Trade with Correlation*

Nelson Lind   Natalia Ramondo

Emory University   UCSD and NBER

May 24, 2019

Abstract

We develop a trade model where correlation in productivity between countries generates heterogeneity in trade elasticities. This model approximates the full class of factor demand systems consistent with Ricardian theory and formalizes Ricardo's insight that countries with relatively dissimilar technology gain more from trade. Incorporating this insight entails a simple correction to the sufficient-statistic approach used for macro counterfactuals. The novelty of our results derives from a characterization of correlation that links macro factor demand systems to technological primitives. Finally, our quantitative application shows that heterogeneity in correlation is key to the gains from trade.

JEL Codes: F1. Key Words: international trade; generalized extreme value; Fréchet distribution; gains from trade; gravity.

*We thank our discussants Costas Arkolakis and Arnaud Costinot for their very helpful comments. We benefit from comments from Rodrigo Adao, Roy Allen, Dave Donaldson, Jonathan Eaton, Pablo Fajgelbaum, Stefania Garetto, and Fernando Parro, as well as from participants at various seminars and conferences. All errors are our own.

1 Introduction

Two hundred years ago, Ricardo (1817) proposed the idea that cross-country differences in production technologies can lead to gains from trade. Ricardo's work, which extended Smith (1776)'s idea on specialization to international trade, led to the following insight: Two countries gain more from trade when they have dissimilar production possibilities. Recent evidence suggests that similarity in technology relates to country characteristics—for instance, to the proximity of countries (Keller, 2002; Bottazzi and Peri, 2003; Comin et al., 2013; Keller and Yeaple, 2013). If so, correlation in productivity may lead to heterogeneity in the gains from trade.

The Ricardian trade model in Eaton and Kortum (2002, henceforth, EK)—which gave rise to a rich theoretical and quantitative literature—does not account for correlation in productivity. They assume that productivity is independently distributed Fréchet across countries with common shape, θ. This assumption leads to tractability via a max-stability property: The distribution of the maximum is also Fréchet and its scale is the sum of the scale parameters of the marginals,

$$\mathbb{P}[\max\{A_1, \ldots, A_N\} \leq a] = \exp\left[-\left(\sum_{o=1}^{N} T_o\right) a^{-\theta}\right].$$

This additive structure, a consequence of independence, implies that comparative advantages across countries are symmetric and trade flows between country pairs have a constant elasticity of substitution (CES). As a result, the EK model cannot capture how technological similarities across countries shape the gains from trade—which may be important to understand why countries choose certain trading partners and not others.

In this paper, we develop a Ricardian theory that allows for rich patterns of correlation in technology, yet preserves the max-stability property central to the EK model. Specifically, we drop independence and assume a max-stable multivariate Fréchet distribution for productivity. In this case, the distribution of the maximum is

$$\mathbb{P}[\max\{A_1, \ldots, A_N\} \leq a] = \exp\Big[-G(T_1, \ldots, T_N) a^{-\theta}\Big],$$

for some correlation function G.[1] Countries can now have different weight on the

[1] A correlation function, often referred as a *tail dependence function* or a *extremal index function* in probability and statistics, provides a way of representing a max-stable copula.

scale of the maximum. In this way, our framework generalizes EK, maintains its tractability, and allows us to extend the results of Arkolakis et al. (2012) (henceforth, ACR) to incorporate how technological similarity between countries influences the gains from trade.

The assumption of a max-stable multivariate Fréchet productivity distribution admits a factor demand system with heterogeneous elasticities, and implies expenditure shares that belong to the generalized extreme value (GEV) class (McFadden, 1978). As a result, any trade model that generates a GEV factor demand system is observationally equivalent to a Ricardian model with a max-stable multivariate Fréchet distribution. Our framework further captures general Ricardian theory due to an approximation result: The GEV class approximates *any* factor demand system arising from Ricardian trade—without the need to restrict to Fréchet productivity distributions. Put simply, our framework captures the full aggregate implications of Ricardian trade theory.[2]

Despite this generality, our theory leads to intuitive and tractable counterfactual analysis. Within the class of GEV factor demand systems, we can calculate the gains from trade as a simple adjustment to the CES case: The results of ACR generalize, after a simple correction, to the GEV class. In the Ricardian context, this correction adjusts a country's self-trade share for correlation in technology with the rest of the world, formalizing Ricardo's insight that more dissimilar countries have higher gains from trade.

All in all, our approach uses the class of max-stable copulas to drop the assumption of independent productivity in the EK model. Yet, several new insights follow from only relaxing the dependence structure.

First, we bring existing Ricardian models into a unifying framework. The GEV class accommodates many Ricardian models of trade, such as multi-sector models (Costinot et al., 2012; Costinot and Rodrìguez-Clare, 2014; Levchenko and Zhang, 2014; DiGiovanni et al., 2014; Caliendo and Parro, 2015; Ossa, 2015; Levchenko and Zhang, 2016; French, 2016; Lashkaripour and Lugovskyy, 2017), multinational production models (Ramondo and Rodríguez-Clare, 2013; Alviarez, 2018), global value chain models (Antràs and de Gortari, 2017), and models of trade with domestic geography (Fajgelbaum and Redding, 2014; Ramondo et al., 2016; Redding, 2016).

[2] While in the body of the paper we assume that preferences are CES for comparability to the standard EK framework, this restriction is not necessary for our main results, which rely only on expenditure shares matching import probabilities. In Online Appendix O.2.1, we show how to relax this assumption.

In Section 5, we show that these models, despite their distinct micro-foundations, have identical macro factor demand systems and, hence, identical implications for macro counterfactuals.

Second, we unpack this GEV macro structure by introducing a common foundation based on technological primitives. Concretely, we provide a structure for technology that is *necessary and sufficient* for productivity to be distributed max-stable multivariate Fréchet. Our approach leverages the spectral representation theorem for max-stable processes (De Haan, 1984; Penrose, 1992; Schlather, 2002; Kabluchko, 2009), which generates extreme-value distributions from Poisson processes. This representation enables us to generate disaggregate models consistent with our aggregate framework, as well as build new economic models with rich patterns of correlation. We refer to this structure as the *global innovation representation* because it can be interpreted as the result of adopting technologies—which are a product of global innovations—based on a country's ability to apply each innovation. When countries adopt similar technologies—those with similar attributes—they have correlated productivity.

Finally, the connection between the cross-country productivity distribution and the attributes of innovations—representing the various micro-structures of non-CES Ricardian models (e.g. sectors)—allows us to tie macro substitution patterns, which are relevant for macro counterfactuals, to micro-estimates common in the trade literature.

In Section 6, we explore the empirical relevance of Ricardo's insight that the gains from trade depend on similarity between countries. That is, does allowing for heterogenous substitution patterns through correlation in productivity change the inferred gains from trade—and other counterfactual exercises? We estimate a multi-sector model of trade that allows for distance-dependent sectoral elasticities of substitution across countries. This specification captures the possibility that nearby countries may share similar technology—and therefore have correlated productivity.

Our estimates show that correlation falls with distance. This empirical result affects counterfactuals: Accounting for spatial correlation translates into gains from trade that are much higher—and much more heterogenous—than the gains calculated under independence.

This paper relates to several strands of the literature. First, we naturally relate to the large trade literature using the Ricardian-EK framework (see Eaton and Kortum, 2012, for a survey). More generally, our approach can be applied to any environment that requires Fréchet tools, with the potential of changing some of their quantitative conclusions. In particular, it can be applied to selection models used in the growth literature (Hsieh et al., 2013), and the macro development literature (Lagakos and Waugh, 2013; Bryan and Morten, 2018), as well as to recent trade models used in the urban literature (Ahlfeldt et al., 2015; Monte et al., 2015; Caliendo et al., 2017; Redding and Rossi-Hansberg, 2017).

Second, we relate to papers in the international trade literature that use non-CES factor demand systems.[3] Scarf and Wilson (2005) present a Ricardian model with a demand structure that satisfies the gross substitutability property, and in which productivity has an arbitrary distribution. They show that, in this case, the competitive equilibrium exists and is unique. We restrict our attention to the sub-class of GEV factor demand systems—which includes models used extensively in the trade literature—and show that it can approximate any demand system generated by the Ricardian model.

Building on the early work by Wilson (1980), Adao et al. (2017) show how to calculate macro counterfactual exercises in neoclassical trade models with invertible factor demand systems.[4] They also provide sufficient conditions for non-parametric identification using aggregate trade data. Their approach departs from CES demand—in which independence of irrelevant alternatives (IIA) holds—but does not necessarily lead to closed-form results. By focusing on the subclass of GEV factor demand systems, we operationalize a tractable model of Ricardian comparative advantage where IIA need not hold. Given that restriction, our aggregation result allows us to relate various micro structures to the macro demand systems studied by Adao et al. (2017), and, as a result, to incorporate disaggregate data to identify macro substitution patterns. All in all, our distinct contribution is to provide theoretical guidance on Ricardian micro-foundations underlying the class of GEV factor demand systems, and in this way, create a bridge between the macro results of Adao et al. (2017) and estimates, common in the trade literature, based on micro data.

In that regard, papers such as Caron et al. (2014), Lashkari and Mestieri (2016),

[3] A related trade literature departs from CES with the goal of analyzing endogenous markups and their effects on the gains from trade. See DeLoecker et al. (2016), Feenstra and Weinstein (2017), Bertoletti et al. (2017), and Arkolakis et al. (2017), among others.

[4] Wilson (1980) shows how a multi-country version of the Ricardian model in Dornbusch et al. (1977) can be reduced to analyzing the properties of an exchange economy—-countries trade their labor with each other.

Brooks and Pujolas (2017), Feenstra et al. (2017), and Bas et al. (2017), among others, estimate non-CES demand systems using disaggregate data. Even though they abandon the class of homothetic demand systems, which we do not, they aim, as we do, to show the consequences of abandoning linear gravity systems, and to incorporate detailed micro data to estimate key elasticities (e.g., heterogeneous price and income elasticities).[5] By linking seemingly different micro structures to common primitives of technology, our general framework provides guidance on how to incorporate the micro estimates in this literature into macro counterfactual exercises. In contrast with this literature, in our supply-side framework, substitution patterns come from the degree of technological similarity between countries. As a result, we can incorporate heterogeneous elasticities without relying on demand-side factors.[6]

Finally, our global innovation representation, which characterizes max-stable multivariate Fréchet distributions, relates to the literature on dynamic innovation and knowledge diffusion processes that generate Fréchet productivity—as in Kortum (1997), Eaton and Kortum (1999), Eaton and Kortum (2001), and Buera and Oberfield (2016). This literature uses extreme value theory to generate independent max-stable random variables. Our representation result introduces a new tool to this literature—the spectral representation theorem for max-stable processes. Contrary to the results from extreme value theory used in Kortum (1997)—which generate extreme value distributions as a limit—our approach delivers exact and closed form results while flexibly accommodating statistical dependence.

[5] Caron et al. (2014) use a constant-relative-elasticity-of-income utility functions to link characteristics of goods in production to their characteristics in preferences. Lashkari and Mestieri (2016) use constant-relative-elasticity-of-income-and-substitution (CREIS) utility functions that allow for general relationships between income and price elasticies. Brooks and Pujolas (2017) analyze the expression for gains from trade arising from models with unrestricted utility functions (typically non-homothetic) that generate a non-constant trade elasticity. Feenstra et al. (2017) use a nested CES utility function to estimate micro and macro elasticities of substitution in a multi-sector model. Finally, Bas et al. (2017) break the Pareto assumption in the Melitz model of trade to get country-pair specific aggregate elasticities, which they estimate using sectoral-level trade data.

[6] In Online Appendix O.2, we extend our framework to models where comparative advantage come from demand-side factors as in the Armington model (Anderson, 1979), and from entry of heterogenous firms as in the Krugman-Melitz model (Krugman, 1980; Melitz, 2003). Similarly to ACR, these results make clear which assumptions on economic fundamentals lead to equivalence within a large and useful class of models.

2 Ricardian Model of Trade

Consider a global economy consisting of N countries that produce and trade in a continuum of product varieties $v \in [0,1]$. Consumers have identical CES preferences with elasticity of substitution $\epsilon > -1$, $C_d = \left(\int_0^1 C_d(v)^{\frac{\epsilon}{\epsilon+1}} \mathrm{d}v\right)^{\frac{\epsilon+1}{\epsilon}}$. Given total expenditure of X_d, expenditure on variety v is $X_d(v) \equiv P_d(v)C_d(v) = (P_d(v)/P_d)^{-\epsilon} X_d$ where $P_d(v)$ is the price of the variety, and $P_d = \left(\int_0^1 P_d(v)^{-\epsilon} \mathrm{d}v\right)^{-\frac{1}{\epsilon}}$ is the price level in country d.

We assume that the production function for varieties presents constant returns to scale in labor and depends on both the origin country o where the good gets produced and the destination market d where it gets delivered. For each $v \in [0,1]$, output $Y_{od}(v)$ satisfies

$$Y_{od}(v) = A_{od}(v) L_{od}(v), \tag{1}$$

where $L_{od}(v)$ is the amount of labor used to produce variety v at origin o for delivery to d and $A_{od}(v)$ is the marginal product of labor—referred to as *productivity*. This productivity variable captures both efficiency of production in the origin and inefficiencies associated with delivery to the destination—trade costs.

The marginal cost to deliver a particular variety v to destination d from origin o is

$$c_{od}(v) = \frac{W_o}{A_{od}(v)}, \tag{2}$$

where W_o is the nominal wage in country o. We assume perfect competition so that prices are equal to unit costs. Good v is provided to country d by the cheapest supplier, so its price in the destination market is

$$P_d(v) = \min_{o=1,\dots,N} \frac{W_o}{A_{od}(v)}. \tag{3}$$

As in EK, we capture heterogeneity in production possibilities by modeling productivity as a random draw. We focus on multivariate random variables which satisfy a property known as max stability. The EK model, which is built on independent Fréchet random variables, gets its tractability from this property. By relaxing the independence assumption, we get a flexible, yet tractable, model of trade that captures Ricardo's insight that the degree of technological similarity determines the gains from trade.

2.1 Max-Stable Multivariate Fréchet Productivity

We start by providing a brief overview of max-stable multivariate Type II extreme value (Fréchet) random variables.

Definition 1 (Multivariate θ-Fréchet). *A random vector, $(A_1, \ldots, A_K)$, has a multivariate θ-Fréchet distribution if for any $\alpha_k \geq 0$ with $k = 1, \ldots, K$ the random variable $\max_{k=1,\ldots,K} \alpha_k A_k$ has a Fréchet distribution with shape parameter θ. In this case, the marginal distributions are Fréchet with (common) shape parameter θ and, for each $k = 1, \ldots, K$, satisfy*

$$\mathbb{P}\left[A_k \leq a\right] = \exp\left[-T_k a^{-\theta}\right], \tag{4}$$

for some scale parameter T_k.

This definition implies that a multivariate θ-Fréchet distribution is *max stable*—the maximum has the same marginal distribution (up to scaling). The multivariate θ-Fréchet distribution includes as special cases the independent multivariate Fréchet distribution in EK, and the symmetric multivariate Fréchet distribution used in Ramondo and Rodríguez-Clare (2013).[7] For both special cases, the max-stability property holds and lends the models their tractability.

By working with the class of multivariate θ-Fréchet random vectors, we can put minimal restrictions on dependence and maintain the key property of max-stability.[8] To make headway in that direction, we characterize the joint distribution of a multivariate θ-Fréchet random vector by first defining the function that summarizes its correlation structure.

Definition 2 (Correlation Function). *$G : \mathbb{R}_+^K \to \mathbb{R}_+$ is a correlation function if:*

1. ***(Normalization).*** *$G(0, \ldots, 0, 1, 0, \ldots, 0) = 1$;*

2. ***(Homogeneity).*** *G is homogeneous of degree one;*

3. ***(Unboundedness).*** *$G(x_1, \ldots, x_K) \to \infty$ as $x_k \to \infty$ for any $k = 1, \ldots, K$; and*

[7] Footnote 14 of Eaton and Kortum (2002) considers this symmetric correlation specification. They point out that the restriction to $\rho = 0$ is innocuous in their aggregate trade model because $\rho > 0$ also implies CES expenditure shares. However, any non-symmetric specification for correlation leads to non-CES expenditure, as we show throughout this paper.

[8] Notice that the restriction to a common shape is necessary for max stability; general multivariate Fréchet distributions may have marginal distributions with different shape parameters, in which case the maximum, even under independence, is not distributed Fréchet.

4 (Differentiability). *The mixed partial derivatives of G exist and are continuous up to order K. The k'th partial derivative of G with respect to k distinct arguments is non-negative if k is odd and non-positive if k is even.*

A correlation function is closely related to a max-stable copula and adds a normalization restriction to the definition of a social surplus function in GEV discrete choice models (McFadden, 1978).[9] The normalization restriction allows us to distinguish between absolute advantage—captured by scale parameters—and comparative advantage—captured by a correlation function. Correlation functions reflect comparative advantage because they measure relative productivity levels across varieties and across origin countries within the same destination market.

The next lemma characterizes the joint distribution of any multivariate θ-Fréchet random vector in terms of the scale parameters of its marginal distributions and a correlation function.

Lemma 1 (Correlation Function Representation). *The random vector $(A_1, \ldots, A_K)$ is multivariate θ-Fréchet if and only if there exists scale parameters T_k for $k = 1, \ldots, K$ and a correlation function G such that its joint distribution satisfies*

$$\mathbb{P}\left[A_k \leq a_k, \;\; k = 1, \ldots, K\right] = \exp\left[-G\left(T_1 a_1^{-\theta}, \ldots, T_K a_K^{-\theta}\right)\right]. \tag{5}$$

Proof. The result follows closely Theorem 3.1 of Smith (1984). See Appendix B. □

This standard result from probability theory allows us to parameterize joint distributions using scale parameters and correlation functions. The restrictions defining a correlation function ensure that (5) characterizes a valid multivariate Type II extreme value (Fréchet) distribution.

Importantly, using the characterization in Lemma 1 and the homogeneity property in Definition 2, we get the max-stability property. The maximum of a multivariate θ-Fréchet random vector is θ-Fréchet,

$$\mathbb{P}\left[\max_{k=1,\ldots,K} A_k \leq a\right] = \exp\left[-G\left(T_1, \ldots, T_K\right) a^{-\theta}\right], \tag{6}$$

with scale parameter $G(T_1, \ldots, T_K)$ and shape parameter θ. Evaluated at the scale

[9] A *max-stable copula* is a copula $C : [0,1]^N \to [0,1]$ satisfying $C(u_1, \ldots, u_N)^\eta = C(u_1^\eta, \ldots, u_N^\eta)$ for $\eta > 0$. The mapping $(u_1, \ldots, u_N) \mapsto e^{-G(-1/\ln u_1, \ldots, -1/\ln u_N)}$ is the max-stable copula associated with a correlation function G.

parameters of the marginal distributions, the correlation function acts as an aggregator that returns the scale parameter of the maximum. Moreover, as in EK, the conditional and unconditional distributions of the maximum are identical,

$$\mathbb{P}\left[\max_{k'=1,\ldots,K} A_{k'} \leq a \mid A_k = \max_{k'=1,\ldots,N} A_{k'}\right] = \mathbb{P}\left[\max_{k'=1,\ldots,K} A_{k'} \leq a\right]. \tag{7}$$

As for EK, this result is crucial for tractability because it ensures that expenditure shares simply reflect the probability of importing from an origin country.[10]

To fix ideas, consider the special case of independent θ-Fréchet productivity, used by EK. Independence implies that the correlation function is additive,

$$\mathbb{P}[A_{1d}(v) \leq a_1, \ldots, A_{od}(v) \leq a_N] = \prod_{o=1,\ldots,N} \mathbb{P}[A_{od}(v) \leq a_o] = \exp\left(-\sum_{o=1}^{N} T_{od} a_o^{-\theta}\right).$$

The max-stability property holds since

$$\mathbb{P}\left[\max_{o=1,\ldots,N} A_{od}(v) \leq a\right] = \exp\left[-\left(\sum_{o=1}^{N} T_{od}\right) a^{-\theta}\right].$$

An additive correlation function imposes a strong assumption, namely that comparative advantages across countries are symmetric. By breaking this symmetry, our model accommodates heterogeneity in comparative advantage, and, as we show in Section 3, allows us to formalize how similarity in technology matters for the gains from trade.

In Section 4, we further show how a correlation function can be constructed from fundamentals, and provide an economic justification for any choice of G.

2.2 Prices and Trade Shares

We now characterize import price distributions and expenditure shares under the assumption that productivity is multivariate θ-Fréchet. The marginal cost to deliver a particular variety v to destination d from origin o is given by (2). The joint distribution of productivity determines the joint distribution of potential import prices, as we show next.

[10] Appendix A formally presents this and other properties of Fréchet random variables which we use throughout the paper.

Proposition 1 (Potential Import Price Distribution). *If productivity has a multivariate θ-Fréchet distribution, then the joint distribution of prices presented to destination market d is given by a multivariate Weibull distribution satisfying*

$$\mathbb{P}[P_{1d}(v) \geq p_1, \ldots, P_{Nd}(v) \geq p_N] = \exp\left[-G^d\left(T_{1d}W_1^{-\theta}p_1^{\theta}, \ldots, T_{Nd}W_N^{-\theta}p_N^{\theta}\right)\right].$$

Proof. See Appendix C. □

For each origin o, the marginal distribution of prices, $\mathbb{P}[P_{od}(v) \geq p] = \exp\left[-T_{od}W_o^{-\theta}p^{\theta}\right]$, is a Weibull distribution with scale $T_{od}W_o^{-\theta}$ and shape θ.[11] The correlation function G^d determines the dependence structure of potential import prices across origins.

Define *bilateral import price indices* as $P_{od} \equiv \left(\int_0^1 P_{od}(v)^{-\epsilon}dv\right)^{-\frac{1}{\epsilon}}$. Proposition 1 together with Appendix Lemma A.1 implies that

$$P_{od} = \gamma T_{od}^{-1/\theta} W_o, \tag{8}$$

with $\gamma > 0$ (defined in Proposition 2). We can further map the scale parameters T_{od}, which are bilateral cost shifters, into standard variables in the trade literature: an origin-country *productivity index*, $A_o \equiv T_{oo}^{1/\theta}$, and an *iceberg trade cost index*, $\tau_{od} \equiv (T_{oo}/T_{od})^{1/\theta}$. The variable A_o measures a country's ability to produce goods in their domestic market, while τ_{od} measures efficiency losses associated with delivering goods to market d—the standard iceberg-type trade costs. Re-writing (8) using these indices yields $P_{od} = \gamma\tau_{od}W_o/A_o$.

Given the distribution of potential import prices, a country imports each variety from the cheapest source. The max-stability property for the productivity distribution, together with the previous characterization of the potential import price distribution, leads to closed-form results for trade shares and the price index.

Proposition 2 (Generalized EK). *Suppose productivity has a multivariate θ-Fréchet distribution with $\theta > \epsilon$. Then:*

[11] A random vector $(B_1, \ldots, B_K)$ is *multivariate Weibull* if its marginal distributions are Weibull: $\mathbb{P}[B_k \leq b] = 1 - e^{-S_k b^{\theta_k}}$ for some scale $S_k > 0$ and shape $\theta_k > 0$ across $k = 1, \ldots, K$. Note that if $(A_1, \ldots, A_k)$ is θ-Fréchet, then the vector $(A_1^{-1}, \ldots, A_K^{-1})$ is multivariate Weibull and its marginals have common shape $\theta_k = \theta$, for each $k = 1, \ldots, K$.

1. *The share of varieties that destination d imports from o is*

$$\pi_{od} = \frac{T_{od} W_o^{-\theta} G_{od}}{\sum_{o'=1}^{N} T_{o'd} W_o^{-\theta} G_{o'd}}, \tag{9}$$

where

$$G_{od} \equiv G_o^d \left(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta}\right) \quad \textit{and} \quad G_o^d \left(x_1, \ldots, x_N\right) \equiv \frac{\partial G^d \left(x_1, \ldots, x_N\right)}{\partial x_o}; \tag{10}$$

2. *The distribution of prices among goods imported into country d from o is identical to the distribution of prices in d;*

3. *Total expenditure by country d on goods from country o is $X_{od} = \pi_{od} X_d$; and*

4. *The price index in country d is*

$$P_d = \gamma G^d \left(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta}\right)^{-\frac{1}{\theta}}, \tag{11}$$

where $\gamma \equiv \Gamma\left(\frac{\theta - \epsilon}{\theta}\right)^{-\frac{1}{\epsilon}}$ and $\Gamma(\cdot)$ is the gamma function.

Proof. See Appendix D. □

First, the formula for the expenditure share, π_{od}, has the same form as choice probabilities in GEV discrete choice models (McFadden, 1978), with $T_{od} W_o^{-\theta}$ taking the place of choice-specific utility.

Second, using (9) and (11), *correlation-adjusted expenditure shares* are CES,

$$\pi_{od}^{*} \equiv \frac{\pi_{od}}{G_{od}} = T_{od} \left(\gamma \frac{W_o}{P_d}\right)^{-\theta}. \tag{12}$$

As a result, these shares constitute a gravity system, as defined by ACR, and are sufficient statistics for real import prices, $\pi_{od}^{*} = (P_{od}/P_d)^{-\theta}$.[12]

Third, as in EK, the distribution of prices among goods actually imported into market d is identical to the overall distribution of prices in d. As a result, expenditure shares are equal to the share of varieties imported into d from o.

[12]Notice that $G^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}) = \sum_{o'=1}^{N} T_{o'd} W_o^{-\theta} G_{o'd}$ so that the denominator of the expenditure share in (10) is $(P_d/\gamma)^{-\theta}$.

Finally, the price level in each destination market is determined by aggregating import price indices using the correlation function. In analogy to the discrete choice literature, welfare calculations depend crucially on the specification of this function.

2.3 GEV Factor Demand Systems

What macro substitution patterns does this theory generate? To answer this question, we first establish, in Corollary 1, that the Ricardian model with multivariate θ-Fréchet productivity implies expenditure shares in the GEV class. We next establish in Proposition 3 that the factor demand systems generated by the Ricardian model with θ-Fréchet-distributed productivity can approximate any factor demand system generated by stochastic productivity. That is, our framework is consistent with any Ricardian model with constant returns to scale in production, competitive markets, and a single factor of production in each country.[13]

First, we define a *factor demand system* for destination d as a collection of expenditure share functions $\{\pi_{od}\}_{o=1}^{N}$ such that for each $o = 1, \ldots, N$ the function $\pi_{od} : \mathbb{R}_+^N \times \mathbb{R}_+ \to [0, 1]$ is homogenous of degree zero and for any vector of wages $\mathbf{W} \equiv (W_1, \ldots, W_N) \in \mathbb{R}_+^N$ and level of expenditure $X_d \geq 0$, $\sum_{o=1}^{N} \pi_{od}(\mathbf{W}, X_d) = 1$. Next, we define the class of GEV factor demand systems.

Definition 3 (GEV Factor Demand System). *The collection $\{\pi_{od}^{GEV}\}_{o=1}^{N}$ is a generalized extreme value (GEV) factor demand system for destination d if there exists a shape parameter $\theta > 0$, scale parameters $\{T_{od}\}_{o=1}^{N}$, and a correlation function G^d satisfying*

$$\pi_{od}^{GEV}(\mathbf{W}, X_d) = \frac{T_{od} W_o^{-\theta} G_o^d(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta})}{\sum_{o'} T_{o'd} W_{o'}^{-\theta} G_{o'}^d(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta})}, \tag{13}$$

for all $o = 1, \ldots, N$.

The GEV class is homothetic, and closely related to the functional form for choice probabilities in GEV discrete choice models (McFadden, 1978), differing only by the normalization restriction in Definition 2.

[13] As mentioned in Footnote 2, the assumption on CES preferences is not crucial. In Online Appendix O.2.1, we show that if consumer preferences are homothetic and generate demand satisfying a law of large numbers on Borel subsets of the continuum of varieties, expenditure shares equal import probabilities when productivity is max-stable multivariate Fréchet.

An important class of factor demand systems within the GEV class is CES. This class is generated by an additive correlation function, implying expenditure shares of the form

$$\pi_{od}^{\text{CES}}(\mathbf{W}, X_d) = \frac{T_{od}W_o^{-\theta}}{\sum_{o'=1}^{N} T_{o'd}W_{o'}^{-\theta}}. \tag{14}$$

The CES specification includes most of the workhorse models of trade, such as Armington, Melitz, and EK (Arkolakis et al., 2012).

The GEV class, however, is much larger than the CES class. For example, consider the following cross-nested CES (CNCES) correlation function,

$$G^d(x_1, \ldots, x_N) = \sum_{k=1}^{K} \left(\sum_{o=1}^{N} (\omega_{kod}x_o)^{1/(1-\rho_k)} \right)^{1-\rho_k}. \tag{15}$$

The factor demand system implied by this correlation function is

$$\pi_{od}^{\text{CNCES}}(\mathbf{W}, X_d) = \sum_{k=1}^{K} \left(\frac{P_{kod}}{P_{kd}} \right)^{-\sigma_k} \frac{P_{kd}^{-\theta}}{\sum_{k'=1}^{M} P_{k'd}^{-\theta}}, \tag{16}$$

where $P_{kod} \equiv \gamma T_{kod}^{-1/\theta} W_o$, $P_{kd} \equiv \gamma (\sum_{o=1}^{N} P_{kod}^{-\sigma_k})^{-1/\sigma_k}$, and $\sigma_k \equiv \theta/(1-\rho_k)$.

It is clear from comparing (14) with (16) that the GEV class can generate richer patterns of substitution across exporters than CES. Comparing the elasticity of demand in the GEV and CES models, around any observed expenditure share, the difference comes from the correlation function,

$$\frac{\partial \ln \pi_{od}^{\text{GEV}}}{\partial \ln W_o} = \frac{\partial \ln \pi_{od}^{\text{CES}}}{\partial \ln W_o} + \frac{\partial \ln G_{od}}{\partial \ln W_o} \quad \text{and} \quad \frac{\partial \ln \pi_{o'd}^{\text{GEV}}}{\partial \ln W_o} = \frac{\partial \ln \pi_{o'd}^{\text{CES}}}{\partial \ln W_o} + \frac{\partial \ln G_{o'd}}{\partial \ln W_o}, \tag{17}$$

with $\partial \ln \pi_{od}^{\text{CES}} / \partial \ln W_o = -\theta\,(1 - \pi_{od})$, and $\partial \ln \pi_{o'd}^{\text{CES}} / \partial \ln W_o = \theta \pi_{o'd}$.

Our next result states that the GEV factor demand system in (13) matches the expenditure shares of the Ricardian model with multivariate θ-Fréchet productivity in Proposition 2.

Corollary 1 (GEV Equivalence). *For any trade model that generates a GEV factor demand system, there exists a Ricardian model that generates the same factor demand system for some max-stable multivariate Fréchet distribution for productivity.*

A direct implication of Corollary 1 is that the Ricardian model with multivariate θ-Fréchet productivity generates factor demand systems matching many (non-CES) trade models. In particular, many of those models are in the GEV sub-class of CNCES factor demand systems, as we show in Section 5.

We push the result in Corollary 1 one step further by adapting results from the discrete choice literature: GEV random utility models are dense in the space of all random utility models (Dagsvik, 1995). This result for choice probabilities does not directly apply to our model since we have CES demand at the variety level. However, an analogous result holds, as we show next.

Proposition 3 (GEV Approximation). *Let $\{A_{od}(v)\}_{o=1}^N$ have any multivariate distribution whose marginals have finite moment of order ϵ. Denote the factor demand system implied by the Ricardian model when productivity is distributed the same as $\{A_{od}(v)\}_{o=1}^N$ by $\{\pi_{od}\}_{o=1}^N$. Then for any compact $K \subset \mathbb{R}_+^{N+1}$ and any $\epsilon > 0$, there exists a GEV factor demand system, $\{\pi_{od}^{GEV}\}_{o=1}^N$, such that*

$$\sup_{(\mathbf{W},X_d)\in K} \left|\pi_{od}(\mathbf{W},X_d) - \pi_{od}^{GEV}(\mathbf{W},X_d)\right| < \epsilon \quad \forall o = 1,\dots,N.$$

Proof. The proof constructs an approximating GEV factor demand system that converges uniformly to the true demand system. See Appendix E. □

This result derives from the following fact: any multivariate distribution can be approximated by multiplying by independent Fréchet noise with sufficiently low dispersion. The proof of Proposition 3 shows that the factor demand system associated with productivity $\{A_{od}\}_{o=1}^N$ is uniformly approximated by a GEV factor demand system that arises from incorporating a small amount of independent Fréchet noise. That is, restricting to the GEV class amounts to smoothing over the factor demand system to ensure closed-form results.

The key implication of Proposition 3 is that *any* factor demand system generated by Ricardian trade can be approximated by a Ricardian model where productivity has a multivariate θ-Fréchet distribution. Put simply, through this approximation result, our framework encompasses the full macroeconomic implications of Ricardian trade theory.

3 Macro Counterfactuals

We next show that heterogeneity in correlation leads to heterogeneity in the gains from trade and changes calculations of (counterfactual) departures from the current equilibrium. It turns out that calculations using a GEV factor demand system are virtually identical, after a correction for correlation, to the calculations in ACR for trade models with CES factor demand systems. Moreover, the correlation correction only requires data on expenditure shares across countries, preserving the simplicity of the ACR calculation for the gains from trade.

From (9), the self-trade share is

$$\pi_{dd} = \frac{T_{dd} W_d^{-\theta} G_{dd}}{\sum_{o=1}^{N} T_{od} W_o^{-\theta} G_{od}}. \tag{18}$$

Using the expression for the price index in (11), we can write the real wage in country d as

$$\frac{W_d}{P_d} = \gamma^{-1} T_{dd}^{\frac{1}{\theta}} \left(\pi_{dd}^*\right)^{-\frac{1}{\theta}}, \tag{19}$$

where $\pi_{dd}^* \equiv \pi_{dd}/G_{dd}$ is the correlation-adjusted self-trade share defined in (12).

Let $\hat{x} \equiv x'/x$ denote the change from x to x'. Using (19), it is straightforward to show that the change in real wages between two equilibria is given by

$$\frac{\hat{W}_d}{\hat{P}_d} \equiv \frac{W_d'/P_d'}{W_d/P_d} = \left(\hat{\pi}_{dd}^*\right)^{-\frac{1}{\theta}}. \tag{20}$$

That is, in any trade model that implies a GEV factor demand system, a (log) change in equilibrium real wages—triggered by some shock to the model's parameters—is proportional to the (log) change in the correlation-adjusted self-trade share, with the factor of proportionally determined by θ.[14]

3.1 Gains From Trade: Autarky

What are the consequences of correlation in technology for the gains from trade relative to autarky? Intuitively, if two countries had perfectly correlated productivity

[14] In Online Appendix O.1, we define the equilibrium formally and show how to apply exact hat-algebra methods to solve for a change from the current (observed) equilibrium to any counterfactual equilibrium.

draws across varieties, they would offer each other identical prices, and there would be no scope for trade between them. Our correlation structure is able to capture this possibility.

In autarky, country d purchases only its own goods so that $\pi_{dd} = 1$. Moreover, as $\tau_{od} \to \infty$, $T_{od} \equiv (A_o/\tau_{od})^\theta \to 0$ for $o \neq d$, and $G_{dd} = 1$—correlation with other countries is irrelevant in autarky. The expression in (20) collapses to

$$GT_d \equiv \frac{W_d/P_d}{(W_d/P_d)^{\text{Autarky}}} = \left(\frac{\pi_{dd}}{G_{dd}}\right)^{-\frac{1}{\theta}}. \tag{21}$$

This expression generalizes the results of ACR to the class of models with GEV demand systems. With a CES factor demand system, $G_{dd} = 1$, and the gains from trade in (21) simplify to the ones in ACR where two countries with the same self-trade share have the same gains from trade. This expression for gains also admits the possibility that if two countries had the same self trade share, but one country had very similar technology to all other countries—high correlation—their gains from trade would be smaller. In contrast, if that country had dissimilar technology to other countries—low correlation—their gains from trade would be larger. In this way, our framework captures Ricardo's insight on the heterogeneity of gains from trade across countries.

Concretely, consider a three-country world with a correlation function given by

$$G^d(x_1, x_2, x_3) = \left(x_1^{1/(1-\rho)} + x_2^{1/(1-\rho)}\right)^{1-\rho} + x_3,$$

which implies that the joint distribution of productivity across countries is

$$\mathbb{P}[A_{1d}(v) \leq a_1, A_{2d}(v) \leq a_2, A_{3d}(v) \leq a_3] = \exp\left[-\left((T_{1d}a_1^{-\theta})^{\frac{1}{1-\rho}} + (T_{2d}a_2^{-\theta})^{\frac{1}{1-\rho}}\right)^{1-\rho} + T_{3d}a_3^{-\theta}\right].$$

Countries 1 and 2 are technological peers, with the parameter ρ measuring the degree of correlation in their technology. Country 3's productivity is uncorrelated with productivity in countries 1 and 2. After some algebra, we get that[15]

$$G_{od} = \left(\frac{\pi_{od}}{\pi_{1d} + \pi_{2d}}\right)^\rho \quad \text{for} \quad o = 1, 2, \quad \text{and} \quad G_{3d} = 1,$$

[15] $G_{od} = \left((T_{1d}W_1^{-\theta})^{1/(1-\rho)} + (T_{2d}W_2^{-\theta})^{1/(1-\rho)}\right)^{-\rho}(T_{od}W_o^{-\theta})^{\rho/(1-\rho)}$ for $o = 1, 2$ Given that $\pi_{od} = T_{od}W_o^{-\theta}G_{od}/G^d(T_{1d}W_1^{-\theta}, T_{2d}W_2^{-\theta}, T_{3d}W_3^{-\theta})$, we can take the ratio $G_{1d}/G_{2d} = (\pi_{1d}/\pi_{2d})^\rho$ to get G_{od} for $o = 1, 2$. For country 3, $G_{3d} = 1$.

which implies that the gains from trade are

$$GT_d = \left[\pi_{dd}^{1-\rho}\left(\pi_{1d} + \pi_{2d}\right)^{\rho}\right]^{-\frac{1}{\theta}} \quad \text{for} \quad d = 1, 2 \quad \text{and} \quad GT_3 = \pi_{33}^{-\frac{1}{\theta}}.$$

The gains from trade for countries 1 and 2 depend on the degree of correlation in technology, while the gains from trade for country 3 are pinned down by the country's self-trade share. The corrected self-trade shares for country 1 and 2 end up being a Cobb-Douglas combination between each country's expenditure share on its own goods and on the aggregation of its own goods with its peer's goods—the self-trade share if countries 1 and 2 were combined into a single country. When correlation in technology is zero ($\rho = 0$), a correlation correction is unnecessary; for positive correlation, the correction increases effective self trade and implies lower gains from trade; and for perfect correlation ($\rho = 1$), the two countries are effectively a single country and the gains from trade depend on their combined self trade.

3.2 Calculating the Correlation Correction

Given the correlation function, we show that one can calculate the gains from trade directly from expenditure data, generalizing the sufficient-statistic approach in ACR.

Because the demand system is CES after correcting for correlation, correlation-adjusted shares in (12) are sufficient statistics for bilateral import prices. The procedure to compute these adjusted expenditure shares amounts to inverting the demand system in (9).

Using the definition of import price index in (8), and the homogeneity of degree zero of G_o^d, expenditure shares in (9) can be written as

$$\pi_{od} = \left(\frac{P_{od}}{P_d}\right)^{-\theta} G_o^d\left[\left(\frac{P_{1d}}{P_d}\right)^{-\theta}, \ldots, \left(\frac{P_{Nd}}{P_d}\right)^{-\theta}\right].$$

Further using (12) yields the system

$$\pi_{od} = \pi_{od}^{*} G_o^d\left(\pi_{1d}^{*}, \ldots, \pi_{Nd}^{*}\right) \quad \text{for} \quad o = 1, \ldots, N. \tag{22}$$

Given expenditure share data and the correlation function of a single destination, the expression in (22) constitutes a system of N equations in the N unknown correlation-adjusted expenditure shares across origins.[16] As a result, we only need expenditure share data to calculate the gains from trade. We still need, however, estimates of the correlation function; the next section provides results motivating estimation from disaggregate trade data.

4 From Macro to Micro

What are the origins of cross-country correlation in productivity? We next present a structure for technology that is *necessary and sufficient* for productivity to be distributed multivariate θ-Fréchet. This structure can be interpreted as the result of adopting technologies—which are a product of global innovations—based on a country's ability to apply each innovation. When countries adopt similar technologies—those with similar attributes—they have correlated productivity.

Our technology structure satisfies the following two assumptions.

Assumption 1 (Innovation Decomposition). *There exists a measurable space of attributes* $(\mathcal{X}, \mathbb{X})$ *and for each* $v \in [0,1]$ *an infinite, but countable, set of global innovations,* $i = 1, 2, \ldots$, *with global productivity* $Z_i(v) > 0$ *and attributes* $\chi_i(v) \in \mathcal{X}$, *such that*

$$A_{od}(v) = \max_{i=1,2,\ldots} Z_i(v) A_{od}(\chi_i(v)), \tag{23}$$

for some measurable function $\chi \mapsto A_{od}(\chi)$, *and each* $o, d = 1, \ldots, N$.

Assumption 1 states that, for each good v, there is a countable collection of technological innovations that represent physical techniques (i.e., blueprints) for producing a good. Each innovation is characterized by two components. *Global productivity*, $Z_i(v)$, measures the fundamental efficiency of the technique, and is identical across all origins and destinations. *Attributes*, $\chi_i(v)$, represent anything specific to the innovation that is relevant for heterogeneity in productivity across origins and destinations. For instance, one innovation's attribute might be the country where the innovation was first developed. Alternatively, attributes could include the

[16] Note that the correlation adjustment is well defined. The mapping from $\mathbb{R}_+^N$ to $\mathbb{R}_+^N$, defined by the right-hand side of the system in (22), satisfies strict gross substitutability and is homogenous of degree one. As a result, it is injective and there is a unique solution for $\{\pi_{od}^*\}_{o=1}^N$, given $\{\pi_{od}\}_{o=1}^N$ (see, for instance, Berry et al., 2013).

sectors and firms that can use the technique. Generally, attributes capture micro-foundations that underlie country-level productivity, and allow us to develop models in which those micro-foundations determine the production techniques used in each country.

The function $A_{od}(\chi)$ determines how bilateral factors (e.g. proximity) and attributes combine to determine productivity. We refer to the variable $A_{iod}(v) \equiv A_{od}(\chi_i(v))$ as the *spatial applicability* of idea i for production in o and delivery to d. For instance, if an attribute of an innovation were whether or not the innovation is known in each country, applicability would be zero in any country with no knowledge of the innovation; conversely, if applicability were positive, it could depend on the proximity between production location o and destination d.

Our next assumption states that innovations follow a Poisson process over global productivities and attributes.

Assumption 2 (Poisson Innovations). *There exists $\theta > 0$ and a σ-finite measure μ such that $\int_{\mathcal{X}} A_{od}(\chi)^{\theta} d\mu(\chi) < \infty$ and the collection $\{Z_i(v), \chi_i(v)\}_{i=1,2,\ldots}$ consists of the points of a Poisson process with intensity measure $\theta z^{-\theta-1} dz d\mu(\chi)$, i.i.d. over $v \in [0,1]$.*

First, Assumption 2 implies that the expected number of innovations with global productivity above any cut off $\underline{z} > 0$ and attributes in any set $B \in \mathbb{X}$ is

$$\mathbb{E}\left[\sum_{i=1}^{\infty} \mathbf{1}\{Z_i(v) > \underline{z}, \chi_i(v) \in B\right] = \int_B \int_{\underline{z}}^{\infty} \theta z^{-\theta-1} \mathrm{d}z \mathrm{d}\mu(\chi) = \underline{z}^{-\theta} \mu(B).$$

The measure μ is the expected number of innovations with attributes in the set B and global productivity above 1. For example, if $\chi_i(v)$ is a list of sectors that can use the innovation, $\mu(\{s_1, s_2\})$ is the expected number of innovations that can be used by both sectors s_1 and s_2.

Second, conditional on those innovations with global productivity above $\underline{z}$ and in the set B, the likelihood that $Z_i(v) = z > \underline{z}$ is

$$\frac{\partial}{\partial z} \mathbb{P}[Z_i(v) \leq z \mid Z_i(v) > \underline{z}, \chi_i(v) \in B] = \frac{\theta z^{-\theta-1} \mathrm{d}z \mu(B)}{\int_{\underline{z}}^{\infty} \theta z^{-\theta-1} \mathrm{d}z \mu(B)} = \frac{\theta \underline{z}^{\theta}}{z^{\theta+1}}.$$

That is, global productivities are independent of attributes, and, conditional on being above $\underline{z}$, are distributed Pareto with lower bound $\underline{z}$ and shape θ. As a consequence, spatial applicabilities—which arise from an innovation's attributes—

are independent of global productivities.[17] The key is that the measure μ, which determines the joint distribution of spatial applicability across origin countries, is relatively unrestricted and allows for rich patterns of correlation through the inclusion of innovation attributes.[18]

One can interpret Assumption 2 as arising from some random discovery process as in Eaton and Kortum (1999, 2010). In our static framework, we interpret i as indexing the collection of all innovations up until the present. We do not explicitly model how innovation occurs, and simply take as given the set of innovations. The key difference between our setup and the Poisson process in Eaton and Kortum (1999) is the inclusion of attributes. Rather than assuming that innovations are country specific, innovations—which represent physical methods to produce a good—are globally applicable. Origin countries adopt whichever innovation is most efficient for them depending on how attributes determine the spatial applicability of the innovation.

The following theorem characterizes multivariate θ-Fréchet distributions and is a consequence of the spectral representation theorem for max-stable processes (De Haan, 1984; Penrose, 1992; Schlather, 2002; Kabluchko, 2009).

Theorem 1 (Global Innovation Representation). *For each d, productivity across origins is multivariate θ-Fréchet if and only if it satisfies Assumptions 1 and 2. In this case, we say that productivity has a global innovation representation. Its joint distribution is*

$$\mathbb{P}\left[A_{1d}(v) \leq a_1, \ldots, A_{Nd}(v) \leq a_N\right] = \exp\left[-G^d\left(T_{1d}a_1^{-\theta}, \ldots, T_{Nd}a_N^{-\theta}\right)\right], \tag{24}$$

with scale $T_{od} \equiv \int_{\mathcal{X}} A_{od}(\chi)^{\theta} d\mu(\chi)$, for $o = 1, \ldots, N$, and correlation function

$$G^d(x_1, \ldots, x_N) \equiv \int_{\mathcal{X}} \max_{o=1,\ldots,N} \frac{A_{od}(\chi)^{\theta}}{T_{od}} x_o d\mu(\chi). \tag{25}$$

Proof. Sufficiency follows directly from Campbell's theorem (Kingman (1992)). Necessity follows from Theorem 1 in Kabluchko (2009), which states that any θ-Fréchet process has a spectral representation. See Appendix F. □

This characterization of productivity establishes primitive assumptions on the technology

[17] Any random variable defined as a measurable function of $\chi_i(v)$ has its stochastic properties derived from the measure space $(\mathcal{X}, \mathbb{X}, \mu)$. We denote by $\mathbb{P}$ the probability measure of the underlying probability space on which global productivities and attributes are jointly defined.

[18] In our examples, we use Fréchet distributions for spatial applicability only because they lead to closed-form solutions.

structure that are necessary and sufficient for θ-Fréchet-distributed productivity across origin countries. In this way, θ-Fréchet productivity can always be interpreted as arising from the spatial applicability of global technologies. Intuitively, both absolute advantage (the scale parameters) and comparative advantage (the correlation function) are the result of the ability of exporters to adopt innovations.

The result in Theorem 1 provides a method to compute scale parameters and correlation functions: They are simply the first moments of spatial applicability and the expected value of the maximum of spatial applicability (after scaling). Put differently, Theorem 1 gives guidance on how to construct max-stable copulas.

Concretely, a symmetric multivariate Fréchet distribution (as in Ramondo and Rodríguez-Clare (2013)) arises from assuming that the spatial applicability of individual technologies is independent across o and distributed Fréchet with scale S_{od} and shape $\sigma > \theta$. Then, Theorem 1 establishes that the scale parameters equal $T_{od} = \int_{\mathcal{X}} A_{od}(\chi)^{\theta} \mathrm{d}\mu(\chi) = \Gamma(1-\theta/\sigma) S_{od}^{\theta/\sigma} \mu(\mathcal{X})$, where we use Appendix Lemma A.1 to compute the integral. The scale depends on the scales of the marginal distributions of spatial applicability and is proportional to the expected number of innovations with productivity above 1, $\mu(\mathcal{X})$. The correlation function (the expectation in (25)) is derived as follows. From Appendix Lemma A.1, $(A_{od}(\chi_i(v))^{\theta}/T_{od})x_o$ is σ/θ-Fréchet with scale $x_o^{\sigma/\theta}/\mu(\mathcal{X})$. Due to independence and max-stability, the maximum over o is also σ/θ-Fréchet and its scale is the sum of the underlying scale parameters. Using Appendix Lemma A.1 to compute the integral in (25) yields

$$G^d(x_1, \ldots, x_N) = \left(\sum_{o=1}^{N} x_o^{\frac{1}{1-\rho}} \right)^{1-\rho}, \quad \text{where} \quad \rho \equiv 1 - \frac{\theta}{\sigma}, \tag{26}$$

and the implied joint distribution of productivity follows from (24) in Theorem 1.

The correlation function in (26) takes the form of a CES aggregator. The coefficient ρ measures the degree of correlation, which arises from dispersion in spatial applicability, controlled by the shape parameter σ. As $\sigma \to \theta$, dispersion in applicability is high, and $\rho \to 0$. Intuitively, when applicability becomes very fat tailed, it dominates the contribution of the common global component of productivity. In this limiting case, productivity is independent and the correlation function is additive due to the assumption that applicability is independent across countries. In contrast, as $\sigma \to \infty$, dispersion in applicability becomes negligible and $\rho \to 1$. In this case, applicability becomes deterministic and heterogeneity in productivity is entirely determined by the global component, $Z_i(v)$. As a result, productivity becomes perfectly correlated across countries.

This example provides intuition on how Theorem 1 generates varying degrees of correlation in productivity from underlying assumptions on the spatial applicability of technologies across the globe. High dispersion in spatial applicability dampens the importance of the common global component of productivity and reduces correlation, while the opposite is true when dispersion in spatial applicability is low.

Theorem 1 also allows us to relate macro-level multivariate θ-Fréchet distributions to the underlying attributes of innovations. As a consequence, the theorem implies that we can always incorporate additional (potentially observable) micro-level variables into the model via innovation attributes.

4.1 Disaggregation

We now consider the implications of Theorem 1 when (some) innovation attributes are observable. The following corollary establishes that the joint distribution of productivity, conditional on some observable component of attributes, $k_i(v)$, is also multivariate θ-Fréchet.

Corollary 2 (Disaggregation). *Suppose that productivity has a global innovation representation with attributes $\chi_i(v) = \{k_i(v), \varepsilon_i(v)\} \in \{1, \ldots, K\} \times \mathcal{E}$, and intensity $dm_k(\varepsilon)$. Then, micro-level productivity,*

$$A_{kod}(v) \equiv \max_{i=1,2,\cdots|k_i(v)=k} Z_i(v) A_{od}(k_i(v), \varepsilon_i(v)),$$

is multivariate θ-Fréchet and independent across k. The scale parameters are $T_{kod} = \int_{\mathcal{E}} A_{kod}(\varepsilon)^\theta dm_k(\varepsilon)$, and the within-$k$ correlation function is

$$G^{kd}(x_1, \ldots, x_N) = \int_{\mathcal{E}} \max_{o=1,\ldots,N} \frac{A_{od}(k,\varepsilon)^\theta}{T_{kod}} x_o dm_k(\varepsilon).$$

Characteristics, $k_i(v)$, represent potentially observable components of the innovation's attributes, while the remaining portion $\varepsilon_i(v)$ are unobservable. Producttivity is independent across and correlated within characteristics. In this way, Corollary 2 provides an interpretation for the attributes in Theorem 1—they are latent factors that generate correlation across countries.

This result connects the macro model studied so far to potential underlying micro-foundations. We can disaggregate country-level productivity by conditioning on innovation attributes, and the resulting disaggregate productivity distribution is also θ-Fréchet.

When innovation characteristics are observable, we can use this result to incorporate disaggregate expenditure data (e.g. sectoral trade data). Define expenditure on goods produced using innovations with $k_i(v) = k$ as

$$X_{kod} \equiv \int_0^1 X_d(v) \mathbf{1}\left\{ P_{kod}(v) = \min_{k'=1,\ldots,K, o'=1,\ldots,N} P_{k'o'd}(v) \right\} \mathrm{d}v,$$

where $P_{kod}(v) \equiv W_o / A_{kod}(v)$.[19] Because the within-k distribution of productivity is multivariate θ-Fréchet, the factor demand system for k-goods is GEV,

$$\frac{X_{kod}}{X_{kd}} = \frac{T_{kod} W_o^{-\theta} G_o^{kd}(T_{k1d} W_1^{-\theta}, \ldots, T_{kNd} W_N^{-\theta})}{\sum_{o'=1}^N T_{ko'd} W_{o'}^{-\theta} G_{o'}^{kd}(T_{k1d} W_1^{-\theta}, \ldots, T_{kNd}(\chi) W_N^{-\theta})}, \tag{27}$$

with $X_{kd} \equiv \sum_{o=1}^N X_{kod}$. Additionally, since productivity is independent across k, the share of total expenditure on k-goods is CES,

$$\frac{X_{kd}}{X_d} = \frac{P_{kd}^{-\theta}}{\sum_{k'=1}^K P_{k'd}^{-\theta}}, \tag{28}$$

where $P_{kd} = \gamma G^{kd}(T_{k1d} W_1^{-\theta}, \ldots, T_{kNd} W_N^{-\theta})^{-1/\theta}$. Inspection of (27) shows that correlation manifests in the trade data in terms of substitution patterns within characteristics (e.g., sectors).

We now provide a concrete example of applying Corollary 2 to generate a cross-nested CES (CNCES) factor demand system, as in (16). This example is important because many Ricardian models in the literature are instances of a CNCES factor demand system—as we show in Section 5.[20]

To get a CNCES factor demand system, we build on our previous example where independent spatial applicability led to symmetric correlation in productivity. Assume that the distribution of spatial applicability conditional on a given characteristic

[19] When $A_{kod}(v) = 0$, $P_{kod}(v) = \infty$.

[20] In Appendix G, we provide an example of a generalized mixed CES factor demand system that includes both CNCES and mixed-CES, as in Adao et al. (2017), as special cases.

is independent σ_k-Fréchet across origins and $i.i.d$ across innovations. The scale is S_{kod}. Under this assumption, Corollary 2 implies that characteristic-level productivity has a joint distribution over o that is symmetric multivariate θ-Fréchet with correlation function as in (26). The correlation coefficient is $\rho_k = 1 - \theta/\sigma_k$, and scales are $T_{kod} = \Gamma(1 - \theta/\sigma_k) S_{kod}^{\theta/\sigma_k}$. Due to independence across k, the joint distribution of productivity across origins and characteristics is

$$\mathbb{P}[A_{kod}(v) \leq a_{ko} \quad \forall k = 1, \ldots, K, o = 1, \ldots, N] = \exp\left[\sum_{k=1}^{K}\left(\sum_{o=1}^{N}(T_{kod}a_{ko}^{-\theta})^{\frac{1}{1-\rho_k}}\right)^{1-\rho_k}\right].$$

The k-level expenditure shares in (27) become

$$\frac{X_{kod}}{X_d} = \left(\frac{P_{kod}}{P_{kd}}\right)^{-\sigma_k} \frac{P_{kd}^{-\theta}}{\sum_{k'=1}^{M} P_{k'd}^{-\theta}} X_d, \tag{29}$$

where $P_{kod} \equiv \gamma T_{kod}^{-1/\theta} W_o$ and $P_{kd} = \left(\sum_{o=1}^{N} P_{kod}^{-\sigma_k}\right)^{-1/\sigma_k}$. The first term on the right-hand side of (29) is expenditure within k and is CES with elasticity σ_k. The second term refers to between-k expenditure and is also CES. The aggregate demand system is cross-nested CES, as in (16), with each nest corresponding to one of the finite number of innovation characteristics.

In summary, Corollary 2 shows how to incorporate disaggregate data into our macroeconomic framework.

4.2 Aggregation

We next provide an aggregation result for recovering the country-level correlation function associated with the disaggregation in Corollary 2.

Corollary 3 (Aggregation). *Under the hypotheses and notation of Corollary 2, country-level productivity is*

$$A_{od}(v) = \max_{k=1,\ldots,K} A_{kod}(v),$$

which is multivariate θ-Fréchet with scale $T_{od} = \sum_{k=1}^{N} T_{kod}$ and correlation function

$$G^d(x_1, \ldots, x_N) = \sum_{k=1}^{K} G^{kd}\left(\omega_{k1d}x_1, \ldots, \omega_{kNd}x_N\right),$$

for aggregation weights $\omega_{kod} \equiv T_{kod}/T_{od}$.

Country-level productivity comes simply from adopting the most efficient innovation. Due to independence of productivity over characteristics, the macro scale parameters are the sum of the characteristic-level scales, and the country-level correlation function comes from aggregating within-characteristic correlation using relative scales as aggregation weights.

Suppose that we have an estimate of the within-k correlation function. How can we use Corollary 3 to recover country-level correlation-adjusted expenditure shares and the country-level correlation function in order to perform counterfactual analysis?

Aggregation weights, ω_{kod}, are key for the aggregation procedure. Because correlation-adjusted expenditure shares are proportional to scale parameters, these weights can be recovered by adjusting disaggregate expenditure shares for correlation.

The logic for computing correlation-adjusted trade shares at the micro level follows the derivations in Section 3.2. Gravity holds after correcting for correlation,

$$\pi^*_{kod} \equiv \frac{\pi_{kod}}{G_{kod}} = T_{kod}\left(\gamma \frac{W_o}{P_d}\right)^{-\theta}, \tag{30}$$

where $\pi_{kod} \equiv X_{kod}/X_d$ and $G_{kod} \equiv G^{kd}_o(T_{k1d}W_1^{-\theta}, \ldots, T_{kNd}W_N^{-\theta})$. We can then recover k-level correlation-adjusted expenditure shares by solving

$$\pi_{kod} = \pi^*_{kod} G^{kd}_o(\pi^*_{k1d}, \ldots, \pi^*_{kNd}).$$

As (30) shows, correlation-adjusted shares are proportional to scale parameters and $T_{od} = \sum_{k=1}^{K} T_{kod}$. Hence, the country-level correlation-adjusted shares and the aggregation weights satisfy

$$\pi^*_{od} = \sum_{k=1}^{K} \pi^*_{kod}, \quad \text{and} \quad \omega_{kod} = \frac{\pi^*_{kod}}{\pi^*_{od}}.$$

Aggregation weights equal the ratio of disaggregate to aggregate correlation-adjusted shares. In addition to being sufficient statistics for real import prices, these shares are sufficient statistics for the aggregation weights.

For example, in the CNCES case, $\pi^*_{kod} = X_{kod}^{1-\rho_k} X_{kd}^{\rho_k} / X_d$, which leads to

$$\pi^*_{od} = \sum_{k=1}^{K} \frac{X_{kod}^{1-\rho_k} X_{kd}^{\rho_k}}{X_d}, \quad \text{and} \quad \omega_{kod} = \frac{X_{kod}^{1-\rho_k} X_{kd}^{\rho_k}}{\sum_{k'=1}^{K} X_{k'od}^{1-\rho_{k'}} X_{k'd}^{\rho_{k'}}}.$$

Given an estimate of the k-level correlation coefficient, ρ_k, these quantities simply reflect observed trade flows; the country-level correlation function follows from (15).

Together, these results allow us to pass seamlessly between the micro and macro levels. Corollary 2 states that we can relate a given macro model with multivariate θ-Fréchet productivity to an underlying disaggregate model in which productivity also has a multivariate θ-Fréchet distribution. In turn, Corollary 3 shows how to recover the correlation-adjusted trade shares and correlation function of the macro model. The link comes from maximizing productivity across k within an origin country.

The natural consequence is that we can estimate the model using disaggregate data (e.g., sectoral data), and perform macro counterfactual analysis using the results in Section 3. Thanks to these results, we can connect (Ricardian) micro foundations to macro substitution patterns, as the applications in the next section show.

In summary, the key role of Theorem 1 is to tie cross-country correlation in productivity to potential micro-foundations. It characterizes correlation as reflecting the extent to which countries adopt similar innovations—with similar attributes. In turn, Corollary 2 and Corollary 3 guide the development of quantitative macro models based on disaggregate factors, and as such, they also guide the incorporation of disaggregate data into estimation.

5 Applications

We now present applications that extend the Ricardian model of trade in EK to multiple sectors (Caliendo and Parro, 2015), multinational production (Ramondo and Rodríguez-Clare, 2013), domestic geography (Ramondo et al., 2016), global value chains (Antràs and de Gortari, 2017), and intermediate inputs (Eaton and Kortum, 2002; Alvarez and Lucas, 2007). All of these models deliver a GEV factor demand system and satisfy the gross substitutes property.[21] These applications also provide concrete examples of the results in the previous section linking county-level correlation to micro characteristics.

5.1 Multiple Sectors

Assume that each country is composed of multiple sectors, $s = 1, \ldots, S$. Caliendo and Parro (2015) assume that consumers in destination d have Cobb-Douglas preferences so that sectoral shares X_{sd}/X_d are exogenous, and assume that productivity within each sector across origins is distributed independent Fréchet with shape θ_s and scale $\widetilde{A}_{so}$. Given trade costs τ_{sod}, the share of d's sector-s expenditure on goods from origin o is

$$\frac{X_{sod}}{X_{sd}} = \frac{\left(\tau_{sod}\frac{W_o}{A_{so}}\right)^{-\theta_s}}{\sum_{o'=1}^{N}\left(\tau_{so'd}\frac{W_{o'}}{A_{so'}}\right)^{-\theta_s}}, \tag{31}$$

where $A_{so} \equiv \widetilde{A}_{so}^{1/\theta}$. Due to the Cobb-Douglas assumption, sectoral expenditure shares are exogenous.

By assuming that productivity is correlated within each sector, we can generate a similar factor demand system, but with endogenous sectoral shares. Suppose that productivity $A_{sod}(v)$ for good v in sector s is a random vector drawn from a multivariate θ-Fréchet distribution with scale parameter T_{sod} and sector-level correlation function,

$$G^{sd}(x_1, \ldots, x_N) = \left(\sum_{o=1}^{N} x_o^{1/(1-\rho_s)}\right)^{1-\rho_s}, \tag{32}$$

The parameter ρ_s measures the degree of correlation across origin countries in each sector. Expenditure shares at the sector level are

$$\pi_{sod} = \left(\frac{P_{sod}}{P_{sd}}\right)^{-\frac{\theta}{1-\rho_s}} \left(\frac{P_{sd}}{P_d}\right)^{-\theta},$$

where $P_{sod} \equiv \gamma T_{sod}^{1/\theta} W_o$, $P_{sd} \equiv (\sum_{o=1}^{N} P_{sod}^{-\frac{\theta}{1-\rho_s}})^{-\frac{1-\rho_s}{\theta}}$, and P_d is the aggregate price

[21] A notable exception—with strong complementarities—is Fally and Sayre (2018). They build a model of trade in scarce and spatially concentrated commodities which implies an import demand system with very low elasticities of substitution. They estimate gains from trade that are much larger than ACR.

index in country d, $P_d = \left(\sum_s P_{sd}^{-\theta}\right)^{-1/\theta}$. This factor demand system matches (31) for $\theta/(1-\rho_s) = \theta_s$, and $P_{sod}/\gamma = \tau_{sod} W_o / A_{so}$. Sectoral shares depend on real sectoral prices with elasticity of substitution θ. As the parameter θ goes to zero, this sectoral CNCES model converges to match the Cobb-Douglas case in Caliendo and Parro (2015).[22,23]

Applying the results from Corollary 3, country-level productivity $A_{od}(v)$ is distributed multivariate θ-Fréchet with scale parameters given by $T_{od} = \sum_{s=1}^{S} T_{sod}$ and correlation function,

$$G^d(x_1, \cdots, x_N) = \sum_{s=1}^{S} \left(\sum_{o=1}^{N} (\omega_{sod} x_o)^{1/(1-\rho_s)} \right)^{1-\rho_s}, \tag{33}$$

with $\omega_{sod} = T_{sod}/T_{od}$. The inner sum indicates that when sectors are present in multiple countries, they induce correlation across origins. The parameter ω_{sod} measures the extent to which sector s matters for trade flows from o to d—it reflects sectoral trade costs and comparative advantage. In turn, aggregate expenditure shares constitute a GEV factor demand system, as in (16).

5.2 Multinational Production

Assume that productivity depends on the home country j of a firm. The micro correlation function is CNCES as in (32), and the implied macro correlation function is as in (33) with s replaced by j for both functions. The parameter ρ_j measures correlation across production locations for firms with home country j.

The expenditure share on goods produced in o for d by firms from j is

$$\pi_{jod} = \left(\frac{P_{jod}}{P_{jd}} \right)^{-\frac{\theta}{1-\rho_j}} \left(\gamma \frac{P_{jd}}{P_d} \right)^{-\theta}, \tag{34}$$

where $P_{jod} \equiv T_{jod}^{-1/\theta} W_o$, and $P_{jd} \equiv \left(\sum_{o=1}^{N} P_{jod}^{-\frac{\theta}{1-\rho_j}}\right)^{-\frac{1-\rho_j}{\theta}}$. The expenditure share on goods produced in o for d follows (16) with j replacing k. In this model, production locations can use technology from a common home country, which

[22] The Cobb-Douglas restriction, however, entails that key cross-price elasticities characterizing the macro demand system are not identified from between-sector variation, as pointed out by Adao et al. (2017). With CNCES, both θ and ρ_s can be identified from between- and within-sector variation, respectively.

[23] French (2016) uses CES expenditure shares across sectors, but he restricts the elasticities of substitution for each sector to be the same, $\rho_s = \rho$.

induces correlation. The factor demand system in (34) matches the one in Ramondo and Rodríguez-Clare (2013) for $\rho_j = \rho$ and $T_{jod}^{-1/\theta} = \tau_{od} h_{jo}/A_j$.

5.3 Multiple Regions

Assume that each country n is composed of R_n regions. Denote productivity in region r by $A_{rnd}(v)$. Productivity across regions within a country is symmetric multivariate θ-Fréchet with correlation parameter ρ_n and scale T_{rnd}, while productivity is independent across countries. The within-n correlation function is

$$G^{nd}(x_1, \ldots, x_{R_n}) = \left(\sum_{r=1}^{R_n} x_r^{1/(1-\rho_n)} \right)^{1-\rho_n}. \tag{35}$$

Workers are mobile across regions within a country and the country wage is W_n. For import price index $P_{rnd} = \gamma T_{rnd}^{-1/\theta} W_n$, the trade share from region r in n into destination d is

$$\pi_{rnd} = \left(\frac{P_{rnd}}{P_{nd}} \right)^{-\frac{\theta}{1-\rho_n}} \frac{P_{nd}^{-\theta}}{\sum_{n'} P_{n'd}^{-\theta}} \quad \text{with} \quad P_{nd} = \left(\sum_{r'=1}^{R_n} P_{r'nd}^{-\frac{\theta}{1-\rho_n}} \right)^{-\frac{1-\rho_n}{\theta}}. \tag{36}$$

The first fraction on the right-hand side of (36) is the probability of importing from region r in country n conditional on importing from some region in country n, while the second fraction is the probability of importing from country n into d.

Because regions are unique to countries, country-level productivity—which is just the maximum across regions within each country—is independent with scale $T_{nd} = \left(\sum_{r=1}^{R_n} T_{rnd}^{1/(1-\rho_n)} \right)^{1-\rho_n}$. In turn, the country-level factor demand system is CES,

$$\pi_{nd} = \sum_{r=1}^{R_n} \pi_{rnd} = \frac{T_{nd} W_n^{-\theta}}{\sum_{n'} T_{n'd} W_{n'}^{-\theta}}.$$

By assuming that $\rho_n = 0$, for all $n = 1, \ldots, N$, this case matches the one in Ramondo et al. (2016) .

5.4 Global Value Chains

We now show that the model of global value chains in Antràs and de Gortari (2017) generates a GEV demand system. That is, it has the same macroeconomic implications as a model without global value chains, but in which productivity follows a multivariate θ-Fréchet distribution with an appropriately chosen correlation function.

Assume that production is done in K stages, $k = 1, \ldots, K$, where $k = K$ is the final stage of production (e.g., assembly), takes the Cobb-Douglas form, and labor is the only factor of production. Let $\ell = [\ell(1), \ldots, \ell(K)]$ index a path of locations across production stages.

The unit cost of the input bundle used for goods produced following the production path ℓ is given by

$$c_\ell = W_{\ell(K)} \prod_{k=1}^{K-1} \left(\frac{W_{\ell(k)}}{W_{\ell(K)}} \right)^{\alpha_k},$$

with $\alpha_k > 0$ and $\sum_{k=1}^{K-1} \alpha_k < 1$. The unit cost of good v is $c_\ell / A_{\ell d}(v)$. The variable $A_{\ell d}(v)$ denotes the marginal product of the input bundle when good v is produced along ℓ and delivered to d. This variable is distributed independent θ-Fréchet across ℓ with scale $T_{\ell d}$. The likelihood of a particular production path ℓ destined to country d is given by

$$\pi_{\ell d} = \frac{T_{\ell d} c_\ell^{-\theta}}{\sum_{\ell'} T_{\ell' d} c_{\ell'}^{-\theta}}. \tag{37}$$

This factor demand share matches the one in Antràs and de Gortari (2017) for $T_{\ell d} = \tau_{\ell(K),d}^{-\theta} T_{\ell(K)}^{1-\sum_{k=1}^{K-1}\alpha_k} \prod_{k=1}^{K-1} (\tau_{\ell(k),\ell(k+1)})^{-\theta\alpha_k} T_{\ell(k)}^{\alpha_k}$ where τ_{ij} is an iceberg cost of transporting goods from country i to country j, and T_i is a productivity index for country i. Aggregate trade shares from country o to d are obtained by summing $\pi_{\ell d}$ over production paths with last production stage in country o—i.e., $\ell(K) = o$.

A macro model where productivity is multivariate θ-Fréchet with scale $T_{\ell d}$ and correlation function given by

$$G^d(x_1, \cdots, x_N) = \sum_\ell x_{\ell(K)} \prod_{k=1}^{K-1} \left(\frac{x_{\ell(k)}}{x_{\ell(K)}} \right)^{\alpha_k},$$

implies a factor demand system equivalent to the the one in the model with global value chains.

5.5 Intermediate Inputs

Suppose that each variety is used to produce an aggregate intermediate input. In turn, firms produce varieties using a Cobb-Douglas production function in labor and this good. Hence, the unit cost of the input bundle in country o is

$$c_o = BW_o^{\beta} P_o^{1-\beta},$$

where $0 < \beta \leq 1$ is the share of labor in production, B is a positive constant, and P_o is the price level in o. If productivity is independent Fréchet with scale A_o and trade costs are τ_{od}, expenditure shares are

$$\pi_{od} = \frac{A_o \tau_{od}^{-\theta} c_o^{-\theta}}{\sum_{o'} A_{o'} \tau_{o'd}^{-\theta} c_{o'}^{-\theta}}. \tag{38}$$

The price index in country d is defined implicitly by

$$P_d = \gamma B \left(\sum_o A_o \tau_{od}^{-\theta} (W_o^{\beta} P_o^{1-\beta})^{-\theta} \right)^{-\frac{1}{\theta}}.$$

It is easy to see that a model without this input-output loop, but with a multivariate θ-Fréchet distribution of productivity with scale $T_{od} \equiv (A_o/\tau_{od})^{\theta/\beta}$ and correlation functions for each country implicitly defined by the system

$$G^d(x_1, \ldots, x_N) = \sum_{o=1}^{N} x_o^{\beta} G^o(x_1, \ldots, x_N)^{1-\beta},$$

generates the same factor demand system as the model with intermediate inputs.

6 Quantitative analysis

This section quantifies the gains from trade when productivity is correlated across space. We consider a sectoral version of the Ricardian model of trade in Section 2, which includes the cross-nested CES sectoral model in Section 5.1 as a special case. The novelty of this quantitative model is that it captures Ricardo's insight that the degree of technological similarity determines the gains from trade. Based exclusively on supply factors, this sectoral model allows for bilateral heterogeneity in trade elasticities based on observable variables (e.g. proximity).[24] The choice of this multi-sector model allows us to highlight the importance of using disaggregate data to estimate elasticities that are key for macro counterfactuals—such as the gains from trade.

6.1 Specification

Denote each sector by $s = 1, \ldots, S$. We assume that within-sector productivity is multivariate θ-Fréchet with scale T_{sod} and correlation function defined implicitly, following Hanoch (1975) and Sato (1977), by

$$1 = \sum_{o=1}^{N} \left(\frac{x_{so}}{G^{sd}(x_{s1}, \ldots, x_{sN})} \right)^{\frac{1}{1-\rho_{sod}}}. \tag{39}$$

The parameters, $\{\rho_{sod}\}_{o=1}^{N}$, capture heterogeneity in correlation across origins within sector s and destination d. The correlation function satisfying (39) results in iso-elastic sectoral expenditure shares,

$$\pi_{sod} \equiv \frac{X_{sod}}{X_d} = \frac{\sigma_{sod} \left(\frac{P_{sod}}{P_{sd}} \right)^{-\sigma_{sod}}}{\sum_{o'=1}^{N} \sigma_{so'd} \left(\frac{P_{so'd}}{P_{sd}} \right)^{-\sigma_{so'd}}} \left(\frac{P_{sd}}{P_d} \right)^{-\theta}, \tag{40}$$

where $\sigma_{sod} \equiv \frac{\theta}{1-\rho_{sod}}$. The sectoral price index, P_{sd}, is defined implicitly as

$$1 = \sum_{o} \left(\frac{P_{sod}}{P_{sd}} \right)^{-\sigma_{sod}}, \tag{41}$$

with $P_{sod} \equiv \gamma T_{sod}^{-1/\theta} W_o$. When $\rho_{sod} = 0$, productivity is independent across sectors and origins, and we get the CES sectoral model with $\sigma_{sod} = \theta$. When $\rho_{sod} = \rho_s$, we get the sectoral CNCES model without spatial correlation, in Section 5.1.

Next, we compute correlation-adjusted sectoral expenditure shares and the gains

[24] As we show in Online Appendix O.2.1, one can allow for complementarity or substitutability across sectors from preferences. In fact, the sectoral model we present below is isomorphic to a model in which preferences across sectors are CES, $u(C_{1d}, \ldots, C_{Sd}) = \left(\sum_{s=1}^{S} C_{sd}^{\frac{\epsilon}{\epsilon+1}} \right)^{\frac{\epsilon+1}{\epsilon}}$, with C_{sd} an aggregate of sectoral products, and with the parameter $\epsilon > -1$ the expenditure elasticity of substitution between sectors. Sectoral goods are substitutes as long as $\epsilon > 0$ so that ϵ takes exactly the role of θ below. Sectoral goods are complements if $\epsilon < 0$, a possibility only allowed if the model had a preferences interpretation. Our estimates suggest substitution, not complementarity, across sectors—that is, our estimates of the parameter θ are positive.

from trade in closed form. As we explain in Section 4.2, because disaggregate adjusted shares are sufficient statistics for real import prices, we can calculate the correlation correction by inverting the demand system. Dividing (40) by σ_{sod}, summing over origins, and using (41) gives $(P_{sod}/P_{sd})^{-\sigma_{sod}}$. Further using sectoral shares, we get

$$\pi^*_{sod} = \left(\frac{X_{sod}/\sigma_{sod}}{\sum_{o'=1}^{N} X_{so'd}/\sigma_{so'd}} \right)^{\theta/\sigma_{sod}} \frac{X_{sd}}{X_d}, \tag{42}$$

where the first term on the right-hand side is $(P_{sod}/P_{sd})^{-\theta}$, and the second terms is $(P_{sd}/P_d)^{-\theta}$. Using the results in Section 4.2, $\pi^*_{od} = \sum_s \pi^*_{sod}$, and further applying (21), the gains from trade are

$$GT_d = \left(\sum_s \pi^*_{sod} \right)^{-1/\theta}. \tag{43}$$

Once we have estimates of σ_{sod} and θ, we can compute the gains from trade using sectoral expenditure data. We turn next to the estimation of these parameters.

6.2 Estimation

We estimate our quantitative model using two sequential gravity regressions. The first step uses variation in trade flows and tariffs across origin countries within each sector to identify bilateral sectoral trade elasticities. The second step uses variation across sectors and destination markets to identify the shape parameter θ.

Letting t index years, we impose additional assumptions on the structure of trade costs and spatial correlation patterns. First, using the decomposition of scale parameters into productivity and trade cost indices in Section 2.2, we can re-write sectoral import prices indices as $P_{sodt} = \gamma \tau_{sodt} W_{ot}/A_{sot}$. Second, we assume that trade costs depend on gravity covariates and tariffs,

$$\ln \tau_{sodt} = \delta_s' \text{Geo}_{od} + \ln(1 + t_{sodt}) + \varepsilon^1_{sdt} + \varepsilon^2_{sodt}. \tag{44}$$

The variable t_{sodt} is an ad-valorem effective tariff on sector-s goods shipped from o to d at time t. Geo_{od} includes variables such as distance and time differences between trading partners, as well as dummies indicating whether the two countries share a border, language, and legal origins. We further allow for sector-specific coefficients, δ_s. The variable ε^1_{sdt} captures unobserved sector-destination-year components of trade costs, while ε^2_{sodt} captures additional unobserved components of trade costs across origin countries. Finally, we assume that elasticities have a sector component and a spatial component proxied by a non-linear function of bilateral distance,[25]

$$\sigma_{sod} = \bar{\sigma}_s + \tilde{\sigma}_1 \text{Dist}_{od} + \tilde{\sigma}_2 \text{Dist}^2_{od}. \tag{45}$$

Using (40), we get our first-step within-sector gravity equation,

$$\ln \pi_{sodt} = S_{sot} + D_{sdt} + B_{sod} - \bar{\sigma}_s \ln(1 + t_{sodt}) + (\alpha_{sot} + \beta_{sdt} - \ln(1 + t_{sodt})) \left(\tilde{\sigma}_1 \text{Dist}_{od} + \tilde{\sigma}_2 \text{Dist}^2_{od}\right) + u_{sodt}, \tag{46}$$

where $S_{sot} \equiv \bar{\sigma}_s \alpha_{sot}$, $D_{sdt} \equiv \bar{\sigma}_s \beta_{sdt} - \bar{\sigma}_s \varepsilon^1_{sdt} - \ln \sum_{o'} \sigma_{so'd} \left(P_{so'dt}/P_{sdt}\right)^{-\sigma_{so'd}}$, $B_{sod} \equiv \ln \sigma_{sod} - \sigma_{sod} \delta'_s \text{Geo}_{od}$, $\alpha_{sot} \equiv \ln(W_{ot}/A_{sot})$, $\beta_{sdt} \equiv \ln(P_{sdt}/\gamma)$, and $u_{sodt} = -\sigma_{sod} \varepsilon^2_{sodt}$. The coefficients from the interaction of tariffs with distance and distance squared allow us to estimate distance-dependent elasticities of substitution. The exclusion restriction for identification is that variation in tariffs across origin countries is exogenous conditional on the covariates included in (46).

To estimate the between-sector elasticity, θ, we use a second-step regression that relies on variation across sectors and inferred within-sector relative prices from the first step. This second regression comes from destination d's expenditure on sector-s goods, $X_{sdt}/X_{dt} = (P_{sdt}/P_{dt})^{-\theta}$, where we can write

$$\ln \frac{P_{sdt}}{P_{dt}} = \ln \frac{P_{sdt}}{P_{sodt}} + \ln \frac{P_{sodt}}{P_{dt}} = \ln \gamma - \ln \frac{P_{sodt}}{P_{sdt}} + \ln \frac{W_{ot}}{A_{sot}} - \ln P_{dt} + \ln \tau_{sodt}.$$

Given estimates of σ_{sod} from our first step, $\hat{\sigma}_{sod}$, we combine (40) with the definition of the sectoral price index in (41) to get an estimate of P_{sodt}/P_{sdt},

$$\widehat{\frac{P_{sodt}}{P_{sdt}}} = \left(\frac{\pi_{sodt}/\hat{\sigma}_{sod}}{\sum_{o'=1}^{N} \pi_{so'dt}/\hat{\sigma}_{so'd}} \right)^{-\frac{1}{\hat{\sigma}_{sod}}}.$$

We estimate the parameter θ from the coefficient on $\ln(1 + t_{sodt})$ in the following

[25] This assumption is motivated by the literature that documents that technology diffusion follows a spatial pattern. Keller (2002) estimates that a 1,200-kilometer increase in distance leads to a 50 percent drop in technology diffusion. Similarly, Bottazzi and Peri (2003), using patent data, find a strong geographic decay in technology diffusion between European regions. Comin et al. (2013) document that the lower the spatial distance to another country's technology, the higher the rate of adoption. Relatedly, Keller and Yeaple (2013) link the gravity patterns observed in flows of firms across countries to multinational firms transferring knowledge from their parent firm to their affiliates abroad, with this transfer being easier to nearby locations.

regression,

$$\ln \frac{X_{sdt}}{X_{dt}} = a_{sot} + b_{dt} + \theta \ln \left(\widehat{P_{sodt}/P_{sdt}} \right) - \theta \delta_s' \text{Geo}_{od} - \theta \ln(1 + t_{sodt}) + v_{sodt}, \tag{47}$$

where $a_{sot} \equiv \theta \ln(A_{sot}/W_{ot})$, $b_{dt} \equiv \theta \ln P_{dt}$, and $v_{sodt} \equiv -\theta(\varepsilon^1_{sdt} + \varepsilon^2_{sodt})$. Identification comes from controlling for within-sector relative prices using our first-step estimates. The identification assumption is that the unobserved component of trade costs is orthogonal to tariffs conditional on the other covariates.

We estimate (46) and (47) by Ordinary Least Squares (OLS), using sectoral tariff data constructed by aggregating 4-digit SITC tariff data, from COMTRADE, and sectoral trade flow data, from the World Input-Output Database (WIOD), for 1996-2007. Appendix H describes the data construction and sample restrictions in detail.

We consider three cases. First, we estimate the CES model where productivity is independent across sectors and origin countries, $\rho_{sod} = 0$, and get an estimate of θ from our first step (since $\sigma_{sod} = \theta$). Second, we estimate the CNCES model where correlation is only sector specific, but common across o and d, $\rho_{sod} = \rho_s$, so that $\sigma_{sod} = \bar{\sigma}_s$. Finally, we allow for bilateral correlation within sector and estimate our model with σ_{sod} specified in (45).

Appendix Figure I.1 presents OLS estimates of the elasticity of substitution, σ_{sod}, as a function of geographical distance. The spatial pattern that emerges is clear: Substitutability decreases with distance, indicating that productivity is less correlated between countries that are further away from each other.[26] Additionally, Appendix Table I.1 presents our OLS estimates of θ for each of the three specifications.

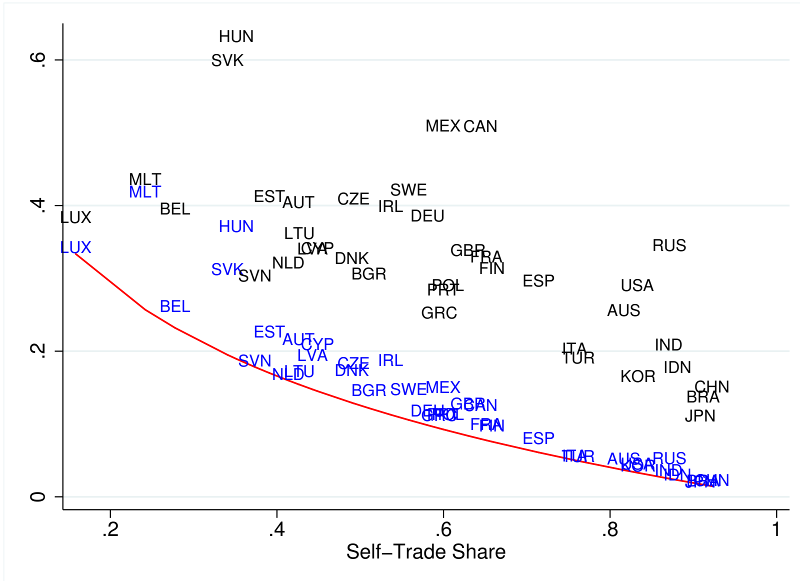
We next use these estimates to perform various counterfactual exercises.

6.3 The Gains from Trade

Figure 1 shows the gains from trade calculated using the model with no correlation (CES), with within-sector correlation (CNCES), and with bilateral distance-dependent correlation, respectively. The figure shows that differences are large across countries with similar self-trade shares. For instance, Mexico and Germany have a similar self-trade share of around 60 percent. However, once we account for spatial heterogeneity

[26] Appendix Figure I.2 shows estimates for the within-sector elasticity of substitution in the CNCES model with $\sigma_{sod} = \bar{\sigma}_s$.

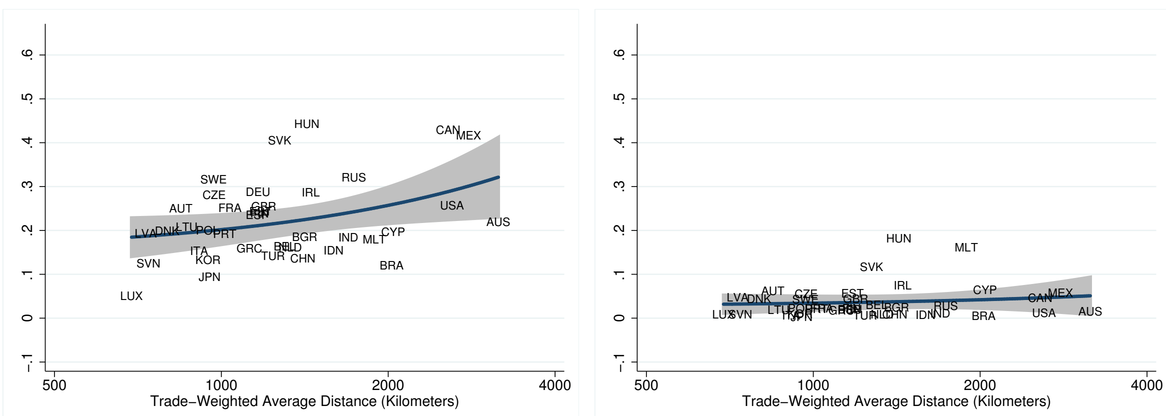
Figure 1: Gains from Trade and Self-Trade Share, 2007.



Notes: Black data: $GT_d^{spatial}$. Blue data: GT_d^{CNCES}. Red line: GT_d^{CES}.

Figure 2: Gains from Trade and Distance, 2007. Percent differences from CES.

(a) Spatial (b) CNCES



Notes: (2a) Black data: $100 \times (GT_d^{CNCES} - GT_d^{CES})/GT_d^{CES}$. (2b): Black data: $100 \times (GT_d^{spatial} - GT_d^{CES})/GT_d^{CES}$. Blue line: log-linear fit. Grey band: 95% point-wise confidence interval.

in correlation, their gains are very different. In contrast, the CES model would predict equal gains from trade for the two countries, while the CNCES model delivers gains that are only slightly different.

Perhaps not surprisingly, Figure 2 shows that the model with bilateral correlation implies that countries that are on average further away from their trading partners have higher gains from trade, while the CNCES model—with sectoral elasticities that are constant over space—fails to do that. Our model interprets countries that are further away from each other—with lower sectoral elasticities—as having more dissimilar technologies and, hence, higher gains from trade. This result captures Ricardo's second insight that gains from trade are higher when countries trade with technological dissimilar partners.

6.4 NAFTA, the Rise of Chinese Imports, and U.S. Protectionism

Next, we consider the implications of our quantitative model on various counterfactual scenarios, and compare them with the implications from the CNCES model. We use the procedure outlined in Section 3.

We first consider a scenario in which the United States increases trade costs with Mexico and Canada simultaneously by x percent, with $x \in [5, 50]$. Figure 3 shows the implications for real wages for the model with distance-dependent correlation and the CNCES model. The difference between the two models comes from adding bilateral correlation in addition to within-sector correlation. Differences in the predicted real wages can be large, particularly for large changes in trade costs.

Correlation matters—and more so for large changes in trade costs—through two potentially offsetting effects that shape the gains from trade: a price effect and a wage effect. The price effect is direct: Increasing trade costs on Mexican goods increases prices for U.S. consumers. The size of this effect is just the elasticity of the price in the United States to the price of imports from Mexico, which equals the expenditure share.[27] In contrast, the wage effect is indirect and operates through the market clearing condition for the United States,

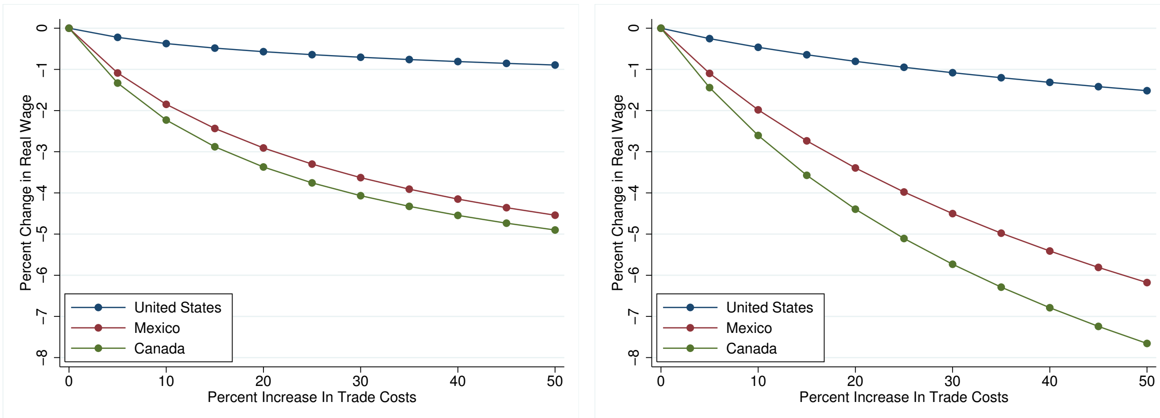
$$W_{USA}L_{USA} = \sum_{d=1}^{N} \frac{P_{USA,d}^{-\theta} G_{USA}^{d}(P_{1d}^{-\theta}, \dots, P_{Nd}^{-\theta})}{G^{d}(P_{1d}^{-\theta}, \dots, P_{Nd}^{-\theta})} X_d.$$

If the change in trade costs with Mexico induces U.S. consumers to substitute expenditure away from Mexican goods and towards U.S. goods, labor demand would increase and U.S. wages would increase. How rapidly expenditure shifts away from Mexican goods and towards U.S. goods depends on the correlation in technology. If the United States and Mexico were very close neighbors (high correlation), then U.S. and Mexican goods would be substitutable and labor demand

27 $\frac{\partial \ln P_{USA}}{\partial \ln \tau_{MEX,USA}} = \frac{\partial \ln G^{USA}(P_{1,USA}^{-\theta}, \dots, P_{N,USA}^{-\theta})^{-\frac{1}{\theta}}}{\partial \ln P_{MEX,USA}} = \frac{P_{MEX,USA}^{-\theta} G_{MEX,USA}}{P_{USA}^{-\theta}} = \pi_{MEX,USA}.$

Figure 3: Effects of NAFTA Reversal on Real Wages.

(a) Cross-nested CES (CNCES) Model (b) Spatial Model



Notes: Effects of an unilateral increase in trade costs for goods from Canada and Mexico into the United States.

in the United States would be sensitive to changes in trade costs for imports from Mexico. The wage effect would be large and would offset the losses coming from increasing prices in the United States. In contrast, if trade between (most parts of) Mexico and the United States occurs over long distances, accounting for spatial correlation would reduce the substitutability between their goods. As a result, the offsetting wage effect would be smaller in the model with heterogenous spatial correlation. As we see in Figure 3, accounting for spatial correlation in productivity leads to an increase in the losses from increasing U.S. trade costs with NAFTA partners. The gap between the models depends on the size of the increase in trade costs. For small changes in trade costs, the gap is small because general equilibrium effects on wages are small, while the direct price effect—with an elasticity equal to the expenditure share—is the same between models.

Our second counterfactual considers the implications of the rise of Chinese manufacturing imports for the U.S. real wage. We compute the change in the U.S. real wage in each year between the observed outcome (i.e., the real wage implied by each model given the data) and a scenario in which we fix China's trade costs at the level of 2003 for the sector "Machinery, Equipment, and Manufacturing n.e.c.". We choose this sector because expenditure by the United States on manufacturing goods from China increased threefold between 2003 and 2007 (see Appendix Figure I.3). Additionally, this sector's implied trade costs decrease sharply in the spatial model, but they are relatively stable for the CNCES model, as shown in Figure 4a. For the remaining sectors, the implied trade costs between the two models are similar (not shown).[28] Figure 4b shows the difference in real wages between the actual and counterfactual scenario: The sharp decrease in trade costs implied by the spatial model from 2003 on translates into large increases in the U.S. real wage. The CNCES model fails to capture the collapse in trade costs within this sector, and, as a consequence, suggests lower gains for the United States from the rise of manufacturing imports from China.

Our final counterfactual considers a series of trade protection exercises where the United States unilaterally increases trade costs by five percent, one trading partner at the time. We compute the counterfactual change in real wages for both the spatial model and the CNCES model. The presence of bilateral correlation changes the rankings of the countries that provoke the largest change in U.S. real wages. While the CNCES model implies that real wages in the United States would decrease the most with increases in trade costs from Canada, the spatial model predicts that China would have the largest impact. The intuition behind these results is similar to the intuition for other counterfactuals: The direct price effect is measured by observed expenditure shares, while the indirect wage effect depends on substitutability of goods between trading partners and, therefore, on the patterns of spatial correlation. Summing up, these counterfactual exercises illustrate that the welfare implications of a trade shock can change substantially once we account for richer patterns of heterogeneity in correlation of productivity. These results reflect Ricardo's insight that differences in technological similarity across trade partners matter for the gains from trade.

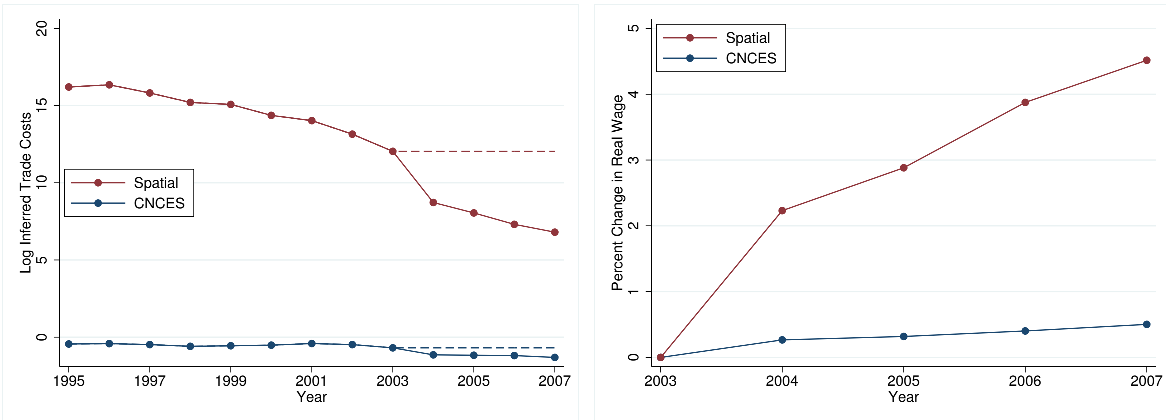
7 Conclusions

This paper is motivated by the old Ricardian idea that a country gains from trading with those countries who are technologically dissimilar. We develop a Ricardian theory of trade that allows for rich patterns of correlation in technology between countries yet retains all the tractability of EK-type tools. Importantly, our structure generates the class of GEV factor demand systems and, as such, approximates *any*

[28]Trade costs are calculated using the ratio $(\pi^*_{sodt}/\pi^*_{soot})^{-1/\theta} = \tau_{sodt}P_{ot}/P_{dt}$, for each s and t. Since we can only uncover within-destination relative prices from our expenditure data, we use GDP deflators from the Penn World Tables (9.0) to adjust this quantity by country price levels and recover sectoral trade costs.

Figure 4: The Rise of Chinese Manufacturing Imports.

(a) Implied trade costs $\tau_{CHN,USA}$, S10 (b) Implied change in U.S. real wage



Notes: S10 refers to the sector "Machinery, Equipment, and Manufacturing n.e.c.". (4a) shows trade costs in S10, implied by the spatial and CNCES model, respectively—dash lines show the counterfactual levels of trade costs. (4b) shows the percent changes in the U.S. real wage, implied by the spatial and CNCES model, respectively.

Ricardian model. The gains from trade coming from a GEV factor demand system can be written as a simple correction to self-trade shares.

We provide a structure for technology, based on the adoption of innovations, that is necessary and sufficient to generate max-stable multivariate Fréchet productivity with a general dependence structure. Moreover, the theory, by relating macro substitutability patterns to underlying factors, provides guidance on incorporating standard micro estimates into macro counterfactual exercises.

Our quantitative application to a multi-sector trade model reveals that differences in correlation across countries matter: Gains are much more heterogeneous across countries than the case of independent productivity. These results suggest that our framework has the potential to change quantitative conclusions in any literature applying Fréchet tools.

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.

Ahlfeldt, G., S. Redding, D. Sturm, and N. Wolf (2015). The economics of density: Evidence from the berlin wall. *Econometrica 83*(6), 2127–2189.

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

Alviarez, V. (2018). Multinational production and comparative advantage. *Manuscript UBC*.

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review 69*(1), 106–116.

Antràs, P. and A. de Gortari (2017). On the geography of global value chains. *Mimeo, Harvard University*.

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodriguez-Clare (2017). The elusive pro-competitive effects of trade. *Review of Economic Studies, forthcoming*.

Arkolakis, C., A. Costinot, and A. Rodríguez-clare (2012). New trade models, same old gains? *The American Economic Review*, 94–130.

Bas, M., T. Mayer, and M. Thoenig (2017). From micro to macro: Demand, supply, and heterogeneity in the trade elasticity. *Journal of International Economics 108*, 1–21.

Berry, S., A. Gandhi, and P. Haile (2013). Connected substitutes and invertibility of demand. *Econometrica 81*(5), 2087–2111.

Bertoletti, P., F. Etro, and I. Simonovska (2017). International trade with indirect additivity. *American Economic Journal: Microeconomics, forthcoming*.

Bottazzi, L. and G. Peri (2003). Innovation, demand, and knowledge spillovers: Evidence from european patent data. *European Economic Review 47*, 687–710.

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

Bryan, G. and M. Morten (2018). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy, forthcoming*.

Buera, F. J. and E. Oberfield (2016). The global diffusion of ideas. NBER Working Paper Series *21844*.

Caliendo, L., F. P. annd Esteban Rossi-Hansberg, and P.-D. Sarte (2017). The impact of regional and sectoral productivity changes on the u.s. economy. *Review of Economic Studies, forthcoming*.

Caliendo, L. and F. Parro (2015). Estimates of the trade and welfare effects of nafta. *The Review of Economic Studies 82*(1), 1–44.

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

Comin, D., M. Dmitriev, and E. Rossi-Hansberg (2013). The spatial diffusion of technology. *Unpublished*.

Costinot, A., D. Donaldson, and I. Komunjer (2012). What goods do countries trade? a quantitative exploration of ricardo's ideas. *Review of Economic Studies 79*, 581–608.

Costinot, A. and A. Rodrìguez-Clare (2014). Trade theory with numbers: Quantifying the consequences of globalization. *Handbook of International Economics, Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff eds. 4*(4), 197–261.

Dagsvik, J. K. (1995). How large is the class of generalized extreme value random utility models? *Journal of Mathematical Psychology 39*(1), 90–98.

De Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability 12*(4), 1194–1204.

DeLoecker, J., P. Goldberg, A. Khandelwal, and N. Pavcnik (2016). Prices, markups and trade reform. *Econometrica 84*(2), 445Ð510.

DiGiovanni, J., A. Levchenko, and J. Zhang (2014). The global welfare impact of china: Trade integration and technological change. *American Economic Journal: Macroeconomics 6*(3), 153–183.

Dornbusch, R., S. Fisher, and P. Samuelson (1977). Comparative advantage, trade and payments in a ricardian model with a continuum of goods. *American Economic Review 67*, 823–839.

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.

Eaton, J. and S. Kortum (2010). Technology in the global economy: A framework for quantitative analysis. *Unpublished manuscript*.

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

Fajgelbaum, P. and S. Redding (2014). External integration, structural transformation and economic development: Evidence from argentina 1870-1914. *NBER Working Paper 20217*.

Fally, T. and J. Sayre (2018). Commodity trade matters. *Mimeo UC Berkeley*.

Feenstra, R. C., P. Luck, M. Obstfeld, and K. N. Russ (2017). In search of the armington elasticity. *The Review of Economics and Statistics 0*(0), null.

Feenstra, R. C. and D. E. Weinstein (2017). Globalization, markups, and the U.S. price level. *Journal of Political Economy 125*(4), 1040–1074.

French, S. (2016). The composition of trade flows and the aggregate effects of trade barriers. *Journal of International Economics 98*, 114–137.

Hanoch, G. (1975). Production and demand models with direct or indirect implicit additivity. *Econometrica: Journal of the Econometric Society*, 395–419.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2013). The allocation of talent and u.s. economic growth. Technical report, National Bureau of Economic Research.

Kabluchko, Z. (2009). Spectral representations of sum-and max-stable processes. *Extremes 12*(4), 401.

Keller, W. (2002). Geographic localization of international technology diffusion. *American Economic Review 92*, 120–42.

Keller, W. and S. R. Yeaple (2013). The gravity of knowledge. *American Economic Review 103*(4), 1414.

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

Kortum, S. (1997). Research, patenting, and technological change. *Econometrica 65*(6).

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *The American Economic Review 70*(5), 950–959.

Lagakos, D. and M. E. Waugh (2013). Selection, agriculture, and cross-country productivity differences. *The American Economic Review 103*(2), 948–980.

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

Lashkaripour, A. and V. Lugovskyy (2017). Industry-level scale economies: from micro-estimation to macro-implications. *Mimeo, Indiana University*.

Levchenko, A. and J. Zhang (2014). Ricardian productivity differences and the gains from trade. *European Economic Review 65*, 45–65.

Levchenko, A. and J. Zhang (2016). The evolution of comparative advantage: Measurement and welfare implications. *Journal of Monetary Economics 78*, 96–111.

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

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695–1725.

Monte, F., S. Redding, and E. Rossi-Hansberg (2015). Commuting, migration and local employment elasticities. *NBER Working Paper 21706*.

Ossa, R. (2015). Why trade matters after all. *Journal of International Economics 97*(2), 266–277.

Penrose, M. D. (1992). Semi-min-stable processes. *The Annals of Probability 20*(3), 1450–1463.

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

Ramondo, N., A. Rodríguez-Clare, and M. Saborío-Rodríguez (2016). Trade, domestic frictions, and scale effects. *American Economic Review 106*(10), 3159–84.

Redding, S. (2016). Goods trade, factor mobility and welfare. *Journal of International Economics 101*, 148–167.

Redding, S. and E. Rossi-Hansberg (2017). Quantitative spatial economics. *Annual Review of Economics 9*, 21–58.

Ricardo, D. (1817). *On the Principles of Political Economy and Taxation* (1 ed.). John Murray, London.

Rudin, W. et al. (1964). *Principles of mathematical analysis*, Volume 3. McGraw-hill New York.

Sato, R. (1977). Homothetic and non-homothetic CES production functions. *The American Economic Review 67*(4), 559–569.

Scarf, H. and C. A. Wilson (2005). Uniqueness of equilibrium in the multi-country ricardo model. *in Frontiers in Applied General Equilibrium Modeling, ed. T. J. Kehoe, T. N. Srinivasan, and J. Whalley*, 24–44.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes 5*(1), 33–44.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations* (1 ed.). W. Strahan, London.

Smith, T. E. (1984). A choice probability characterization of generalized extreme value models. *Applied Mathematics and Computation 14*(1), 35–62.

Wilson, C. A. (1980). On the general structure of ricardian models with a continuum of goods: Applications to growth, tariff theory, and technical change. *Econometrica 48*(7), 1675–1702.

A Properties of Fréchet Random Variables

Lemma A.1. *Let X be distributed Fréchet with scale $A > 0$ and shape $\alpha > 0$. Then if $\alpha > 1$, $\mathbb{E}[X] = \Gamma(1 - 1/\alpha)A^{1/\alpha}$. Also, for any $B > 0$ and $\beta > 0$, BX^β is Fréchet with scale $AB^{\alpha/\beta}$, shape α/β, and $\mathbb{E}[BX^\beta] = \Gamma(1 - \beta/\alpha)BA^{\beta/\alpha}$.*

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty z\frac{\partial}{\partial z}\mathbb{P}\left[X \leq z\right]\mathrm{d}z = \int_0^\infty z\frac{\partial}{\partial z}e^{-Az^{-\alpha}}\mathrm{d}z \\
&= \int_0^\infty ze^{-Az^{-\alpha}}\alpha Az^{-\alpha-1}\mathrm{d}z = \int_0^\infty t^{-1/\alpha}e^{-t}\mathrm{d}tA^{1/\alpha} = \Gamma(1-1/\alpha)A^{1/\alpha},
\end{aligned}$$

and

$$\mathbb{P}[BX^\beta \leq z] = \mathbb{P}[X \leq (z/B)^{1/\beta}] = e^{-A(z/B)^{-\alpha/\beta}} = e^{-AB^{\alpha/\beta}z^{-\alpha/\beta}}.$$

The previous result implies that $\mathbb{E}[BX^\beta] = \Gamma(1 - \beta/\alpha)BA^{\beta/\alpha}$. □

Lemma A.2. *Let $\{X_i\}_{i=1}^N$ be α-Fréchet with scale parameters $\{A_i\}_{i=1}^N$ and correlation function $G : \mathbb{R}_+^N \to \mathbb{R}_+$. Then, for any $B_i \geq 0$ $i = 1, \ldots, N$ and $\beta > 0$, the random vector $\{B_iX_i^\beta\}_{i=1}^N$ is α/β-Fréchet with scale parameters of $\{A_iB_i^{\alpha/\beta}\}_{i=1}^N$ and correlation function G.*

Proof.

$$\begin{aligned}
\mathbb{P}[B_iX_i^\beta \leq y_i, i = 1, \ldots, N] &= \mathbb{P}[X_i \leq (y_i/B_i)^{1/\beta}, i = 1, \ldots, N] \\
&= \exp\left[-G(A_1(y_1/B_1)^{-\alpha/\beta}, \ldots, A_N(y_N/B_N)^{-\alpha/\beta})\right] \\
&= \exp\left[-G(A_1B_1^{\alpha/\beta}y_1^{-\alpha/\beta}, \ldots, A_NB_N^{\alpha/\beta}y_N^{-\alpha/\beta})\right].
\end{aligned}$$

□

Lemma A.3. *Let $\{X_i\}_{i=1}^N$ be θ-Fréchet with scale parameters $\{T_i\}_{i=1}^N$ and correlation function $G : \mathbb{R}_+^N \to \mathbb{R}_+$. Then, the random variable $\max_{i=1,\ldots,N} X_i$ is θ-Fréchet with scale $G(T_1, \ldots, T_N)$. Moreover, let $\{\mathcal{I}_j\}_{j=1}^M$ be any partition of $\{1, \ldots, N\}$ and define the random variable $\{Y_1, \ldots, Y_M\}$ as*

$$Y_j = \max_{i\in\mathcal{I}_j} X_i.$$

Let $j : \{1, \ldots, N\} \to \{1, \ldots, M\}$ be the unique mapping such that $j = j(i)$ if and only if $i \in \mathcal{I}_j$. Define $\tilde{T}_j = G(T_1 \mathbf{1}\{1 \in \mathcal{I}_j\}, \ldots, T_N \mathbf{1}\{N \in \mathcal{I}_j\})$ and $\omega_i = \frac{T_i}{\tilde{T}_j} \mathbf{1}\{i \in \mathcal{I}_j\}$. Then,

1. *$\{Y_1, \ldots, Y_M\}$ is θ-Fréchet with correlation function $H : \mathbb{R}_+^M \to \mathbb{R}_+$ satisfying*

$$H(z_1, \ldots, z_M) = G(\omega_1 z_{j(1)}, \ldots, \omega_N z_{j(N)});$$

2. $$\mathbb{P}\left[Y_j = \max_i X_i\right] = \frac{\sum_{i \in \mathcal{I}_j} T_i G_i(T_1, \ldots, T_N)}{G(T_1, \ldots, T_N)},$$

   where $G_i(x_1, \ldots, x_N) \equiv \partial G(x_1, \ldots, x_N)/\partial x_i$;

3. *For any $j = 1, \ldots, M$, the distribution of Y_j conditional on the event $Y_j = \max_{i=1,\ldots,N} X_i$ is identical to the distribution of $\max_{i=1,\ldots,N} X_i$,*

$$\mathbb{P}\left[Y_j \leq y \mid Y_j = \max_i X_i\right] = e^{-G(T_1,\ldots,T_N)y^{-\theta}} = \mathbb{P}\left[\max_{i=1,\ldots,N} X_i \leq y\right].$$

Proof. We first prove part (1). Let $\{\mathcal{I}_j\}_{j=1}^M$ be a partition of $\{1, \ldots, N\}$ and define $Y_j = \max_{i \in \mathcal{I}_j} X_i$. Let the function $j : \{1, \ldots, N\} \to \{1, \ldots, M\}$ satisfy $i \in \mathcal{I}_{j(i)}$ for all $i = 1, \ldots, N$. Note that there is a unique function satisfying this condition since $\{\mathcal{I}_j\}_{j=1}^M$ is a partition of $\{1, \ldots, N\}$. Then,

$$\begin{aligned} \mathbb{P}\left[Y_j \leq y_j, \forall j = 1, \ldots, M\right] &= \mathbb{P}\left[X_i \leq y_j, \forall i \in \mathcal{I}_j, \forall j = 1, \ldots, M\right] \\ &= e^{-G(T_1 y_{j(1)}^{-\theta}, \ldots, T_N y_{j(N)}^{-\theta})}. \end{aligned}$$

Therefore $\{Y_1, \ldots, Y_M\}$ is θ-Fréchet. Its scale parameters are

$$\lim_{y_k \to \infty, k \neq j} G(T_1 y_{j(1)}^{-\theta}, \ldots, T_N y_{j(N)}^{-\theta}) = G(T_1 \mathbf{1}\{1 \in \mathcal{I}_j\}, \ldots, T_N \mathbf{1}\{N \in \mathcal{I}_j\}) = \tilde{T}_j,$$

and its correlation function must then be

$$G(T_1/\tilde{T}_{j(1)} z_{j(1)}, \ldots, T_N/\tilde{T}_{j(N)} z_{j(N)}) = G(\omega_1 z_{j(1)}, \ldots, \omega_N z_{j(N)}) = H(z_1, \ldots, z_M).$$

Note that if we take $M = 1$ so that $\mathcal{I}_1 = \{1, \ldots, N\}$ we get

$$\begin{aligned} \mathbb{P}\left[\max_{i=1,\ldots,N} X_i \leq y\right] &= \mathbb{P}\left[Y_1 \leq y\right] = \mathbb{P}\left[Y_j \leq y, \forall j = 1, \ldots, M\right] \\ &= e^{-G(T_1 y^{-\theta}, \ldots, T_N y^{-\theta})} = e^{-G(T_1,\ldots,T_N) y^{-\theta}}. \end{aligned}$$

That is, $\max_{i=1,\ldots,N} X_i$ is a θ-Fréchet random variable with scale $G(T_1,\ldots,T_N)$ and shape θ.

Next we prove part (2). We have

$$\begin{aligned}
&\mathbb{P}\left[\max_i X_i \le y \text{ and } Y_j = \max_i X_i\right] = \mathbb{P}\left[Y_j \le y \text{ and } Y_j = \max_i X_i\right] \\
&= \mathbb{P}\left[Y_j \le y \text{ and } X_i \le Y_j, \forall i = 1,\ldots,N\right] = \mathbb{P}\left[Y_j \le y \text{ and } X_i \le Y_j, \forall i \notin \mathcal{I}_j\right] \\
&= \int_0^y \mathbb{P}\left[X_i \le t, \forall i \notin \mathcal{I}_j \mid Y_j = t\right] \frac{\partial}{\partial t}\mathbb{P}[Y_j \le t]\mathrm{d}t \\
&= \int_0^y \frac{\partial}{\partial t}\, \mathbb{P}\left[X_i \le z, \forall i \notin \mathcal{I}_j, \text{ and } X_i \le t, \forall i \in \mathcal{I}_j\right]\Big|_{z=t} \mathrm{d}t \\
&= \int_0^y \sum_{i\in\mathcal{I}_j} \frac{\partial}{\partial y_i}\, e^{-G(T_1 y_1^{-\theta},\ldots,T_N y_N^{-\theta})}\Big|_{y_i=t,\forall i=1,\ldots,N} \mathrm{d}t \\
&= \int_0^y \sum_{i\in\mathcal{I}_j} e^{-G(T_1 y_1^{-\theta},\ldots,T_N y_N^{-\theta})} G_i(T_1 y_1^{-\theta},\ldots,T_N y_N^{-\theta}) T_i \theta y_i^{-\theta-1}\Big|_{y_i=t,\forall i=1,\ldots,N} \mathrm{d}t \\
&= \int_0^y e^{-G(T_1,\ldots,T_N)t^{-\theta}} \sum_{i\in\mathcal{I}_j} T_i G_i(T_1,\ldots,T_N)\theta t^{-\theta-1}\mathrm{d}t \\
&= \frac{\sum_{i\in\mathcal{I}_j} T_i G_i(T_1,\ldots,T_N)}{G(T_1,\ldots,T_N)} \int_0^y e^{-G(T_1,\ldots,T_N)t^{-\theta}} G(T_1,\ldots,T_N)\theta t^{-\theta-1}\mathrm{d}t \\
&= \frac{\sum_{i\in\mathcal{I}_j} T_i G_i(T_1,\ldots,T_N)}{G(T_1,\ldots,T_N)} e^{-G(T_1,\ldots,T_N)y^{-\theta}},
\end{aligned}$$

where $G_i(x_1,\ldots,x_N) = \partial G(x_1,\ldots,x_N)/\partial x_i$. Let $y \to \infty$ to get

$$\mathbb{P}\left[Y_j = \max_i X_i\right] = \frac{\sum_{i\in\mathcal{I}_j} T_i G_i(T_1,\ldots,T_N)}{G(T_1,\ldots,T_N)}.$$

Finally, we can prove part (3) using the previous results:

$$\begin{aligned}
\mathbb{P}\left[\max_i X_i \le y \mid Y_j = \max_i X_i\right] &= \frac{\mathbb{P}\left[\max_i X_i \le y \text{ and } Y_j = \max_i X_i\right]}{\mathbb{P}\left[Y_j = \max_i X_i\right]} \\
&= \frac{\frac{\sum_{i\in\mathcal{I}_j} T_i G_i(T_1,\ldots,T_N)}{G(T_1,\ldots,T_N)} e^{-G(T_1,\ldots,T_N)z^{-\theta}}}{\frac{\sum_{i\in\mathcal{I}_j} T_i G_i(T_1,\ldots,T_N)}{G(T_1,\ldots,T_N)}} \\
&= e^{-G(T_1,\ldots,T_N)z^{-\theta}} \\
&= \mathbb{P}\left[\max_i X_i \le y\right].
\end{aligned}$$

□

B Proof of Lemma 1

First, we show that if productivity is θ-Fréchet, then there must exist a correlation function $G^d : \mathbb{R}_+^N \to \mathbb{R}_+$ such that (5) is the joint distribution of productivity across origins.

Consider any $(x_1, \ldots, x_N) \in \mathbb{R}_+^N$. Then $x_o^{1/\theta} \geq 0$ for each o. From the definition of a multivariate θ-Fréchet random variable, the random variable $\max_{o=1,\ldots,N} x_o^{1/\theta} A_{od}(v)$ must be distributed as a θ-Fréchet random variable. That is, there exists some $T > 0$ such that

$$\mathbb{P}\left[\max_{o=1,\ldots,N} x_o^{1/\theta} A_{od}(v) \leq a\right] = e^{-Ta^{-\theta}}.$$

Let $T^d : \mathbb{R}_+^N \to \mathbb{R}_+$ be the map $(x_1, \ldots, x_N) \mapsto T$. We then have that for any $(x_1, \ldots, x_N) \in \mathbb{R}_+^N$

$$\mathbb{P}\left[\max_{o=1,\ldots,N} x_o^{1/\theta} A_{od}(v) \leq a\right] = \exp\left[-T^d(x_1, \ldots, x_N) a^{-\theta}\right].$$

Note that the joint distribution of productivity can be written as

$$\begin{aligned}
\mathbb{P}[A_{1d}(v) \leq a_1, \ldots, A_{Nd}(v) \leq a_N] &= \mathbb{P}[A_{1d}(v)/a_1 \leq 1, \ldots, A_{Nd}(v)/a_N \leq 1] \\
&= \mathbb{P}\left[\max_{o=1,\ldots,N} A_{od}(v)/a_o \leq 1\right].
\end{aligned}$$

Choosing $x_o = a_o^{-\theta}$ and $a = 1$ we can use the properties of our function T^d and get

$$\mathbb{P}\left[\max_{o=1,\ldots,N} A_{od}(v)/a_o \leq 1\right] = \exp\left[-T^d(a_1^{-\theta}, \ldots, a_N^{-\theta})\right].$$

Therefore, the joint distribution of productivity satisfies

$$\mathbb{P}[A_{1d}(v) \leq a_1, \ldots, A_{Nd}(v) \leq a_N] = e^{-G^d(T_{1d}a_1^{-\theta}, \ldots, T_{Nd}a_N^{-\theta})},$$

for the function $G^d : \mathbb{R}_+^N \to \mathbb{R}_+$ defined by $(x_1, \ldots, x_N) \mapsto T^d(x_1/T_{1d}, \ldots, x_N/T_{Nd})$.

We now show that this G^d is a correlation function. First we show that it must be homogenous. Fix $(x_1, \dots, x_N) \in \mathbb{R}_+^N$ and let $\lambda > 0$. We have

$$\begin{aligned}
\exp\left[-G^d(\lambda x_1, \dots, \lambda x_N)\right] &= \mathbb{P}[T_{1d}A_{1d}(v)^{-\theta} \geq \lambda x_1, \dots, T_{Nd}A_{Nd}(v)^{-\theta} \geq \lambda x_N] \\
&= \mathbb{P}[(x_1/T_{1d})^{1/\theta}A_{1d}(v) \leq \lambda^{-1/\theta}, \dots, (x_N/T_{Nd})^{-1/\theta}A_{Nd}(v) \leq \lambda^{-1/\theta}] \\
&= \mathbb{P}[\max_{o=1,\dots,N}(x_o/T_{od})^{-1/\theta}A_{od}(v) \leq \lambda^{-1/\theta}] \\
&= \exp\left[-T^d(x_1/T_{1d}, \dots, x_N/T_{Nd})\lambda\right] \\
&= \exp\left[-\lambda G^d(x_1, \dots, x_N)\right],
\end{aligned}$$

so that $G^d(\lambda x_1, \dots, \lambda x_N) = \lambda G^d(x_1, \dots, x_N)$ as desired.

Now consider the normalization restriction. Fix o. The distribution of $A_{od}(v)$ is

$$\exp\left(-T_{od}a^{-\theta}\right) = \mathbb{P}[A_{od}(v) \leq a] = \mathbb{P}\left[\max_{n=1,\dots,N} x_n^{1/\theta}A_{nd}(v) \leq a\right],$$

for the choice of $x_n = 0$ for $n \neq o$ and $x_o = 1$. But then,

$$\begin{aligned}
\exp\left(-T_{od}a^{-\theta}\right) &= \exp\left[-T^d(x_1, \dots, x_N)a^{-\theta}\right] \\
&= \exp\left[-T^d(0, \dots, 0, 1, 0, \dots, 0)a^{-\theta}\right] \\
&= \exp\left[-G^d(0, \dots, 0, T_{od}, 0, \dots, 0)a^{-\theta}\right] \\
&= \exp\left[-G^d(0, \dots, 0, 1, 0, \dots, 0)T_{od}a^{-\theta}\right],
\end{aligned}$$

where the last equality comes from the homogeneity of G^d. We therefore must have $G^d(0, \dots, 0, 1, 0, \dots, 0) = 1$ as desired.

The unboundedness restriction follows from the limiting properties of joint distributions. Fix o. Then,

$$\begin{aligned}
\lim_{x_o \to \infty} e^{-G^d(x_1, \dots, x_N)} &= \lim_{x_o \to \infty} \mathbb{P}[T_{1d}A_{1d}(v)^{-\theta} \geq x_1, \dots, A_{Nd}(v) \geq x_N] \\
&= \lim_{x_o \to \infty} \mathbb{P}[T_{1d}^{-1/\theta}A_{1d}(v) \leq x_1, \dots, T_{Nd}^{-1/\theta}A_{Nd}(v) \leq x_N] = 0.
\end{aligned}$$

Therefore, $\lim_{x_o \to \infty} G^d(x_1, \dots, x_N) = \infty$ as desired.

Finally, the differentiability restrictions are necessary because the productivity distribution is continuous and therefore has a joint density function. Smith (1984) shows that the differentiability condition is necessary for this joint density to exist.

Therefore, the function G^d must be a correlation function, and we have proven that if productivity is θ-Fréchet then there exists a correlation function $G^d : \mathbb{R}_+^N \to \mathbb{R}_+$ such that (5) holds.

We now prove the converse. Let $T_{od} > 0$ for each $o = 1, \ldots, N$, and let $G^d : \mathbb{R}_+^N \to \mathbb{R}_+$ be a correlation function. Suppose that $\{A_{od}(v)\}_{o=1}^N$ satisfies

$$\mathbb{P}[A_{1d}(v) \leq a_1, \ldots, A_{Nd}(v) \leq a_N] = \exp\left[-G^d(T_{od}a_1^{-\theta}, \ldots, T_{Nd}a_N^{-\theta})\right].$$

We want to show that $\{A_{od}(v)\}_{o=1}^N$ is θ-Fréchet. Let $(x_1, \ldots, x_N) \in \mathbb{R}_+^N$ and consider the distribution of $\max_{o=1,\ldots,N} x_o A_{od}(v)$,

$$\begin{aligned}
\mathbb{P}[\max_{o=1,\ldots,N} x_o A_{od}(v) \leq a] &= \mathbb{P}[x_1 A_{1d}(v) \leq a, \ldots, x_N A_{Nd}(v) \leq a] \\
&= \mathbb{P}[A_{1d}(v) \leq a/x_1, \ldots, A_{Nd}(v) \leq a/x_N] \\
&= \exp\left[-G^d(T_{od}x_1^{\theta}a^{-\theta}, \ldots, T_{Nd}x_N^{\theta}a^{-\theta})\right] \\
&= \exp\left[-G^d(T_{od}x_1^{\theta}, \ldots, T_{Nd}x_N^{\theta})a^{-\theta}\right],
\end{aligned}$$

where the last equality uses the homogeneity of G^d. Therefore, $\max_{o=1,\ldots,N} x_o A_{od}(v)$ is a θ-Fréchet random variable with scale parameter $G^d(T_{od}x_1^{\theta}, \ldots, T_{Nd}x_N^{\theta})$. As a result, we conclude that $\{A_{od}(v)\}_{o=1}^N$ is θ-Fréchet.

C Proof of Proposition 1

Perfect competition implies that potential import prices are

$$P_{od}(v) = \frac{W_o}{A_{od}(v)}.$$

Then,

$$\begin{aligned}
\mathbb{P}[P_{1d}(v) \geq p_1, \ldots, P_{Nd}(v) \geq p_N] &= \mathbb{P}[P_{1d}(v)/W_1 \geq p_1/W_1, \ldots, P_{Nd}(v)/W_N \geq p_N/W_N] \\
&= \mathbb{P}[1/A_{1d}(v) \geq p_1/W_1, \ldots, 1/A_{Nd}(v) \geq p_N/W_N] \\
&= \mathbb{P}[A_{1d}(v) \leq W_1/p_1, \ldots, A_{Nd}(v) \leq W_N/p_N] \\
&= \exp\left[-G^d(T_{1d}W_1^{-\theta}p_1^{\theta}, \ldots, T_{Nd}W_N^{-\theta}p_n^{\theta})\right].
\end{aligned}$$

D Proof of Proposition 2

The proof follows directly from the properties of θ-Fréchet random variables. The probability that variety v is imported by destination d from origin o is

$$\pi_{od} \equiv \mathbb{P}[P_{od}(v) \geq P_{o'd}(v) \quad \forall o' \neq o] = \frac{T_{od} W_o^{-\theta} G_o^d(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta})}{G^d(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta})},$$

by Proposition 1 and Lemma A.3. The distribution of prices among goods imported by destination d from country o satisfies

$$\mathbb{P}\left[P_{od}(v) \geq p \mid P_{od}(v) = \min_{o'=1,\ldots,N} P_{o'd}(v)\right] = \mathbb{P}\left[\min_{o'=1,\ldots,N} P_{o'd}(v) \geq p\right] = e^{-G(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta}) p^\theta},$$

by Proposition 1 and Lemma A.3. The price index in destination d is then

$$P_d = \left[\int_0^1 \min_{o=1,\ldots,N} P_{od}(v)^{-\epsilon} \mathrm{d}v\right]^{-\frac{1}{\epsilon}} = \left[\mathbb{E}\left(\min_{o=1,\ldots,N} P_{od}(v)^{-\epsilon}\right)\right]^{-\frac{1}{\epsilon}} = \gamma G^d \left(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta}\right)^{-\frac{1}{\theta}},$$

where $\gamma = \Gamma\left(\frac{\theta-\epsilon}{\theta}\right)^{-\frac{1}{\epsilon}}$, $\Gamma(\cdot)$ is the gamma function, and the last equality follows from the fact that $\min_{o=1,\ldots,N} P_{od}(v)^{-\epsilon} = (\max_{o=1,\ldots,N} 1/P_{od}(v))^\epsilon$ is a Fréchet random variable with scale $G^d\left(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta}\right)$ and shape $\theta/\epsilon > 1$ due to the assumption that $\theta > \epsilon$ and due to Lemma A.1.

E Proof of Proposition 3

First, the set of varieties from o imported to d is $\{v \in [0,1] \mid W_o/A_{od}(v) = \min_{o'} W_{o'}/A_{o'd}(v)\}$ and for any variety in this set, expenditure is

$$X_d(v) = \left(\frac{W_o/A_{od}(v)}{P_d}\right)^{-\epsilon} X_d.$$

Any v not in this set must get imported from a different origin. The price index is

$$P_d = \left[\int_0^1 \left(\min_{o'} W_{o'}/A_{o'd}(v)\right)^{-\epsilon} \mathrm{d}v\right]^{-\frac{1}{\epsilon}},$$

so that we can write the expenditure share as

$$
\begin{aligned}
\pi_{od}(\mathbf{W}, X_d) &\equiv \int_0^1 \frac{X_d(v)}{X_d} \mathbf{1}\left\{W_o/A_{od}(v) = \min_{o'} W_{o'}/A_{o'd}(v)\right\} \mathrm{d}v \\
&= \frac{\int_0^1 (W_o/A_{od}(v))^{-\epsilon} \mathbf{1}\left\{W_o/A_{od}(v) = \min_{o'} W_{o'}/A_{o'd}(v)\right\} \mathrm{d}v}{\int_0^1 \left(\min_{o'} W_{o'}/A_{o'd}(v)\right)^{-\epsilon} \mathrm{d}v} \\
&= \frac{\mathbb{E}\left[(W_o/A_{od}(v))^{-\epsilon} \mathbf{1}\left\{W_o/A_{od}(v) = \min_{o'} W_{o'}/A_{o'd}(v)\right\}\right]}{\mathbb{E}\left[\left(\min_{o'} W_{o'}/A_{o'd}(v)\right)^{-\epsilon}\right]}.
\end{aligned}
$$

We need to show that there exists a correlation function that approximates this factor demand system. The proof is similar to the proof of Theorem 1 in Dagsvik (1995), differing in the functional form of the demand system to be approximated.

We start by constructing an approximating GEV factor demand system that generates the same price level as multiplying productivity by independent Fréchet noise. Consider the random vector $\{A_{od}(v)U_{od}(v)\}_{o=1}^N$ where $U_{od}(v)$ is some θ-Fréchet noise with unit scale that is independent across o and independent of of $\{A_{od}(v)\}_{o=1}^N$. Under this modified productivity distribution, potential import prices are $P_{od}(v) = W_o/(A_{od}(v)U_{od}(v))$ and $\{P_{od}(v)^{-\epsilon}\}_{o=1}^N \mid \{A_{od}\}_{o=1}^N$ is θ/ϵ-Fréchet with scale $(A_{od}(v)/W_o)^\theta$ by Lemma A.1 and independent across o. As a consequence, $P_d(v)^{-\epsilon} \mid \{A_{od}\}_{o=1}^N$ is also θ/ϵ-Fréchet and has scale $\sum_{o=1}^N (A_{od}(v)/W_o)^\theta$. The associated price level is

$$
\begin{aligned}
P_d &= \left\{\mathbb{E}\left[\mathbb{E}\left(P_d(v)^{-\epsilon} \mid \{A_{od}(v)\}_{o=1}^N\right)\right]\right\}^{-\frac{1}{\epsilon}} \\
&= \mathbb{E}\left[\Gamma(1-\epsilon/\theta)\left(\sum_{o=1}^N (A_d(v)/W_o)^\theta\right)^{\frac{\epsilon}{\theta}}\right]^{-\frac{1}{\epsilon}} \\
&= \Gamma(1-\epsilon/\theta)^{-\frac{1}{\epsilon}} G^d(T_{1d}W_1^{-\theta}, \ldots, T_{Nd}W_N^{-\theta})^{-\frac{1}{\theta}},
\end{aligned}
$$

for $T_{od} \equiv \mathbb{E}A_{od}(v)^\theta$, and

$$
G^d(x_1, \ldots, x_N) \equiv \left[\mathbb{E}\left(\sum_o A_{od}(v)^\theta x_o/T_{od}\right)^{\frac{\epsilon}{\theta}}\right]^{\frac{\theta}{\epsilon}}.
$$

Note that this price level is identical to assuming that productivity is θ-Fréchet with scale T_{od} and correlation function G^d. It also approximates the true price level. In particular,

$$
\begin{aligned}
P_d &= \left[\Gamma\left(\frac{\theta-\epsilon}{\theta}\right)\mathbb{E}\left(\sum_o (A_{od}(v)/W_o)^\theta\right)^{\frac{\epsilon}{\theta}}\right]^{-\frac{1}{\epsilon}} \\
&\overset{\theta\to\infty}{\to} \left[\mathbb{E}\left(\max_o A_{od}(v)/W_o\right)^\epsilon\right]^{-\frac{1}{\epsilon}} = \left[\mathbb{E}\left(\min_o W_o/A_{od}(v)\right)^{-\epsilon}\right]^{-\frac{1}{\epsilon}}.
\end{aligned}
$$

That is, the price level implied by either multiplying by θ-Fréchet noise or by assuming θ-Fréchet productivity with this correlation function converges point-wise to the price level associated with the true productivity distribution.

The implied GEV factor demand system is

$$
\begin{aligned}
\pi_{od}^{GEV}(\mathbf{W}, X_d;\theta) &= \frac{T_{od}W_o^{-\theta}G_o^d(T_{1d}W_1^{-\theta},\ldots,T_{Nd}W_N^{-\theta})}{G^d(T_{1d}W_1^{-\theta},\ldots,T_{Nd}W_N^{-\theta})} \\
&= \frac{\mathbb{E}\left[\left(\sum_{o'}(A_{o'd}(v)/W_{o'})^\theta\right)^{\frac{\epsilon}{\theta}-1}(A_{od}(v)/W_o)^\theta\right]}{\mathbb{E}\left(\sum_{o'}(A_{o'd}(v)/W_{o'})^\theta\right)^{\frac{\epsilon}{\theta}}} \\
&\overset{\theta\to\infty}{\to} \frac{\mathbb{E}\left[(W_o/A_{od}(v))^{-\epsilon}\mathbf{1}\left\{W_o/A_{od}(v) = \min_{o'} W_{o'}/A_{o'd}(v)\right\}\right]}{\mathbb{E}\left[(\min_{o'} W_{o'}/A_{o'd}(v))^{-\epsilon}\right]} = \pi_{od}(\mathbf{W}, X_d).
\end{aligned}
$$

That is, the implied GEV factor demand system converges point-wise to the true demand system. To establish uniform convergence across $(\mathbf{W}, X_d) \in K$, for $K \subset \mathbb{R}_+^{N+1}$ compact, note that if the sequence $\{\pi_{od}^{GEV}(\mathbf{W}, X_d;\theta_j)\}_{j=1}^\infty$ is convergent, there exists a positive sequence $\{\theta_k\}_{k=1}^\infty$ that diverges such that $\{\pi_{od}^{GEV}(\mathbf{W}, X_d;\theta_k)\}_{k=1}^\infty$ is monotone and converges. Then, since $\pi_{od}(\mathbf{W}, X_d)$ is continuous, we can apply Theorem 7.13 in Rudin et al. (1964) to establish uniform convergence.

F Proof of Theorem 1

Sufficiency follows from Campbell's theorem (see Kingman (1992)). Under Assumption 1,

$$
\begin{aligned}
&\mathbb{P}\left[A_{od}(v) \leq a_o, \forall o = 1, \ldots, N\right] \\
&\mathbb{P}\left[\max_{i=1,2,\ldots} Z_i(v) A_{od}(\chi_i(v)) \leq a_o, \forall o = 1, \ldots, N\right] \\
&\mathbb{P}\left[Z_i(v) A_{od}(\chi_i(v)) \leq a_o, \forall o = 1, \ldots, N, \forall i = 1, 2, \ldots\right] \\
&\mathbb{P}\left[Z_i(v) \leq \min_{o=1,\ldots,N} a_o / A_{od}(\chi_i(v)), \forall i = 1, 2, \ldots\right] \\
&\mathbb{P}\left[Z_i(v) > \min_{o=1,\ldots,N} a_o / A_{od}(\chi_i(v)), \text{ for no } i = 1, 2, \ldots\right].
\end{aligned}
$$

This last expression is a void probability. Under Assumption 2 we can compute it by applying Campbell's theorem,

$$
\begin{aligned}
&\mathbb{P}\left[Z_i(v) > \min_{o=1,\ldots,N} a_o / A_{od}(\chi_i(v)), \text{ for no } i = 1, 2, \ldots\right] \\
&= \exp\left[-\int_{\mathcal{X}} \int_{\min_{o=1,\ldots,N} a_o / A_{od}(\chi)}^{\infty} \theta z^{-\theta-1} \mathrm{d}z \mathrm{d}\mu(\chi)\right] \\
&= \exp\left[-\int_{\mathcal{X}} \max_{o=1,\ldots,N} A_{od}(\chi)^{\theta} a_o^{-\theta} \mathrm{d}\mu(\chi)\right] \\
&= \exp\left[-\int_{\mathcal{X}} \max_{o=1,\ldots,N} \frac{A_{od}(\chi)^{\theta}}{T_{od}} T_{od} a_o^{-\theta} \mathrm{d}\mu(\chi)\right],
\end{aligned}
$$

for $T_{od} \equiv \int_{\mathcal{X}} A_{od}(\chi)^{\theta} \mathrm{d}\mu(\chi)$. This final expression is the joint distribution of a multivariate θ-Fréchet random variable with scale parameters T_{od} for each $o = 1, \ldots, N$ and correlation function $G^d(x_1, \ldots, x_K) \equiv \int_{\mathcal{X}} \max_{o=1,\ldots,N} \frac{A_{od}(\chi)^{\theta}}{T_{od}} x_o \mathrm{d}\mu(\chi)$.

Necessity follows from Theorem 1 in Kabluchko (2009), which states that any θ-Fréchet process has a spectral representation. Let $\{A_{od}\}_{o=1,\ldots,N}$ be a θ-Fréchet process on $\{1, \ldots, N\}$—that is, a multivariate θ-Fréchet random vector. Then there exists a σ-finite measure space $(\mathcal{X}, \mathcal{F}, \mu)$, spectral functions $\{A_{od}(\chi)\}_{o=1,\ldots,N}$ with $\int_{\mathcal{X}} A_{od}(\chi) \mathrm{d}\chi < \infty$, and a Poisson process $\{Z_i, \chi_i\}_{i=1,2,\ldots}$ with intensity $\theta z^{-\theta-1} \mathrm{d}z \mathrm{d}\mu(\chi)$ such that $A_{od} = \max_{i=1,2,\ldots} Z_i A_{od}(\chi_i)$. Taking $\{A_{od}(v)\}_{o=1,\ldots,N}$ across $v \in [0,1]$ to be $i.i.d.$ copies of $\{A_{od}\}_{o=1,\ldots,N}$ completes the proof.

G Generalized Mixed CES

We now provide a concrete example of applying Corollary 2 to generate a *generalized mixed-CES factor demand system*. This model includes the mixed-CES model estimated in Adao et al. (2017) as a special case, as well as many existing models based on EK, as presented in Section 5.

We get the generalized mixed-CES factor demand system by specifying a particular structure for the innovation's attributes. Let $\chi_i(v) = \{k_i(v), \sigma_i(v), \beta_i(v), \nu_i(v)\}$ where $k_i(v) \in \{1, \ldots, K\}$ is a discrete characteristic (e.g., a sector) and assume that $A_{od}(k_i(v), \varepsilon_i(v)) \mid \sigma_i(v) = \sigma, \beta_i(v) = \beta, k_i(v) = k$ is σ-Fréchet and independent across o with scale equal to $\Gamma(1-\theta/\sigma)^{-\sigma/\theta} e^{\beta' x_{od} + u_{kod}}$. The random coefficient $\sigma_i(v)$ parameterizes within-k dispersion in spatial applicability, and $\beta_i(v)$ is a vector of random coefficients for the dependence of spatial applicability on a vector of covariates, x_{od}.

Under this assumption, the distribution of productivity conditional on $\sigma_i(v) = \sigma, \beta_i(v) = \beta, k_i(v) = k$ is θ-Fréchet with CES correlation function with correlation coefficient $\rho(\sigma) = 1 - \theta/\sigma$ and scale

$$T_{kod}(\sigma, \beta) = e^{(1-\rho(\sigma))(\beta' x_{od} + u_{kod})}.$$

Within-(k, σ, β) expenditure shares are

$$\frac{X_{kod}(\sigma, \beta)}{\sum_{o'=1}^{N} X_{ko'd}(\sigma, \beta)} = \frac{e^{\beta' x_{od} + u_{kod}} W_o^{-\sigma}}{\sum_{o'=1}^{N} e^{\beta' x_{o'd} + u_{ko'd}} W_{o'}^{-\sigma}},$$

while between-(σ, k, β) expenditure levels are

$$\sum_{o'=1}^{N} X_{ko'd}(\sigma, \beta) = \frac{P_{kd}(\sigma, \beta)^{-\theta} f(\sigma, \beta \mid k) m_k}{\sum_{k=1}^{K} \int_{\mathbb{R}^J} \int_{\theta}^{\infty} P_{kd}(\sigma, \beta)^{-\theta} f(\sigma, \beta \mid k) m_k \mathrm{d}\sigma \mathrm{d}\beta} X_d,$$

where $P_{kd}(\sigma, \beta) \equiv \gamma \left[\sum_{o=1}^{N} e^{\beta' x_{od} + u_{kod}} W_o^{-\sigma}\right]^{-1/\sigma}$.

In the limit as $\theta \to 0$, the density of expenditure across characteristics converges to the likelihood that an innovation has $k_i(v) = (\sigma, \beta, k)$. As a result, the share of expenditure by d on goods produced in o with innovations of type k is mixed-CES,

$$\frac{X_{kod}}{X_d} = \int_{\mathbb{R}^J} \int_0^{\infty} \frac{e^{\beta' x_{od} + u_{kod}} W_o^{-\sigma}}{\sum_{o'=1}^{N} e^{\beta' x_{o'd} + u_{ko'd}} W_{o'}^{-\sigma}} f(\sigma, \beta \mid k) \mathrm{d}\sigma \mathrm{d}\beta \frac{m_k}{\sum_{k'=1}^{K} m_{k'}},$$

where $X_{kod} \equiv \int_{\mathbb{R}^J} \int_0^\infty X_{kod}(\sigma, \beta) \mathrm{d}\sigma \mathrm{d}\beta$. This limiting case corresponds to the mixed-CES factor demand system used by Adao et al. (2017), for $K = 1$, and x_{od} containing origin and destination dummy variables as well as the logarithm of per capita income in country o.

H Data Construction

For our quantitative analysis, we use trade flow data from the World Input-Output Database (WIOD), tariff data from the United Nations Comtrade Database, and gravity covariates from Centre D'Études Prospectives et d'Informations Internationales (CEPII). When calculating the trade costs implied by this data, we use GDP deflator data from the Penn World Tables (PWT), version 9.0.

H.1 Map from SITC Codes to WIOD Sectors

The WIOD data allows us to compute the total value of trade between a sample of 40 countries across 25 sectors from 1995 through 2011. The sector classification in this data set comes from aggregating underlying data classified according to the third revision of the International Standard Industrial Classification (ISIC). The Comtrade tariff data is classified according to the second revision of the Standard International Trade Classification (SITC). In order to merge these data sources, we construct a mapping that assigns SITC codes to WIOD sectors.

First, we match ISIC and SITC definitions using existing correspondences of each standard to Harmonized System (HS) product definitions. These correspondences come from the World Bank's World Integrated Trade Solution (WITS).[29] This merge matches on 5,701 products out of 5,705 total HS products. We drop the four unmatched products. This creates a HS product dataset with 764 SITC codes and 35 ISIC codes. Note that there are 925 SITC codes in the tariff data to be classified into WIOD sectors.

Next, we map the ISIC definitions in this merge to the 25 WIOD sectors using the relation between ISIC codes and the WIOD sectors. This leaves products in the ISIC code 99 ("Goods not elsewhere classified") without a WIOD sector definition.

At this point, there are two issues we must address: (1) classifying SITC codes that have products in multiple WIOD sectors; and (2) classifying the SITC codes in the tariff data that were either matched to ISIC code 99 or were not matched to any ISIC code. We use a most-common-sector rule and manual classification based on SITC codes to resolve these two issues and arrive at a mapping from SITC codes to WIOD sectors.

[29] They are available at https://wits.worldbank.org/product_concordance.html.

We proceed as follows. First, we determine the most common WIOD sector classification (including "unclassified") at the HS product level of each 4-digit SITC code within the merge. We re-classify all products within an 4-digit SITC sector as belonging to the most common WIOD sector, and break ties manually. This step resolves issue (1) and leaves us with 764 4-digit SITC codes mapped to a unique WIOD sector, and 161 4-digit SITC codes left unclassified.

Second, we resolve issue (2) by refining the map by using the most common classification of HS products within each 3-digit SITC code, again breaking ties manually. In this step, we only use the most-common classification at the 3 digit level to classify previously unclassified 4-digit SITC codes, filling in the map. This step mostly resolves issue (2), leaving only 12 4-digit SITC codes unclassified. We complete the map by manually classifying ten of these remaining codes, while choosing to leave codes 9110 ("Postal packages not classified according to kind") and 9310 ("Special transactions, commodity not classified according to class") unclassified.

H.2 Construction of Sectoral Trade Flow and Tariff Data

With this mapping from (all but two) 4-digit SITC codes to WIOD sectors, we next aggregate the Comtrade tariff data to the WIOD sector level. First, we compute the average applied tariff and total value of trade within the Comtrade data by SITC code, exporter, importer, and year. We then compute the average tariff and total trade value by WIOD sector, exporter, importer, and year, using the value of total trade in each SITC code and year as weights when calculating averages, and dropping codes 9110 and 9310.

Next, we merge these data with the WIOD data. The WIOD data give us the amount of imports by each sector and country across sectors of all other countries. We first aggregate this input-output data to get total expenditure by each importer across the sectors of each exporting country. This aggregation gives a balanced bilateral dataset of trade flows across 25 sectors for each exporter-importer pair from 1995 to 2011. The data contains 40 countries and a rest-of-world aggregate (1,681 pairs per sector, including self trade).

We merge this data with our tariff data at the WIOD sector, exporter, importer, and year level. The two dataset intersect from 1995 through 2007. For each year, we drop any observations in the tariff data that are not in the WIOD data. This eliminates countries without WIOD bilateral data. We set tariffs for self trade to zero. Additionally, we have no tariff data for the rest-of-world aggregate and Romania, and limited data for Taiwan. We drop these three entities leaving us with a sample of bilateral trade flows and tariffs between 38 countries. Finally, we do not have tariff data for sectors 15 through 25 (non-traded sectors), so we also drop them from the data.

The resulting dataset has many trade zeros and missing tariff observations. To address this potential issue, we aggregate together WIOD sectors to get the final ten sector definitions we use in our quantitative analysis. Specifically, we combine the "Coke, Refined Petroleum and Nuclear Fuel" and "Chemicals, Rubber, and Plastics" WIOD sectors to form our "Fuel, Chemicals, Rubber, and Plastics" sector. Also, we combine the "Machinery, n.e.c.," "Electrical and Optical Equipment," "Transport Equipment," and "Manufacturing, Nec; Recycling" sectors to get our "Machinery, Equipment, and Manufacturing n.e.c." sector. We compute the aggregate value of trade within each of our sectors across bilateral pairs and years, and compute average tariffs for each of our sectors across bilateral pairs and years using total global trade in each WIOD sector and year as weights.

This aggregation results in a balanced dataset of trade flows and tariffs across 10 sectors and 38 countries (1,444 exporter-importer pairs) from 1995 to 2007. The share of trade zeros is 1.3 percent, and the share of missing tariff observations is 8.92 percent. Conditional on zero trade, the probability of tariffs being missing is 42.9 percent and conditional on a missing tariff, the probability of a trade zero is 6.3 percent. We finally merge in the CEPII data on geography and other standard gravity covariates.

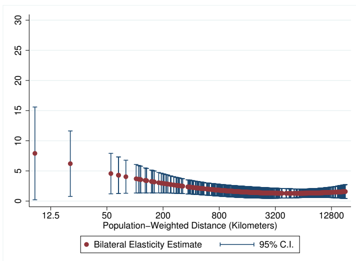
I Additional Tables and Figures

Table I.1: Estimates of the trade elasticity θ, OLS.

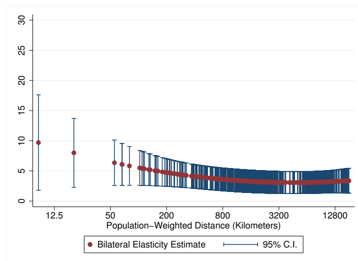
Dep variable	CES $\ln \pi_{sodt}$	Cross-nested CES $\ln \sum_o \pi_{sodt}$	Spatial $\ln \sum_o \pi_{sodt}$
	(1)	(2)	(3)
$\ln(1 + t_{sodt})$	5.523	0.607	0.489
	(0.020)***	(0.025)***	(0.024)***
$\hat{x}^{cnces}_{sodt}$		✓	
$\hat{x}^{spatial}_{sodt}$			✓
Sector-Covariate Interactions	✓	✓	✓
Sector-Origin-Year Effects	✓	✓	✓
Sector-Destination-Year Effects	✓		
Destination-Year Effects		✓	✓
Observations	174,201	174,201	174,201
R-squared	0.82	0.74	0.73
F-statistic	759.2	1,485.4	427.8

Notes: Results in column (1) are from estimating (46) by OLS, assuming that $\rho_{sod} = 0$ so that $\sigma_{sod} = \theta$. Results in columns (2) and (3) are from estimating (47) by OLS, with $\hat{x}^{l}_{sodt} = \ln(\widehat{P_{sod}/P_{sd}})$, $l = spatial, cnces$, from the first-stage gravity equation in (46). In column (2), (45) collapses to $\sigma_{sod} = \bar{\sigma}_s$ so that $\rho_{sod} = \rho_s$, for all o, d. Robust standard errors in parenthesis with levels of significance denoted by *** p < 0.01, and ** p < 0.05 and * p<0.1.

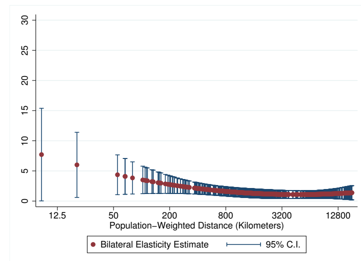
Figure I.1: Elasticities of Substitution and Distance, by sector.



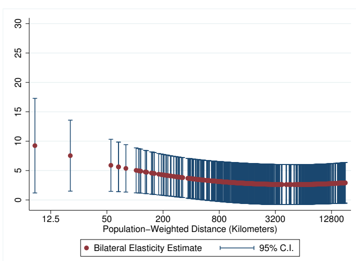
Agriculture, hunting, forestry, and fishing



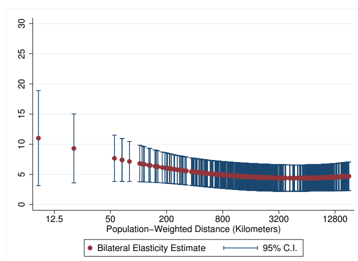
Mining and quarrying



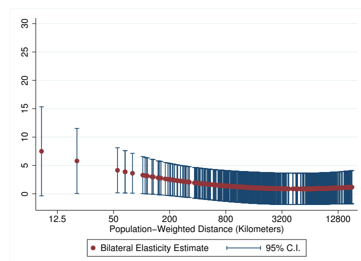
Food, beverages, and tobacco



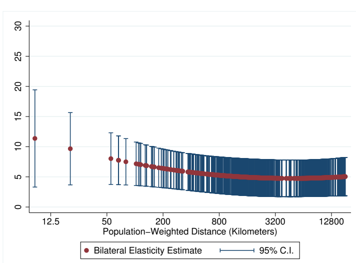
Textiles and leather



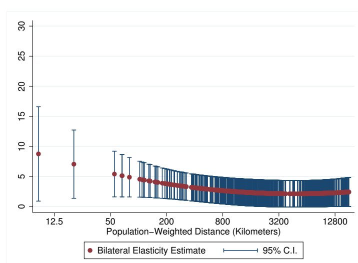
Wood and products of wood and cork



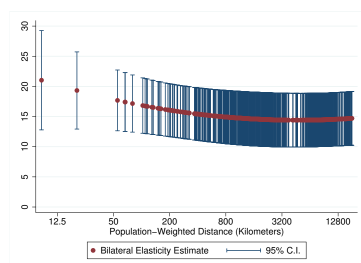
Pulp, paper, publishing and printing



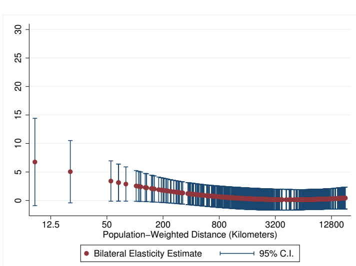
Fuel, chemicals, rubber, and plastics



Other non-metallic mineral products



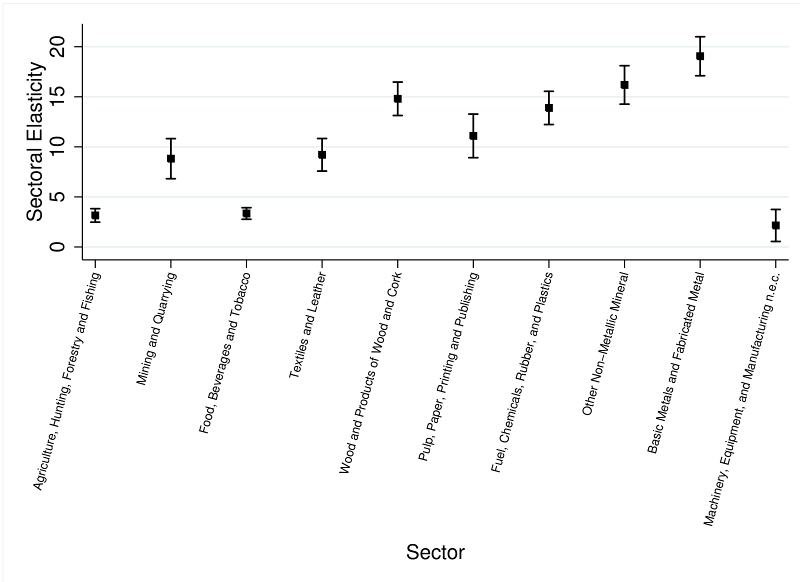
Basic metals and fabricated metal products



Machinery, equipment, and manufacturing n.e.c.

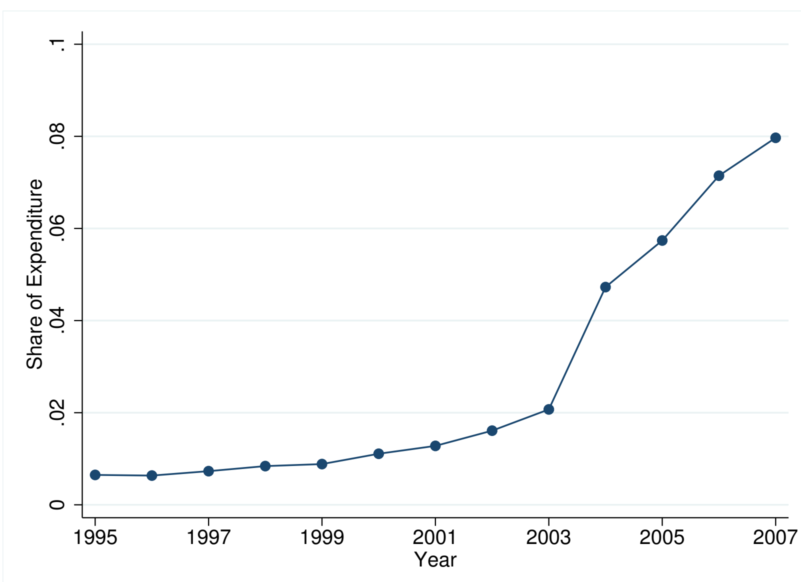
Notes: Results from estimating (46) by OLS.

Figure I.2: Elasticity of substitution, by sector. Cross-nested CES estimation.



Notes: Results from estimating (46) by OLS assuming that $\sigma_{sod} = \bar{\sigma}_s$.

Figure I.3: U.S. imports from China in "Machinery, Equipment, & Manuf n.e.c.".



Notes: Data from WIOD.