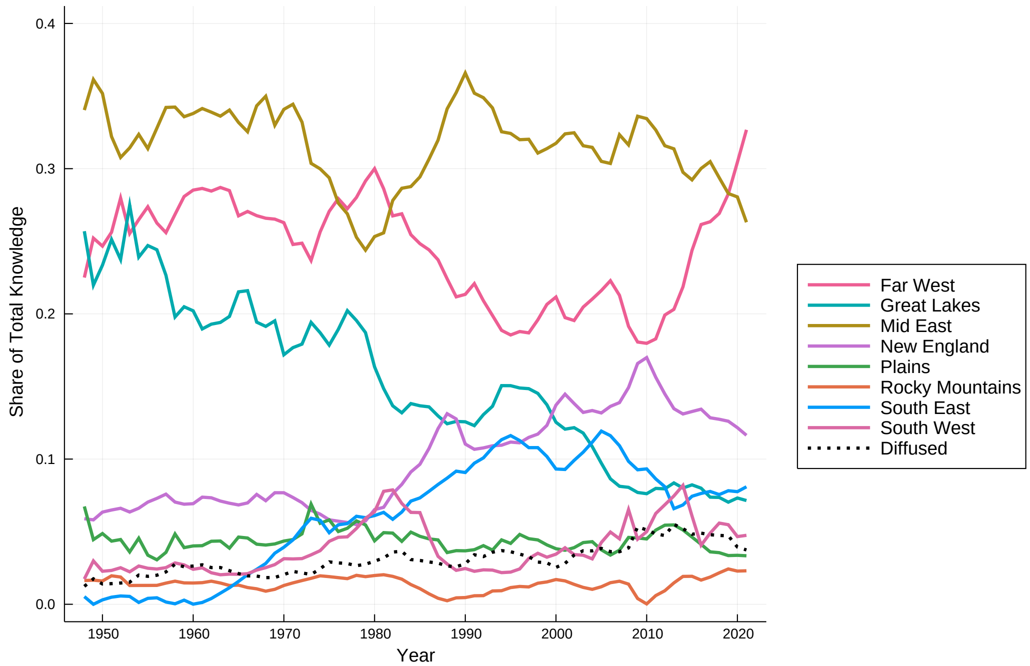
$$\pi_\ell(t) = K_\ell^{ND}(t) W_\ell(t)^{-\theta} + K^D(t)\mathbb{W}(t)^{-\theta} \left[\frac{W_\ell(t)}{\mathbb{W}(t)}\right]^{-\sigma}. \tag{26}$$

We combine the law of motions in (23) and (25), the moment condition in (24), and the expression for expenditure in (26) to implement an estimation procedure for $\{\delta_\ell\}_{\ell\in\mathbb{L}}$ and $\{\lambda_\ell\}_{\ell\in\mathbb{L}}$, and in this way obtain the dynamics of knowledge, $\{K^D(t)\}_{t=1}^T$ and $\{K_\ell^{ND}(t)\}_{t=1}^T$ for each ℓ. We set $\theta = 5$, a value standard in the trade literature. We use data on income shares to proxy for expenditure — as implied by the model's frictionless equilibrium condition in (18) — and real GDP per capita to proxy for real wages (i.e. our cost shifters), for eight regions within the United States over the period from 1948 to 2020. Appendix D presents the details of the estimation procedure. We also show that the special cases that lead to CES fit the data worse than our baseline specification.

First, given $\theta = 5$, we estimate $\sigma = 24.8$ implying that $\rho = 0.798$. This value suggests high similarity for applicability among US regions in the second half of the twentieth century.

Second, Table 1 reports estimates for innovation and diffusion rates for each US region. The Far West, which includes California, together with the Mid East, which includes the New York area, have the highest innovation rates, while the region of the Rocky Mountains has the lowest innovation rates. Diffusion rates are more similar across regions, but higher for low innovation regions.

Figure 3: Evolution of diffused and non-diffused knowledge shares, by region.



Notes: Solid lines are estimates of non-diffused knowledge, $K_\ell^{ND}(t)$, while the dash line represents diffused knowledge, $K^D(t)$, which is common for all regions. Both are as shares of total knowledge $\sum_\ell K_\ell^{ND}(t) + K^D(t)$.

Finally, what do these estimates imply for the stock of diffused and non-diffused knowledge in each region over time? Figure 3 shows the stocks of non-diffused knowledge for each region and diffused knowledge as a share total knowledge, $\sum_\ell K_\ell^{ND}(t) + K^D(t)$, at each point in time. Some interesting patterns arise: The estimation captures the decline in the share of the Great Lakes region, which mainly includes the Rust Belt, as a engine of innovation in the last part of the twenty century; the steep rise of the West Coast in the 2010's, presumably due to Silicon valley, and the rise of the South East between 1965-1995, presumably due to firms moving there. Finally, the share of diffused knowledge increases from 1.2 to 3.7 percent of total US knowledge.

Admittedly one can have richer diffusion and innovation structures, as well as richer spatial patterns (e.g. trade costs) to take to, for instance, bilateral expenditure data across countries, or any other geographical unit. Our objective in performing this simple estimation exercise has been to highlight, through the lense of a particular model of knowledge, the information about knowledge flows embedded in expenditure data. We did not need to use patent and citation data, or any other data directly measuring technology creation and adoption, which is much more difficult to obtain not only for a large set of countries but also for long time series. The link between knowledge and expenditure is made possible thanks to our result linking a structure for ideas, which encompasses many growth models of innovation and diffusion, to head-to-head competition models with GEV expenditure.

5 Conclusion

The trade literature has produced extremely rich estimates of substitution elasticities for international expenditure patterns (see Broda et al., 2008; Costinot and Rodrìguez-Clare, 2014; Feenstra et al., 2018; 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 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 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 expenditure patterns 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*.

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 (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.

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. 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.

Flaaen, A., A. Hortaçsu, and F. Tintelnot (2020). The production relocation and price effects of us trade policy: The case of washing machines. *American Economic Review 110*(7), 2103– 27.

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).

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. We first prove sufficiency. Under Assumption 1, the distribution of productivity satisfy

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

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 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,

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

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 $j = 1, \ldots, J$. The distribution of $\max_{j=1,\ldots,J} \upsilon_j Z_{\ell_j}(t_j, v)$ is

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

Therefore, $\max_{j=1,\ldots,J} \upsilon_j Z_{\ell_j}(t_j, v)$ 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 $j = 1, \ldots, L$, we have

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

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 $T_\ell(t) \equiv \int a_\ell^\theta \mathrm{d}M(a_1, \ldots, a_L; t)$ and correlation function $G(x_1, \ldots, x_L; t) \equiv \int \max_{\ell=1,\ldots,N} \frac{a_\ell^\theta}{T_\ell(t)} x_\ell \mathrm{d}M(a_1, \ldots, a_L; t)$.

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

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

Additionally, $\{A_{i\ell}(t, v) : \Omega \to \mathbb{R}\}_{(\ell,t) \in \{1,\ldots,L\} \times \mathbb{R}}$ 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 $\{Z_\ell(t, v)\}_{(\ell,t) \in \mathbb{L} \times \mathbb{R}}$ 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), 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 $s \mapsto A_\ell(t,s)$ 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 $\{Q_i(v), s_i(v)\}_{i=1,2,\ldots}$ for each v with intensity $\theta q^{-\theta-1} \mathrm{d}q \mathrm{d}\mu(s)$ such that $Z_\ell(t,v) = \max_{i=1,2,\ldots} Q_i(v) A_\ell(t, s_i(v))$. 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 $\{Q_i(v), s_i(v)\}_{i=1,2,\ldots}$ 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 $i = 1, 2, \ldots$ 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, $t_i^*(v)$ 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 $B \in \mathcal{B}(\mathbb{R})$.

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

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

and we also have

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

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

□

B Proof of Proposition 2

Proof. Using the definition of the correlation function G in (9), we calculate

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

with

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

Using (14), we have

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

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

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

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 Proposition 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

$$M(a_1,\ldots,a_L;t) = \mathbb{P}\left[A_{i1}(t,v)\le a_L,\ldots,A_{iL}(t,v)\le a_L \mid t_i^*(v)\le t\right]\Lambda(t) \tag{27}$$
$$= \int \exp\left[-\sum_{\ell=1}^{L}\left(\frac{a_\ell}{\phi_\ell}\right)^{-\sigma}\right]\mathrm{d}\mathcal{F}(\sigma,\phi_1,\ldots,\phi_L;t)\Lambda(t),$$

where $\mathcal{F}$ is a distribution function for each t, and $\phi^{1/\sigma}$ is the scale of Fréchet applicability. Here, we are simply adding some smoothing to the $\max$ operator in (5) as follows. 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 (5) to get

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

where ρ is defined as $\rho\equiv 1-\theta/\sigma$. In practice, this smoothing is without loss of generality because this expression limits to (5) if we consider $\sigma\to\infty$ and let $M(a_1,\ldots,a_L;t)\to\lim_{\sigma\to\infty}\mathcal{F}(\sigma,a_1,\ldots,a_L;t)\Lambda(t)$. It also ensures that we have convenient closed forms for expenditure shares,

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

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 (27). 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

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

Using these results, we can derive (11),

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

and (13),

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

D All-Or-Nothing Diffusion Model: Estimation

Multiplying (26) by $W_\ell(t)^\theta$, solving for $K_\ell^{ND}(t)$, and using $\rho = 1 - \theta/\sigma$ yields

$$
K_\ell^{ND}(t) = \underbrace{\pi_\ell(t) W_\ell(t)^\theta}_{\equiv C_\ell(t)} - \underbrace{\left(\frac{W_\ell(t)^{-\frac{\theta}{1-\rho}}}{\mathbb{W}(t)^{-\frac{\theta}{1-\rho}}}\right)^{\rho}}_{\equiv B_\ell(t)} K^D(t). \tag{28}
$$

As a result, given θ, ρ, and $K^D(t)$, we can recover non-diffused knowledge stocks from observations of income shares and real wages. The variables $C_\ell(t)$ and $B_\ell(t)$ are sufficient statistics for the data that we use to infer knowledge stocks, given θ and ρ.

Using (28) in (23) yields

$$
K^D(t+1) = K^D(t) + \sum_{\ell \in \mathbb{L}} \delta_\ell \left[C_\ell(t) - B_\ell(t) K^D(t)\right],
$$

which allows us to calculate the path for diffused knowledge given any diffusion rates, $\{\delta_\ell\}_{\ell \in \mathbb{L}}$, and an initial stock of diffused knowledge, $K^D(1)$. As a consequence, we can recover all knowledge stocks from the data for any values of θ, ρ, $\{\delta_\ell\}_{\ell \in \mathbb{L}}$, and $K^D(1)$.

We proceed to estimate these parameters based on the assumption that innovation rates are random with conditional expectation of $\mathbb{E}_t \alpha_\ell(t+1) = \lambda_\ell$. Recall that

$$
K_\ell^{ND}(t+1) = (1-\delta_\ell) K_\ell^{ND}(t) + \alpha_\ell(t+1)\left[K^D(t) + K_\ell^{ND}(t)\right],
$$

so that we can calculate shocks to innovation rates as

$$
u_\ell(t+1) \equiv \alpha_\ell(t+1) - \lambda_\ell = \frac{K_\ell^{ND}(t+1) - (1-\delta_\ell) K_\ell^{ND}(t)}{K_\ell^{ND}(t) + K^D(t)} - \lambda_\ell.
$$

Using the sample analog of $\mathbb{E} u_\ell(t+1) = 0$, we can estimate λ_ℓ as

$$
\hat{\lambda}_\ell = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{K_\ell^{ND}(t+1) - (1-\delta_\ell) K_\ell^{ND}(t)}{K_\ell^{ND}(t) + K^D(t)},
$$

and estimate shocks to innovation as

$$\hat{u}_\ell(t+1) = \alpha_\ell(t+1) - \hat{\lambda}_\ell.$$

We can now estimate the mean squared error in innovation rates as $\frac{1}{N(T-1)} \sum_{t=1}^{T-1} \sum_{\ell \in \mathbb{L}} \hat{u}_\ell(t+1)^2$. Note that, as we have concentrated out average innovation rates, this estimation criterion can be calculated for any given values of θ, ρ, $\{\delta_\ell\}_{\ell \in \mathbb{L}}$, and $K^D(1)$.

However, the model implies additional inequality restrictions that must be accounted for when estimating parameters. In particular, each $K_\ell^{ND}(t)$ must be non-negative. However, because these non-diffused knowledge stocks are decreasing in diffused knowledge (given the sufficient statistics $C_\ell(t)$ and $B_\ell(t)$), imposing this non-negativity constraint is equivalent to imposing the following upper bound on $K^D(t)$:

$$K^D(t) \leq \overline{K}^D(t) \equiv \min_{\ell \in \mathbb{L}} \frac{C_\ell(t)}{B_\ell(t)}.$$

When $K^D(t) = \overline{K}^D(t)$, there is some non-diffused knowledge stock that exactly equals zero. The right panel of Figure 4 shows the evolution of the estimates diffused knowledge and the upper bound.

In turn, this upper bound on diffused knowledge restricts what values of diffusion rates and initial diffused knowledge are consistent with the data. Under the assumption that

$$\sum_{\ell \in \mathbb{L}} \delta_\ell B_\ell(t) < 1,$$

the whole trajectory for diffused knowledge is strictly increasing in the initial knowledge stock. As a consequence, we can replace the restriction that $K^D(t) \leq \overline{K}^D(t)$ for all t with a single upper bound on the initial knowledge stock. For example, since $\rho < 1$ and $\frac{W_\ell(t)^{-\frac{\theta}{1-\rho}}}{\sum_{\ell' \in \mathbb{L}} W_{\ell'}(t)^{-\frac{\theta}{1-\rho}}} < 1$ for each ℓ, a sufficient condition is that $\sum_{\ell \in \mathbb{L}} \delta_\ell < 1$. In this case, for any diffusion rate for each ℓ, the initial condition on diffused knowledge must satisfy

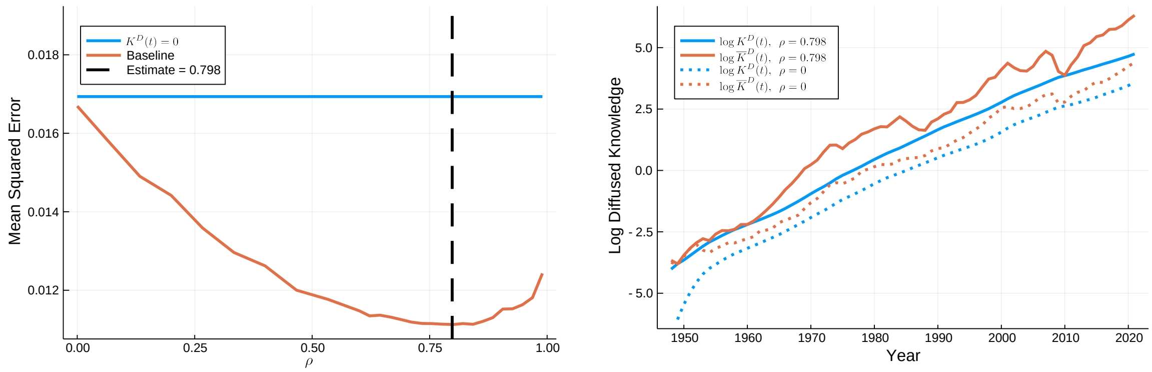
$$K^D(1) \leq \tilde{K}^D(1) \equiv \min_{t=2,\ldots,T} \frac{\overline{K}^D(t) - \sum_{s=1}^{t-1} \prod_{j=s+1}^{t-1} \left(1 - \sum_{\ell \in \mathbb{L}} \delta_\ell B_\ell(j)\right) \sum_{\ell \in \mathbb{L}} \delta_\ell C_\ell(s)}{\prod_{s=1}^{t-1} \left(1 - \sum_{\ell \in \mathbb{L}} \delta_\ell B_\ell(s)\right)}.$$

Any value for initial diffused knowledge above this upper bound implies that $K^D(t) > \overline{K}^D(t)$ for some $t > 1$. Conversely, if $K^D(1) \leq \tilde{K}^D(1)$, then $K^D(t) \leq \overline{K}^D(t)$ for all t. For any given diffusion rates, $\{\delta_\ell\}_{\ell \in \mathbb{L}}$, and the implied upper bound of $\tilde{K}^D(1)$, the set of feasible initial conditions is $K^D(1) \in [0, \tilde{K}^D(1)]$. To estimate $K^D(1)$ given $\{\delta_\ell\}_{\ell \in \mathbb{L}}$, we minimize the mean squared error in innovation rates using the bisection method for minimizing univariate functions on bounded intervals.

Given θ and ρ, it then remains to estimate $\{\delta_\ell\}_{\ell \in \mathbb{L}}$. Due to the non-linearity of the objective in these diffusion rates, we proceed by finding a convex set that bounds all feasible diffusion rates, sampling uniformly from this set, and using the value for $\{\delta_\ell\}_{\ell \in \mathbb{L}}$ that implies the lowest mean squared error in innovation shocks.

Figure 4: Estimation: mean squared errors (MSE) and diffused knowledge.

(a) Goodness of fit. (b) Evolution of diffused knowledge.



Notes: Left panel: Mean squared errors (MSE) for different values of the parameter ρ: baseline model (orange), and model with no diffused knowledge, $K^D(t) = 0$ for all t (blue). Right panel: Estimates of $K^D(t)$ and the upper bound $\overline{K}^D(t)$, for $\rho = 0.798$ and $\rho = 0$.

This bounding approach is based on the following logic. First, any value for diffusion rates must imply that $\tilde{K}^D(1) \geq 0$. If not, then there is no initial condition that does not lead to a negative value for non-diffused knowledge. Note that $\tilde{K}^D(1) \geq 0$ occurs if and only if

$$\sum_{s=1}^{t-1} \prod_{j=s+1}^{t-1} \left(1 - \sum_{\ell \in \mathbb{L}} \delta_\ell B_\ell(j)\right) \sum_{\ell \in \mathbb{L}} \delta_\ell C_\ell(s) \leq \overline{K}^D(t) \quad \text{for} \quad t = 2, \ldots, T.$$

The left hand side of this inequality is a polynomial in $\{\delta_\ell\}_{\ell \in \mathbb{L}}$ that is zero at the origin and approximately linear in a neighborhood of the origin. To get an outer approximation to the set of diffusion rates satisfying these inequalities, we find upper bounds, $\bar{\delta}_\ell$, for diffusion rates such that this inequality binds for some t, verify that the LHS of the inequality is increasing at these values, and restrict our search to the box $\{(\delta_1, \ldots, \delta_N) \mid 0 \leq \delta_\ell \leq \bar{\delta}_\ell\}$. We then sample uniformly from this outer approximation to the set of feasible diffusion rates, and only consider those sampled values for diffusion rates that satisfy these inequalities (ensuring $\tilde{K}^D(1) \geq 0$). For our estimate, we use the sampled value for $(\delta_1, \ldots, \delta_N)$ with the lowest implied mean squared error, $\frac{1}{N(T-1)} \sum_{t=1}^{T-1} \sum_{\ell \in \mathbb{L}} \hat{u}_\ell(t+1)^2$.

We repeat the estimation for several values of $\rho \in [0, 1)$, and choose the estimation with the lowest mean squared error, shown in the left panel of Figure 4. This figure also shows the results of estimating the model with no diffused knowledge, $K^D(t) = 0$ for all t, which makes the parameter ρ irrelevant. Note that CES expenditure arises when either: (1) $K^D(t) = 0$ for all t, or (2) $\rho = 0$. The blue line in the left panel corresponds to the first case, while the y-axis intercept of the orange line corresponds to the second case. For both, we get a worse fit relative to the baseline model with $\rho = 0.798$, which implies non-CES expenditure.

The right panel of Figure 4 shows the trajectories for diffused knowledge (blue) and the upper bounds on diffused knowledge (orange) for our baseline model at the estimate of $\rho = 0.798$ (solid lines) and when restricting to $\rho = 0$ (dotted lines). Since the restriction to zero correlation in productivity implies CES expenditure, the comparison between these two cases reveals the extent to which departures from CES inform inference on diffused knowledge. For $\rho = 0$, diffused knowledge acts as a common component to productivity and is identified solely from the assumption that shocks to innovation rates are un-forecastable. When allowing for $\rho > 0$, the model can capture departures from CES expenditure with larger values of diffused knowledge generating larger departures from CES. We estimate that diffused knowledge is about 150% larger on average under the estimate of $\rho = 0.798$ relative to the CES case with $\rho = 0$. The difference in these estimates comes solely from fitting patterns of non-CES expenditure in the data. We can see this directly from how the upper bound, $\overline{K}^D(t)$, differs between the two cases. Since $C_\ell(t)$ does not depend on ρ, the difference in the upper bound comes entirely from $B_\ell(t)$. This term is a sufficient statistic for how correlation in productivity influences expenditure shares. Note that in both cases, we estimate trajectories for diffused knowledge that hit each upper bound. Therefore, we can directly see how allowing for non-CES expenditure due to correlation in productivity impacts inference on diffused knowledge via the tightening of the upper bound.