Trade with Correlation*

Nelson Lind   Natalia Ramondo

Emory University   UCSD and NBER

October 28, 2018

Abstract

We develop a trade model in which productivity presents an arbitrary pattern of correlation. The model approximates the full class of factor demand systems consistent with Ricardian theory. In particular, our framework formalizes Ricardo's insight that countries gain more from trade with partners that have relatively dissimilar technology. Incorporating this insight entails a simple correction to the sufficient-statistic approach used for macro counterfactuals, and enables a general aggregation result that links macro demand systems to micro estimates. In our quantitative application, we estimate a multi-sector trade model which captures the possibility that nearby countries may share technology, and, hence, have correlated productivity draws. Our estimates suggest that accounting for correlation is key to calculating 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 Roy Allen, Dave Donaldson, Jonathan Eaton, 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. They assume that productivity is independently distributed Fréchet across countries. This assumption leads to tractability via a max-stability property: The maximum is also Fréchet with shape θ and scale that is the sum of the scale parameters of the marginals,

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

The symmetry of this additive structure, a consequence of independence, implies that trading partners are indistinguishable. As a result, the EK model cannot capture how 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 of trade that allows for arbitrary 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. The distribution of the maximum is then

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

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, gives 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 similarity between countries influences the gains from trade.

The assumption of a max-stable multivariate Fréchet productivity distribution implies a factor demand system with expenditure shares that match choice probabilities in generalized extreme value (GEV) discrete choice models (McFadden, 1978). As a result, any trade model that generates a GEV factor demand system is observationally equivalent to a model in our framework for some specification of the correlation function. In fact, our framework captures general Ricardian theory due to an approximation result: The GEV class approximates *any* Ricardian model—without the need to restrict to Fréchet productivity distributions. Hence, any factor demand system generated by the Ricardian trade model can be approximated using a multivariate Fréchet distribution of productivity. 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. We show that within the class of GEV factor demand systems, we can calculate the gains from trade as a simple adjustment to the case of a constant-elasticity-of-substitution (CES) demand system. Specifically, we show that the results of ACR generalize, after a simple correction, to the class of models whose demand systems fit into the GEV form. In the Ricardian context, this correction adjusts a country's self-trade share to account for correlation in technology with the rest of the world, formalizing Ricardo's insight that more dissimilar countries have higher gains from trade. Additionally, for any given pattern of correlation across countries, the adjusted self-trade share is calculated using only data on expenditure shares across countries, preserving the simplicity of ACR.

We can also leverage the max-stability property to get closed-form aggregation, and to get guidance on how to use micro data to discipline macro factor demand systems. This result brings existing Ricardian models into a unifying framework. Specifically, the GEV class accommodates many disaggregate Ricardian models of

[2] While in the body of the paper we assume that preferences are CES, 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 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 multivariate θ-Fréchet. We focus on CES preferences for comparability to the standard EK framework.

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). In Section 6, we show that all these models, once aggregated at the bilateral level, can be represented by a model in which productivity is distributed multivariate Fréchet with a correlation function appropriately chosen.

In Section 7, we assess the empirical relevance of correlation. We estimate a multi-sector model of trade where correlation in productivity can depend on the proximity of countries. This spatial correlation model captures the possibility that nearby countries may share similar technology—and therefore have correlated productivity draws—by allowing for distance-dependent elasticities of substitution across countries, within sectors.

Our estimates show that correlation follows a spatial pattern: Distant countries have productivity draws with lower correlation. This empirical result has important implications for counterfactuals. Accounting for these spatial correlation patterns translates into gains from trade that can be much higher—and much more heterogenous—than the gains calculated without accounting for correlation.

This paper relates to several strands of the literature. First, we naturally make contact with the large trade literature using the Ricardian-EK framework in its various forms (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 (such as Hsieh et al., 2013), and the macro development literature (such as Lagakos and Waugh, 2013; Bryan and Morten, 2018), as well as to recent trade models used in the urban literature (such as Ahlfeldt et al., 2015; Monte et al., 2015; Caliendo et al., 2017), reviewed in Redding and Rossi-Hansberg (2017).

Second, we relate to papers in the international trade literature that use non-CES factor demand systems.[3]

[3] A related trade literature departs from CES with the goal of analyzing endogenous markups

The early work by Wilson (1980) shows how the Ricardian model in Dornbusch et al. (1977), extended to an arbitrary number of countries, can be reduced to analyzing the properties of an exchange economy—-countries trade their labor with each other. Under some restrictions on the properties of the factor demand system (i.e., gross substitutes and homotheticity), some comparative static results are derived, such as the effect of increasing tariffs from zero uniformly across goods. Results, however, are only local. Additionally, Scarf and Wilson (2005) present a Ricardian model with a demand structure that satisfies the gross substitutability property, and in which productivity follows an arbitrary probability 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.

Adao et al. (2017), building on the idea in Wilson (1980), show how to calculate macro counterfactual exercises in neoclassical trade models with invertible factor demand systems, and provide sufficient conditions for non-parametric identification using aggregate trade data. In our paper, we restrict to the GEV class—a subclass with the invertibility property. Given that restriction, our aggregation result allows us to relate various micro structures to the macro demand systems that they study, and, as a result, to use variation in disaggregate data to identify macro substitution patterns. All in all, our paper provides a bridge between the macro results of Adao et al. (2017) and estimates, common in the trade literature, based on micro data.[4]

In that regard, papers such as Caron et al. (2014), Lashkari and Mestieri (2016), 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, at showing the consequences of abandoning the assumptions that lead to linear gravity systems, and at incorporating detailed micro data to estimate key elasticities. They all notice the failure of aggregate theories to incorporate the richness of the micro data (e.g., heterogeneous price and income elasticities across

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] Our paper shares a common theme with Redding and Weinstein (2017) as they develop a framework to aggregate from micro trade transactions to macro trade and prices using the class of nested invertible demand systems.

traded goods), and "fix it" by assuming non-CES demand systems.[5] By linking various micro structures to common primitives of technology, our general framework provides guidance—given by our aggregation result—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—i.e., correlation—between countries. As a result, we can incorporate destination-origin elasticities without relying on demand-side factors and non-CES preferences.[6]

Finally, we relate 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). We make contact with this literature by introducing a *global innovation representation* for productivity. This representation characterizes multivariate Fréchet productivity as the consequence of technology adoption, and provides an economic justification for any choice of the correlation function. In this representation, an (unbounded) collection of innovations for the production of each good exists. The productivity of each innovation depends on a global component—common across countries—that captures the fundamental efficiency of the technology, and a spatial applicability component—unique to each country pair—that captures bilateral factors influencing efficiency. For each good the global component follows a Poisson process, while spatial applicability can have any distribution. With these assumptions, we apply the spectral representation theorem for max-stable processes (De Haan, 1984; Penrose, 1992; Schlather, 2002) and establish that productivity has this global innovation representation if and only if it is distributed max-stable multivariate Fréchet. This result not only provides economic primitives for productivity, but

[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, and in this way explain some "puzzles" observed in the data on trade patterns. 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 show extensions of our framework that include comparative advantage coming from demand-side factors as in the Armington model of trade (Anderson, 1979), and from entry of heterogenous firms as in the Krugman-Melitz model of trade (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.

it also gives a method to compute correlation functions from any underlying assumptions on the applicability of technology across the globe.

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 $\sigma > -1$, $C_d = \left(\int_0^1 C_d(v)^{\frac{\sigma}{\sigma+1}} \mathrm{d}v\right)^{\frac{\sigma+1}{\sigma}}$. Given total expenditure of X_d, expenditure on variety v is $X_d(v) \equiv P_d(v)C_d(v) = (P_d(v)/P_d)^{-\sigma} X_d$ where $P_d(v)$ is the price of the variety, and $P_d = \left(\int_0^1 P_d(v)^{-\sigma} \mathrm{d}v\right)^{-\frac{1}{\sigma}}$ 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 (i.e., 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. As in the original EK model, we assume perfect competition so that prices are equal to unit costs, $P_{od}(v) = c_{od}(v)$. 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 their 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. Importantly, by assuming that random productivity is country-pair specific, we are able to introduce purely Ricardian bilateral motives for 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. We first define a multivariate θ-Fréchet random vector.

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). 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. 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 with independence, is not distributed Fréchet.

To make headway without the independence assumption, 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*
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).[7] The normalization restriction provides us with notation 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.

Next, we characterize 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.

[7] 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.

Importantly, using the characterization in Lemma 1 and the homogeneity property of the correlation function, 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 given by $G\left(T_1,\ldots,T_K\right)$ and shape parameter given by θ. Evaluated at the scale 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,

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

This result is crucial for tractability in EK because it ensures that expenditure shares simply reflect the probability of importing from an origin country. Because this property holds for general multivariate θ-Fréchet random vectors, our model inherits the same tractability. Appendix A formally presents this and other properties of Fréchet random variables which we use throughout the paper.

To fix ideas, consider the special case of independent 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 4, allows us to formalize how similarity in technology matters for the gains from trade.

The next section shows how a correlation function can be constructed from fundamentals, and hence, provides an economic justification for any choice of G.

2.2 Global Innovation Representation of Productivity

We 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. Theorem 1 characterizes multivariate θ-Fréchet distributions as precisely the productivity distributions arising from global innovation.

The result is based on a technology structure that satisfies three assumptions.

Assumption 1 (Innovation Decomposition). *There exists $\theta > 0$ and, for each $v \in [0,1]$, a countable set of global innovations, $i = 1, 2, \ldots$, with global productivity $\{Z_i(v)\}_{i=1,2,\ldots}$ and spatial applicability $\{\{A_{iod}(v)\}_{o=1}^N\}_{i=1,2,\ldots}$ satisfying*

$$A_{od}(v) = \max_{i=1,2,\ldots} \left[Z_i(v) A_{iod}(v)\right]^{1/\theta}. \tag{8}$$

Assumption 2 (Independence). *$\{\{A_{iod}(v)\}_{o=1}^N\}_{i=1,2,\ldots}$ is independent of $\{Z_i(v)\}_{i=1,2,\ldots}$ and i.i.d. over $i = 1, 2, \ldots$ and $v \in [0,1]$ with $\mathbb{E}A_{iod}(v) < \infty$.*

Assumption 3 (Poisson Innovations). *The collection $\{Z_i(v)\}_{i=1,2,\ldots}$ consists of the points of a Poisson process with intensity measure $z^{-2}dz$, and is i.i.d. over $v \in [0,1]$.*

First, Assumption 1 defines a structure for technology that can be interpreted as arising from global innovation and technology adoption. For each good v, there is a countable collection of technological innovations $i = 1, 2, \ldots$ that influence the marginal product of labor. These innovations represent physical techniques (i.e., blueprints) for producing a good. For a given good v, each innovation i has a *global productivity* component, $Z_i(v)$, and an origin-destination specific *spatial applicability* component, $A_{iod}(v)$. The global productivity component measures the fundamental efficiency of technique i, and is identical across all origins and destinations. In turn, the spatial applicability component captures origin-destination specific factors that determine the efficiency of the technique when adopted at origin o to deliver goods to destination d.

Second, the key aspect of Assumption 2 is that it does not impose independence of applicability across origin countries; instead, it allows for arbitrary patterns of correlation. Moreover, the assumption does not impose any particular joint distribution for applicability; this distribution can belong to any family as long as its first moments exists. In our examples, we use Fréchet distributions for spatial applicability only because they lead to closed-form solutions.

Finally, Assumption 3 states that the global productivity component, $Z_i(v)$, follows a non homogenous Poisson process over i's, for each v. This assumption implies that the number of innovations whose global productivity satisfies $a < Z_i(v) \leq b$ is a Poisson random variable. It also implies that the distribution of $Z_i(v)$ over v, conditional on i, is Pareto with shape parameter i and lower bound $i!^{-1/i}$.[8] Loosely speaking, better innovations (i.e., with higher $Z_i(v)$) are less likely to be observed, but conditional on being observed, higher-ranked innovations (i.e., with higher i) have a Pareto distribution with higher lower bound and thinner tail (i.e., lower-variance) across goods v.

One can interpret Assumption 3 as arising from some random discovery process as in Eaton and Kortum (1999, 2010). In our static framework, we can interpret i as indexing the collection of all innovations up until the present. Rather than assuming that innovations are country specific, innovations—which represent physical methods to produce a good—are globally applicable. Origin countries differentially load on global productivity through their individual draw of spatial applicability, $A_{iod}(v)$, and adopt whichever innovation is most efficient for them.

The following theorem is a consequence of the spectral representation theorem for max-stable processes (De Haan, 1984; Penrose, 1992; Schlather, 2002).

Theorem 1 (Global Innovation Representation). *Productivity, $\{A_{od}(v)\}_{o=1}^N$, is multivariate θ-Fréchet if and only if it satisfies Assumptions 1, 2, and 3. In this case, we say that productivity has a global innovation representation.*

Let $\{A_{iod}(v)\}_{o=1}^N$ denote an underlying spatial applicability process. Then, the joint productivity distribution is

$$\mathbb{P}\left[A_{1d}(v) \leq a_1, \ldots, A_{Nd}(v) \leq a_N\right] = \exp\left[-G^d\left(T_{1d}a_1^{-\theta}, \ldots, T_{Nd}a_N^{-\theta}\right)\right], \tag{9}$$

[8] We have that $\mathbb{P}[Z_i(v) > z] = \mathbb{P}[Z_i(v)^{-1} < t]$ for $t = z^{-1}$. Then, since $\{Z_i(v)\}_{i=1,2,\ldots}$ are the points of a non homogenous Poisson process with intensity measure $z^{-2}\mathrm{d}z$, $\{Z_i(v)^{-1}\}_{i=1,2,\ldots}$ are the points of a homogenous Poisson process, and $\mathbb{P}[Z_i(v)^{-1} < t] = \sum_{j=i}^{\infty} \frac{t^j}{j!}e^{-t} = \frac{t^i}{i!} = \frac{z^{-i}}{i!}$. Therefore, $\mathbb{P}[Z_i(v) \leq z] = 1 - \left(\frac{z}{(i!)^{-1/i}}\right)^{-i}$.

with scale $T_{od} \equiv \mathbb{E}A_{iod}(v)$, for $o = 1, \ldots, N$, and correlation function

$$G^d(x_1, \ldots, x_N) \equiv \mathbb{E} \max_{o=1,\ldots,N} \frac{A_{iod}(v)}{T_{od}} x_o. \tag{10}$$

Proof. See Appendix C. □

This characterization of productivity establishes primitive assumptions on the technology structure that are necessary and sufficient for θ-Fréchet-distributed productivity across origin countries.[9] 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 technological innovations.

In fact, 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, assume that the spatial applicability of individual technologies is independent across o and distributed Fréchet with scale $\Gamma(1 - 1/\vartheta)^{-\vartheta} S_{od}$, shape $\vartheta > 1$, and Γ the gamma function. The constant on the scale ensures that productivity remains finite as $\vartheta \to 1$.

First, we compute the scale parameters of productivity. Theorem 1 establishes that the scale parameters equal the first moments of spatial applicability. By Lemma A.1, $T_{od} = S_{od}^{1/\vartheta}$.

Next, we derive the correlation function—i.e., the expectation in (10). From Lemma A.1, $(A_{iod}(v)/T_{od})x_o$ is ϑ-Fréchet with scale x_o^{ϑ}. Due to independence and max-stability, the maximum over o is also ϑ-Fréchet and its scale is the sum of the underlying scale parameters. Using Lemma A.1 to compute the expectation in (10) yields

$$G^d(x_1, \ldots, x_N) = \left(\sum_{o=1}^{N} x_o^{\frac{1}{1-\rho}} \right)^{1-\rho}, \quad \text{where} \quad \rho \equiv 1 - \frac{1}{\vartheta}. \tag{11}$$

The implied joint distribution of productivity follows from (9) in Theorem 1.

[9] It is worth noting that, contrary to the insight first in Kortum (1997), in which the distribution of the best idea belongs to the extreme value family in the limit, this result is exact.

The correlation function in (11) 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 ϑ. As $\vartheta \to 1$, 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 our assumption that applicability is independent across countries. In contrast, as $\vartheta \to \infty$, 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 for 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.

Additionally, the symmetric θ-Fréchet example is a useful building block for generating richer correlation functions, such as a cross-nested CES (CNCES) correlation function. Consider a latent (within country) technology adoption decision where applicability comes from a choice of how to apply each innovation. In particular, firms select an application m across M alternatives so that $A_{iod}(v) = \max_{m=1,\ldots,M} A_{imod}(v)$. Assume that applicability is independent across m and for each m it is ϑ-Fréchet across o with scale $\Gamma(1-1/\vartheta)^{-\vartheta} T_{mod}$ and a correlation function as in (11) with $\rho_m \in [0, 1)$.

First, we compute the scale parameters of the productivity distribution. Due to independence, the maximum of applicability over m is ϑ-Fréchet with scale $\Gamma(1 - 1/\vartheta)^{-\vartheta} \sum_{m=1}^{M} T_{mod}$. From Lemma A.1, productivity has scale $T_{od} = (\sum_{m=1}^{M} T_{mod})^{1/\vartheta}$.

Next, we compute the correlation function. By Lemma A.2, $\{(A_{imod}(v)/T_{od})x_o\}_{o=1}^{N}$ is ϑ-Fréchet with scale $T_{mod}(x_o/T_{od})^{\vartheta}$ and CES correlation function with parameter ρ_m. Due to max-stability, the maximum over o is also ϑ-Fréchet and it has scale equal to $(\sum_{o=1}^{N}(T_{mod}(x_o/T_{od})^{\vartheta})^{\frac{1}{1-\rho_m}})^{1-\rho_m}$. Due to independence, the scale of the maximum over m is simply the sum of these scale parameters. Using Lemma A.1

and focusing on the limiting case of $\vartheta \to 1$, we get $T_{od} = \sum_{m=1}^{M} T_{mod}$ and

$$G^d(x_1, \ldots, x_N) = \sum_{m=1}^{M} \left(\sum_{o=1}^{N} (\omega_{mod} x_o)^{\frac{1}{1-\rho_m}} \right)^{1-\rho_m}, \tag{12}$$

for $\omega_{mod} \equiv T_{mod}/T_{od}$. The relative efficiency of application m for origin o when delivering to destination d determines the weight of application m within the resulting CNCES correlation function. The implied productivity distribution follows from (9) in Theorem 1.

Summing up, Theorem 1 allows us to relate alternative specifications for the correlation function to primitive assumptions on the nature of technological adoption. When developing models based on Fréchet-distributed productivity, one can either use a particular specification for applicability—possibly arising from a model of innovation—and derive the implied correlation function using Theorem 1. Or, alternatively, one can directly specify a correlation function satisfying the restrictions in Definition 2. The key contribution of Theorem 1 is to provide an economic justification for any choice of the correlation function, and to provide a method for computing correlation functions implied by models of technology adoption and innovation.

2.3 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 potential import prices is shown in the next proposition.

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*[10]

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

[10] 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 $k = 1, \ldots, K$.

Proof. See Appendix D. $\square$

The joint distribution of productivity determines the joint distribution of potential import prices. For each origin o, the marginal distribution of prices, $\mathbb{P}[P_{od}(v) \leq p] = 1 - \exp\left[-T_{od}W_o^{-\theta}p^{\theta}\right]$, is a Weibull distribution with scale parameter $T_{od}W_o^{-\theta}$ and shape parameter θ. The correlation function G^d determines the dependence structure.

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

$$P_{od} = \gamma T_{od}^{-1/\theta}W_o, \tag{13}$$

with $\gamma > 0$ defined in Proposition 2. Nominal import prices are proportional to the exporter nominal wage, W_o, and decreasing in the bilateral productivity parameter, T_{od}. The scale parameters of the productivity distribution act as bilateral cost shifters, and can ve mapped 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 (13) 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 $\theta > \sigma$. Then,*

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

$$\pi_{od} = \frac{T_{od}W_o^{-\theta}G_{od}}{\sum_{o'=1}^{N} T_{o'd}W_o^{-\theta}G_{o'd}}, \tag{14}$$

where

$$G_{od} \equiv G_o^d\left(T_{1d}W_1^{-\theta}, \ldots, T_{Nd}W_N^{-\theta}\right) \quad \textit{and} \quad G_o^d\left(x_1, \ldots, x_N\right) \equiv \frac{\partial G^d\left(x_1, \ldots, x_N\right)}{\partial x_o}; \tag{15}$$

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*

$$P_d = \gamma G^d \left(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta} \right)^{-\frac{1}{\theta}}, \tag{16}$$

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

Proof. See Appendix E. □

First, the formula for the expenditure share, π_{od}, takes 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, as we explain in Section 3.

Second, using (14) and (16), correlation-adjusted expenditure shares, defined as $\pi^*_{od} \equiv \pi_{od}/G_{od}$, are CES,

$$\pi^*_{od} = T_{od} \left(\gamma \frac{W_o}{P_d} \right)^{-\theta}. \tag{17}$$

Hence, these shares constitute a gravity system, as defined by ACR, and using the definition in (13), are sufficient statistics for real import prices, $\pi^*_{od} = (P_{od}/P_d)^{-\theta}$. This interpretation will be particularly useful when we compute correlation-adjusted expenditure shares for our counterfactual analysis of Section 4.2.

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. This result follows from the property that the conditional distribution of the maximum of a multivariate Fréchet random vector is identical to its unconditional distribution, as shown in Lemma A.3.

Finally, the price level in each destination market is determined by the correlation function, G^d. In fact, the price level is determined by aggregating import price indices using the correlation function, $P_d = G^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta})^{-\frac{1}{\theta}}$. In analogy to the discrete choice literature, welfare calculations depend crucially on the specification of this function.

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 that match choice probabilities in GEV discrete choice models (McFadden, 1978). We then establish in Proposition 3 that the demand systems generated by the Ricardian model with θ-Fréchet-distributed productivity can approximate the demand systems generated by any stochastic productivity. That is, our framework is consistent with Ricardian trade, as long as the assumptions of constant returns to scale in production, competitive markets, and a single factor of production in each country are maintained.[11]

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 $\pi_{od} : \mathbb{R}_{+}^{N} \times \mathbb{R}_{+} \rightarrow [0,1]$ 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**). *A generalized extreme value (GEV) factor demand system for destination d is a factor demand system, $\{\pi_{od}^{GEV}\}_{o=1}^{N}$, such that there exists a shape parameter $\theta > \sigma$, scale parameters $\{T_{od}\}_{o=1}^{N}$, and a correlation function G^d satisfying*

$$\pi_{od}^{GEV}(\mathbf{W}, X_d) = \frac{T_{od} W_o^{-\theta} G_o^d(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta})}{\sum_{o'} T_{o'd} W_{o'}^{-\theta} G_{o'}^d(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta})}, \tag{18}$$

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

The GEV class is homothetic since expenditure shares do not depend on overall expenditure. It is closely related, as mentioned above, to the functional form for choice probabilities in GEV discrete choice models. It differs slightly, however, because the correlation function is a restricted version of the social surplus function in GEV models due to our normalization restriction in Definition 2.

An important class of models within the GEV class are CES factor demand systems, as in ACR. These models are generated by an additive correlation function,

[11] As mentioned in Footnote 2, CES preferences are not crucial for the results.

implying expenditure shares of the form

$$\pi_{od}^{\text{CES}}(\mathbf{W}, X_d) = \frac{T_{od} W_o^{-\theta}}{\sum_{o'=1}^{N} T_{o'd} W_{o'}^{-\theta}}. \tag{19}$$

The CES specification leads to a gravity system and 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 cross-nested CES (CNCES) correlation function derived in (12),

$$G^d(x_1, \ldots, x_N) = \sum_{m=1}^{M} \left(\sum_{o=1}^{N} (\omega_{mod} x_o)^{1/(1-\rho_m)} \right)^{1-\rho_m}. \tag{20}$$

The factor demand system implied by this correlation function is

$$\pi_{od}^{\text{CNCES}}(\mathbf{W}, X_d) = \sum_{m=1}^{M} \left(\frac{P_{mod}}{P_{md}} \right)^{-\frac{\theta}{1-\rho_m}} \frac{P_{md}^{-\theta}}{\sum_{m'=1}^{M} P_{m'd}^{-\theta}}, \tag{21}$$

where $P_{mod} \equiv \gamma T_{mod}^{-1/\theta} W_o$, and $P_{md} \equiv \gamma (\sum_{o=1}^{N} P_{mod}^{-\frac{\theta}{1-\rho_m}})^{-\frac{1-\rho_m}{/\theta}}$. The first fraction on the right-hand side of (21) represents the probability of importing from country o given that it chose application m. Due to correlation across countries, for a given m, this conditional probability has an elasticity of substitution of $\theta/(1-\rho_m)$. The second fraction on the right-hand side of (21) represents the probability of importing m-goods into d. Due to independenceacross m's, the elasticity of substitution is simply θ.

It is clear from comparing (19) with (21) that the GEV class generates a richer pattern of substitution across exporters than CES. In general, if we compare the elasticity of demand in the GEV and CES models, around any observed expenditure shares, the difference comes from the correlation function,

$$\frac{\partial \ln \pi_{od}^{\text{GEV}}}{\partial \ln W_o} = \frac{\partial \ln \pi_{od}^{\text{CES}}}{\partial \ln W_o} + \frac{\partial \ln G_{od}}{\partial \ln W_o} \quad \text{and} \quad \frac{\partial \ln \pi_{od}^{\text{GEV}}}{\partial \ln W_{o'}} = \frac{\partial \ln \pi_{od}^{\text{CES}}}{\partial \ln W_{o'}} + \frac{\partial \ln G_{od}}{\partial \ln W_{o'}}, \tag{22}$$

with $\partial \ln \pi_{od}^{\text{CES}} / \partial \ln W_o = -\theta (1 - \pi_{od})$, and $\partial \ln \pi_{od}^{\text{CES}} / \partial \ln W_{o'} = \theta \pi_{o'd}$, since $\frac{\partial \ln P_d}{\partial \ln W_o} = \pi_{od}$. While the elasticity with respect to W_o is always negative and more sensitive in the GEV model, the cross elasticity is either negative or positive depending on the strength of the substitution effect coming from correlation.

Our next result states that the GEV factor demand system in (18) 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 a 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 cross-nested CES factor demand systems, as we show in Section 6.[12]

We can 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: The set of factor demand systems generated by any Ricardian model—without restricting to θ-Fréchet productivity distributions—can be approximated arbitrarily well by the class of GEV factor demand systems.

Proposition 3 (**GEV Approximation**). *Let $\{A_{od}(v)\}_{o=1}^{N}$ have any multivariate distribution whose marginals have finite moment of order σ. 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*

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

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

The key implication of Proposition 3 is that *any* factor demand system generated by the Ricardian trade model can be approximated by a Ricardian trade model where productivity has a multivariate θ-Fréchet distribution. Put simply, through

[12] It is worth noting that this subclass can approximate any GEV factor demand system (Fosgerau et al., 2013) and, by extension, any GEV Ricardian model.

this approximation result, our framework encompasses the full macroeconomic implications of Ricardian trade theory.

4 Macro Counterfactuals

We next show that heterogeneity in correlation leads to heterogeneity in the gains from trade and that this heterogeneity affects the calculation of any (counterfactual) departure from the current equilibrium. It turns out that calculations using a GEV 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 of the gains from trade.

From (14), the self-trade share is

$$\pi_{dd} = \frac{T_{dd} W_d^{-\theta} G_{dd}}{\sum_{o=1}^{N} T_{od} W_o^{-\theta} G_{od}}. \tag{23}$$

Using the expression for the price index in (16), we can write the real wage in country d as

$$\frac{W_d}{P_d} = \gamma^{-1} T_{dd}^{\frac{1}{\theta}} \left(\pi_{dd}^*\right)^{-\frac{1}{\theta}}, \tag{24}$$

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

Let $\hat{x} \equiv x'/x$ denote the change from x to x' in an equilibrium outcome due to some change in the model's parameters. Using (24), it is straightforward to show that the change in real wages between two equilibria is given by

$$\frac{\hat{W}_d}{\hat{P}_d} \equiv \frac{W_d'/P_d'}{W_d/P_d} = \left(\hat{\pi}_{dd}^*\right)^{-\frac{1}{\theta}}. \tag{25}$$

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 given by θ.[13]

[13] 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.

4.1 Gains From Trade: Autarky

What are the consequences of correlation in technology for the gains from trade relative to autarky? Intuitively, if two countries have perfectly correlated productivity draws across varieties, they will offer each other identical prices across varieties, 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 $G_{dd} = 1$—i.e., correlation with other countries is irrelevant in autarky. Real wages in autarky are simply

$$\left(\frac{W_d}{P_d}\right)^{\text{Autarky}} = \gamma^{-1} T_{dd}^{\frac{1}{\theta}}. \tag{26}$$

Dividing (24) by (26), the *gains from trade relative to autarky* are

$$GT_d \equiv \frac{W_d/P_d}{(W_d/P_d)^{\text{Autarky}}} = \left(\frac{\pi_{dd}}{G_{dd}}\right)^{-\frac{1}{\theta}}. \tag{27}$$

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 (27) simplify to the ones in ACR where two countries with the same self-trade share have the same gains from trade relative to autarky.

The expression for gains in (27) admits the possibility that if two countries have the same self trade share, but one country has very similar technology to all other countries—i.e., high correlation—their gains from trade will be smaller. In contrast, if that country has dissimilar technology to other countries—i.e., low correlation—their gains from trade will be larger. In this way, our general 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

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

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[14]

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

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—i.e., 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.

4.2 Calculating the Correlation Correction

To make the necessary adjustment for correlated technology, we need to know the correlation structure across countries, which requires estimates of G^d. Given the correlation function, we can then calculate the gains from trade directly from expenditure data, generalizing the sufficient-statistic approach in ACR.

Because the demand system is CES after the correlation correction, correlation-adjusted shares are sufficient statistics for bilateral import prices. The procedure to compute correlation-adjusted expenditure shares amounts to inverting the demand system in (14).

[14] $G_{od} = \left((T_{1d}W_1^{-\theta})^{1/(1-\rho)} + (T_{2d}W_2^{-\theta})^{1/(1-\rho)}\right)^{-\rho} (T_{od}W_o^{-\theta})^{\rho/(1-\rho)}$ 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, $G_{3d} = 1$.

Using the definition of import price index in (13), and the homogeneity of degree zero of G_o^d, expenditure shares in (14) 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 (17) yields the system

$$\pi_{od} = \pi_{od}^* G_o^d \left(\pi_{1d}^*, \ldots, \pi_{Nd}^*\right) \quad \text{for} \quad o = 1, \ldots, N. \tag{28}$$

Given expenditure share data and the correlation function of a single destination, (28) constitute a system of N equations in the N unknown correlation-adjusted expenditure shares across origins.[15] As a result, we only need expenditure share data to calculate the gains from trade.

5 Aggregation

This section provides an aggregation result that links the aggregate trade model we have studied so far to trade models with more disaggregated structures, such as sectors and global value chains.

This aggregation result is important because it allows us to link the macro results of Adao et al. (2017) and estimates based on micro data that are common in the trade literature. In their paper, they show how to estimate non-parametrically invertible demand systems using only aggregate trade data. In practice, however, as they recognize, they need to proceed parametrically (e.g., mixed CES) and estimate the factor demand system without taking a stance on any particular microfoundation. By being less general and restricting to the GEV class of factor demand systems—a sub-class of invertible demand systems with the gross substitute property—our aggregation result enables us to identify macro substitution patterns using disaggregate data (e.g., sectoral data). We provide concrete applications in the next section.

Our aggregation result follows directly from Definition 1: Max-linear combina-

[15] 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 (28), 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).

tions of multivariate θ-Fréchet random vectors are multivariate θ-Fréchet. This property implies that we can aggregate models built on optimizing behavior and θ-Fréchet productivity to get equivalent macro models where productivity is also θ-Fréchet. Put differently, one can always think of macro-level scale parameters and macro-level correlation functions as the result of the micro-level scales and correlation function. The following proposition formalizes this result.

Proposition 4 (Aggregation of the Productivity Process). *Consider a model with M_o micro factors within each origin country $o = 1, \ldots, N$. Let productivity $\{\{A_{mod}(v)\}_{m=1}^{M_o}\}_{o=1}^{N}$ be distributed multivariate θ-Fréchet with $\theta > \sigma$. Denote the associated scale parameters by $\{\{T_{mod}\}_{m=1}^{M_o}\}_{o=1}^{N}$ and correlation function by $F^d : \mathbb{R}_+^{M_1} \times \cdots \times \mathbb{R}_+^{M_N} \to \mathbb{R}_+$.*

Then, aggregate productivity $A_{od}(v) = \max_{m=1,\ldots,M_o} A_{mod}(v)$ is also distributed multivariate θ-Fréchet. The aggregate scale parameters are, for each $o = 1, \ldots, N$, given by

$$T_{od} = F^d\left(\mathbf{0}_1, \ldots, \mathbf{0}_{o-1}, \mathbf{T}_{od}, \mathbf{0}_{o+1}, \ldots, \mathbf{0}_N\right) \equiv F^{od}(\mathbf{T}_{od}), \tag{29}$$

where $\mathbf{0}_o$ is the zero vector of length M_o, and $\mathbf{T}_{od} \equiv (T_{1od}, \ldots, T_{M_o od})$. The aggregate correlation function is given by

$$G^d(x_1, \ldots, x_N) = F^d(\mathbf{\Omega}_{1d} x_1, \ldots, \mathbf{\Omega}_{Nd} x_N), \tag{30}$$

where, for each $o = 1, \ldots, N$, $\mathbf{\Omega}_{od} \equiv (\omega_{1od}, \ldots, \omega_{M_o od})$ is a vector of aggregation weights with elements $\omega_{mod} \equiv T_{mod}/T_{od}$ for $m = 1, \ldots, M_o$.

Proof. This result follows from the max-stability property. See Appendix G. $\square$

Proposition 4 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. The link between the micro and macro levels comes from maximizing productivity across m's within an origin country.

This aggregation result produces aggregate expenditure shares belonging to the GEV class.

Corollary 2 (Aggregation of Expenditure Shares). *Micro expenditure shares are*

$$\pi_{mod} = \frac{T_{mod} W_o^{-\theta} F_{mo}^d\left(\mathbf{T}_{1d} W_1^{-\theta}, \ldots, \mathbf{T}_{Nd} W_N^{-\theta}\right)}{F^d\left(\mathbf{T}_{1d} W_1^{-\theta}, \ldots, \mathbf{T}_{Nd} W_N^{-\theta}\right)} \quad \text{for} \quad m = 1, \ldots, M_o \quad \text{and} \quad o = 1, \ldots, N,$$

where $\mathbf{T}_{od} \equiv (T_{1od}, \ldots, T_{M_o od})$, *and* $F^d_{mo}(\mathbf{X}_1, \ldots, \mathbf{X}_N) \equiv \frac{\partial}{\partial x_{mo}} F^d(\mathbf{X}_1, \ldots, \mathbf{X}_N)$.

Then, aggregate expenditure shares are

$$\pi_{od} \equiv \sum_{m=1}^{M_o} \pi_{mod} = \frac{T_{od} W_o^{-\theta} G^d_o(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta})}{G^d(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta})} \quad \textit{for} \quad o = 1, \ldots, N,$$

where T_{od} *and* G^d *are given, respectively, by* (29) *and* (30) *in Proposition* 4.

This result allows us to pass seamlessly between the micro and macro levels. Its natural consequence is that we can use disaggregate data (e.g., sectoral data) to estimate the micro correlation function and micro scale parameters. We can then apply Proposition 4 to derive the macro correlation function and macro scale parameters, and perform macro counterfactual analysis using the results in Section 4. Thanks to this aggregation result, we can connect (Ricardian) micro foundations to macro substitution patterns, as our applications in the next section show.

Aggregation weights, ω_{mod}, are key for the aggregation procedure. Because correlation-adjusted expenditure is proportional to the scale parameters of both the micro and macro models, these weights can be recovered by computing correlation-adjusted expenditure from disaggregate expenditure shares.

The logic for computing correlation-adjusted trade shares at the micro level follows the derivations in Section 4.2. Gravity holds after correcting for correlation at the micro level,

$$\pi^*_{mod} \equiv \frac{\pi_{mod}}{F_{mod}} = T_{mod} \left(\gamma \frac{W_o}{P_d} \right)^{-\theta}, \tag{31}$$

where $F_{mod} \equiv F^d_{mo}\left(\mathbf{T}_{1d} W_1^{-\theta}, \ldots, \mathbf{T}_{Nd} W_N^{-\theta}\right)$. We can then recover micro correlation-adjusted expenditure shares by solving $\pi_{mod} = \pi^*_{mod} F^d_{mo}\left(\mathbf{\Pi}^*_{1d}, \ldots, \mathbf{\Pi}^*_{Nd}\right)$ where $\mathbf{\Pi}^*_{od} \equiv (\pi^*_{1od}, \ldots, \pi^*_{M_o od})$. Because correlation-adjusted shares are sufficient statistics for micro import prices, this step amounts to inverting the demand system.

We can then compute the correlation-adjusted expenditure shares in the implied macro model. Using (29), (31), and the homogeneity of F^{od}, yields

$$\pi^*_{od} = F^{od}(\mathbf{\Pi}^*_{od}), \quad \text{and} \quad \omega_{mod} = \frac{\pi^*_{mod}}{\pi^*_{od}}. \tag{32}$$

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

6 Applications

We now present applications in the literature 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 illustrate the aggregation result of the previous section. These applications, by belonging to the GEV class, fall into the class of factor demand systems with the gross substitutes property; as such, complementarities are precluded.[16]

6.1 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 with sectoral shares μ_{sd}, and 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 total expenditure on goods in sector s from origin o is

$$\pi_{sod} = \frac{\left(\tau_{sod}\frac{W_o}{A_{so}}\right)^{-\theta_s}}{\sum_{o'=1}^{N}\left(\tau_{so'd}\frac{W_{o'}}{A_{so'}}\right)^{-\theta_s}}\mu_{sd}, \tag{33}$$

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 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 nested CES correlation

[16] 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.

function,

$$F(\mathbf{X}_1,\ldots,\mathbf{X}_N) = \sum_{s=1}^{S}\left(\sum_{o=1}^{N} x_{so}^{1/(1-\rho_s)}\right)^{1-\rho_s}, \tag{34}$$

with $\mathbf{X}_o \equiv (x_{1o},\ldots,x_{So})$. The parameter ρ_s measures the degree of correlation across origin countries in each sector s.

Expenditure shares at the sector level constitute a gravity system given by

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

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 index in country d, $P_d = \left(\sum_s P_{sd}^{-\theta}\right)^{-1/\theta}$.

This demand system matches (33) for $\theta/(1-\rho_s) = \theta_s$, $P_{sod}/\gamma = \tau_{sod}W_o/A_{so}$, and $\mu_{sd} = (P_{sd}/P_d)^{-\theta}$. Sectoral shares now 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).[17,18]

We can get an equivalent macro model by applying Proposition 4. Aggregate 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 as derived in (12),

$$G^d(x_1,\cdots,x_N) = \sum_{s=1}^{S}\left(\sum_{o=1}^{N}(\omega_{sod}x_o)^{1/(1-\rho_s)}\right)^{1-\rho_s}, \tag{35}$$

with $\omega_{sod} = T_{sod}/T_{od}$.[19] The inner sum indicates that latent sectors induce correlation across origins. The parameter ω_{sod} measures the extent to which sector s matters for trade flows from o to d—i.e., it reflects sectoral trade costs and comparative advantage.

In turn, aggregate expenditure shares constitute a GEV factor demand system, as

[17] 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.

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

[19] As the number of micro factors—the M dimension in (12) and the S dimension here—gets large, the CNCES specification can arbitrarily approximate the mixed-CES factor demand system used in Adao et al. (2017). In this way, their empirical application can be interpreted as arising from some disaggregate Ricardian model.

in (21),

$$\pi_{od} \equiv \sum_s \pi_{sod} = \sum_s \frac{\left(\omega_{sod} T_{od} W_o^{-\theta}\right)^{\frac{1}{1-\rho_s}}}{\sum_{o'} \left(\omega_{so'd} T_{o'd} W_{o'}^{-\theta}\right)^{\frac{1}{1-\rho_s}}} \frac{\left[\sum_{o'} \left(\omega_{so'd} T_{o'd} W_{o'}^{-\theta}\right)^{\frac{1}{1-\rho_s}}\right]^{1-\rho_s}}{\sum_{s'} \left[\sum_{o'} \left(\omega_{s'o'd} T_{o'd} W_{o'}^{-\theta}\right)^{\frac{1}{1-\rho_{s'}}}\right]^{1-\rho_{s'}}}. \tag{36}$$

The aggregate model with a CNCES correlation function has the same implications as the multi-sector model with correlation across origin countries within each sector.

6.2 Multinational Production

Assume that productivity depends on the home country j of a firm. The micro correlation function is CNCES as in (34), and the implied macro correlation function is as in (35) 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

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

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 (36) with j replacing s.

This factor demand system matches the one in Ramondo and Rodríguez-Clare (2013) for $\rho_j = \rho$ and $T_{jod}^{1/\theta} = h_{jo} A_j$. That is, if we assume that the distribution of productivity $A_{od}(v)$ has a correlation function given by (35) with j replacing s, our Ricardian trade model matches, at the bilateral level, a Ricardian trade model with multinational production.

6.3 Multiple Regions

Assume that each country o is composed of multiple regions, given by the set R_o. Denote productivity in region r by $A_{rod}(v)$. Productivity across regions and countries is multivariate θ-Fréchet with scale parameter T_{rod} and correlation function

$$F^d(\mathbf{X}_1, \ldots, \mathbf{X}_N) = \sum_o \left(\sum_{r \in R_o} x_{ro}^{1/(1-\rho_o)} \right)^{1-\rho_o}, \tag{37}$$

with $\mathbf{X}_o \equiv (x_{1o}, \ldots, x_{R_o o})$. The expression in (37) indicates that productivity draws are independent across countries and correlated within a country, across regions, with the degree of correlation given by the parameter ρ_o.

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

$$\pi_{rod} = \left(\frac{P_{rod}}{P_{od}} \right)^{-\frac{\theta}{1-\rho_o}} \frac{P_{od}^{-\theta}}{\sum_{o'} P_{o'd}^{-\theta}} \quad \text{with} \quad P_{od} = \left(\sum_{r' \in R_o} P_{r'od}^{-\frac{\theta}{1-\rho_o}} \right)^{-\frac{1-\rho_o}{\theta}}. \tag{38}$$

The first fraction on the right-hand side of (38) is the probability of importing from region r in country o conditional on importing from some region in country o, while the second fraction is the probability of importing from country o into d.

Aggregate productivity is multivariate θ-Fréchet with scale $T_{od} = \left(\sum_{r \in R_o} T_{rod}^{1/(1-\rho_o)} \right)^{1-\rho_o}$, and with an additive correlation function. As a result, the factor demand system is CES,

$$\pi_{od} = \sum_{r \in R_o} \pi_{rod} = \frac{T_{od} W_o^{-\theta}}{\sum_{o'} T_{o'd} W_{o'}^{-\theta}}.$$

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

6.4 Global Value Chains

We now show that the model of global value chains in Antràs and de Gortari (2017) is equivalent to 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

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

with $\alpha_k > 0$ and $\sum_{k=1}^{K-1} \alpha_k < 1$. The unit cost of good v is $c_\ell(\mathbf{W})/A_{\ell d}(v)$. 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

$$\pi_{\ell d} = \frac{T_{\ell d} c_\ell(\mathbf{W})^{-\theta}}{\sum_{\ell'} T_{\ell' d} c_{\ell'}(\mathbf{W})^{-\theta}}. \tag{39}$$

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

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

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

6.5 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

$$c_o(\mathbf{W}) = B W_o^\beta P_o^{1-\beta},$$

where $0 < \beta \leq 1$ is the share of labor in production, B is a positive constant, $\mathbf{W} \equiv [W_1, \ldots, W_N]$, 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

$$\pi_{od} = \frac{A_o \tau_{od}^{-\theta} c_o(\mathbf{W})^{-\theta}}{\sum_{o'} A_{o'} \tau_{o'd}^{-\theta} c_{o'}(\mathbf{W})^{-\theta}}. \tag{40}$$

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.

7 Quantitative analysis

This section quantifies the gains from trade when productivity is correlated across space. We consider a sectoral version of our Ricardian model of trade. Based exclusively on supply factors, this sectoral model allows for bilateral heterogeneity in trade elasticities.[20] These substitution patterns are obtained because we allow for the distribution of productivity to depend on characteristic of the country where goods get produced and of the destination market where goods get delivered.

The choice of the multi-sector Ricardian model of trade not only allows us to assess

[20] 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.

the importance of correlation in the calculation of the gains from trade and other counterfactuals, but it also allows us to highlight the importance of using disaggregate data to estimate elasticities that are key for those counterfactuals. Additionally, we choose this model because it includes as special cases two benchmark models in the trade literature: the CES sectoral model and the cross-nested CES (CNCES) sectoral model. These cases arise when bilateral factors do not influence correlation.

7.1 Specification

Denote each sector by $s = 1, \ldots, S$. We assume that sector-level productivity is multivariate θ-Fréchet with scale T_{sod} and correlation function

$$F^d(\mathbf{X}_1, \cdots, \mathbf{X}_N) = \sum_{s=1}^{S} H^{sd}(x_{s1}, \cdots, x_{sN}). \tag{41}$$

In turn, the function H^{sd} is a correlation function capturing correlation within sector s and destination market d, across origin countries. Following Hanoch (1975) and Sato (1977), this function is defined implicitly by

$$1 = \sum_{o=1}^{N} \left(\frac{x_{so}}{H^{sd}(x_{s1}, \ldots, x_{sN})} \right)^{\frac{1}{1-\rho_{sod}}}. \tag{42}$$

The parameter ρ_{sod} measures correlation across origins within sector s and destination d. The resulting sectoral expenditure shares are iso-elastic,

$$\pi_{sod} = \frac{\sigma_{sod} \left(\frac{P_{sod}}{P_{sd}} \right)^{-\sigma_{sod}}}{\sum_{o'=1}^{N} \sigma_{so'd} \left(\frac{P_{so'd}}{P_{sd}} \right)^{-\sigma_{so'd}}} \mu_{sd}, \tag{43}$$

where $\sigma_{sod} \equiv \frac{\theta}{1-\rho_{sod}}$ is the expenditure elasticity of substitution induced by bilateral correlation, μ_{sd} is the share of expenditure by country d on sector s goods, $\mu_{sd} = (P_{sd}/P_d)^{-\theta}$, and the sectoral price index P_{sd} is defined implicitly as

$$1 = \sum_{o} \left(\frac{P_{sod}}{P_{sd}} \right)^{-\sigma_{sod}}, \tag{44}$$

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 $\sigma_{sod} = \theta$. In this case, there is a single trade elasticity common to all trade flows. When $\rho_{sod} = \rho_s$, for all o, d, we get the sectoral CNCES model without bilateral spatial correlation (as in Section 6).

Next, we compute correlation-adjusted sectoral expenditure shares and the gains from trade in closed form. As we explain in Section 5, because correlation-adjusted shares are sufficient statistics for real sectoral import prices, we can calculate the correlation correction by inverting the demand system. We first solve for within-sector relative prices. Dividing (43) by σ_{sod}, summing over origins, and using (44) gives $(P_{sod}/P_{sd})^{-\sigma_{sod}}$. Further using sectoral shares, μ_{sd}, we get

$$\pi^*_{sod} = \underbrace{\left(\frac{\pi_{sod}/\sigma_{sod}}{\sum_{o'=1}^{N} \pi_{so'd}/\sigma_{so'd}} \right)^{\theta/\sigma_{sod}}}_{(P_{sod}/P_{sd})^{-\sigma_{sod}}} \times \underbrace{\mu_{sd}}_{(P_{sd}/P_d)^{-\theta}} . \tag{45}$$

Our aggregation result in (32) and independence across sectors implies that the function F^{od} is additive, $\pi^*_{od} = \sum_s \pi^*_{sod}$. Using (27), the gains from trade are

$$GT_d = \left(\sum_s \pi^*_{sod} \right)^{-1/\theta}, \tag{46}$$

which can be computed using sectoral expenditure data, once we have estimates of σ_{sod} and θ.

7.2 Estimation

We estimate the multi-sector model with spatial correlation using two sequential gravity regressions. The first step regression uses variation in trade flows and tariffs across origin countries within each sector to identify geographic correlation in productivity. The second step regression uses variation across sectors and destination markets to identify θ.

Letting t index years, we impose additional assumptions on the structure of trade costs and spatial correlation patterns.

First, trade costs depend on gravity covariates, Geo_{od}, and on tariffs, t_{sodt},

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

We include in Geo_{od} variables such as distance and time differences between trading partners, and 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.

Second, we assume that the sectoral elasticity is additively separable into a sector component and a spatial component which we proxy by a non-linear function of bilateral distance,

$$\sigma_{sod} = \bar{\sigma}_s + \tilde{\sigma}_1 \text{Dist}_{od} + \tilde{\sigma}_2 \text{Dist}^2_{od}. \tag{47}$$

This assumption is motivated by the literature that documents that technology diffusion follows a spatial pattern.[21]

Using (43), we get our first-step within-sector gravity equation,

$$\ln \pi_{sodt} = S_{sot} + D_{sdt} + B_{sod} - \bar{\sigma}_s \ln(1+t_{sodt}) + \left(\alpha_{sot} + \beta_{sdt} - \ln(1+t_{sodt})\right)\left(\tilde{\sigma}_1 \text{Dist}_{od} + \tilde{\sigma}_2 \text{Dist}^2_{od}\right) + u_{sodt}, \tag{48}$$

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 distance and distance squared with tariffs allow us to estimate distance-dependent elasticities of substitution across origin-destination pairs. The exclusion restriction for identification is that variation in tariffs across origin countries is exogenous conditional on the covariates included in (48).

To estimate the between-sector elasticity, we use a second step regression that relies on variation across sectors and on inferred within-sector relative prices from the first step. This second regression comes from the condition for destination d's expenditure on sector-s goods, $\mu_{sd} = (P_{sd}/P_d)^{-\theta}$. First, note that

$$\ln \frac{P_{sd}}{P_d} = \ln \frac{P_{sd}}{P_{sod}} + \ln \frac{P_{sod}}{P_d} = -\ln \frac{P_{sod}}{P_{sd}} + \ln \frac{W_o}{A_{so}} - \ln P_d + \ln \tau_{sod}.$$

[21] 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.

To construct an estimating equation from this result, we need estimates of P_{sod}/P_{sd}. Specifically, given estimates of σ_{sod}, we combine (43) with the definition of the sectoral price index in (44) to calculate

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

where $\hat{\sigma}_{sod}$ is our estimate for σ_{sod} from the first step.

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

$$\ln \mu_{sdt} = a_{sot} + b_{dt} + \theta \ln\left(\widehat{P_{sodt}/P_{sdt}}\right) - \theta\delta_s' \text{Geo}_{od} - \theta \ln(1 + t_{sodt}) + v_{sodt}, \tag{49}$$

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.

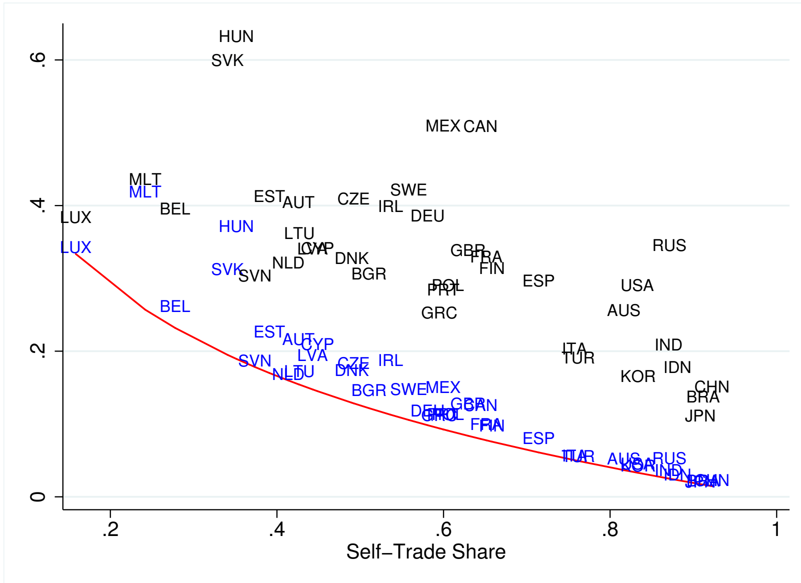
We estimate (48) and (49) by Ordinary Least Squares (OLS). We use sectoral tariff data constructed by aggregating 4-digit SITC tariff data, from COMTRADE, and sectoral trade flow data, from the World Input-Output Database (WIOD). Appendix H describes the data construction and sample restrictions in detail.

We consider three models. First, we estimate the CES model where productivity is independent across sectors and origin countries ($\rho_{sod} = 0$). Because $\sigma_{sod} = \theta$, estimation of the first step gives us an estimate of θ. Second, we estimate the CNCES model where correlation is not bilateral, but rather, symmetric across origin and destination countries within each sector—that is, ρ_{sod} is common across o and d within s, so that $\sigma_{sod} = \bar{\sigma}_s$. Finally, we allow for bilateral spatial correlation and estimate our model with σ_{sod} specified in (47). Note that the CES model is nested within the CNCES model and the CNCES model is nested within the spatial model.

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.[22]

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

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}.

Appendix Table I.1 presents our OLS estimates of θ. Column 1 shows an estimate of $\theta = 5.523$ from the CES model, in line with values in the literature (see Simonovska and Waugh, 2013). Column 2 shows that $\theta = 0.607$ after we account for symmetric correlation across space within each sector (CNCES). Finally, column 3 presents the estimate of $\theta = 0.489$ when we allow for distance-dependent sectoral elasticities.

We next use the estimates of the correlation function and the parameter θ to perform various counterfactual exercises.

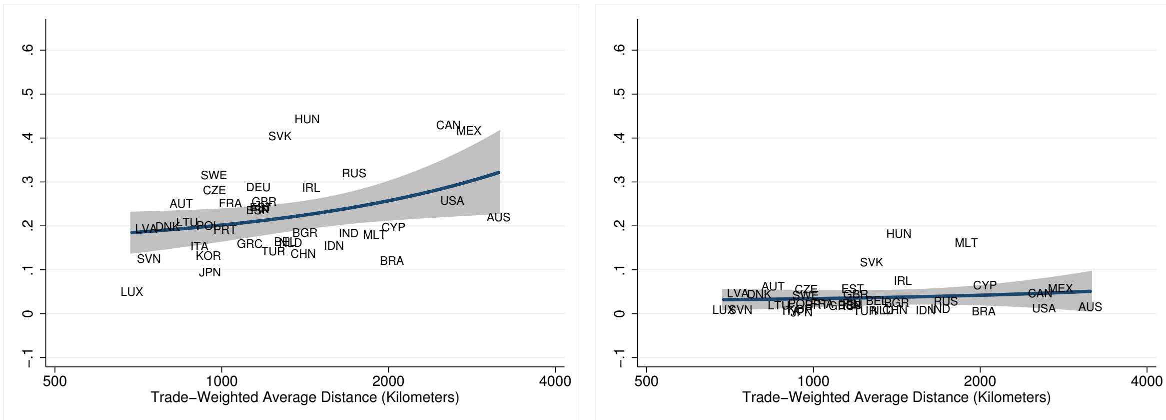
7.3 The Gains from Trade

Figure 1 shows the gains from trade calculated using the model with no correlation (CES), with within-sector symmetric correlation (cross-nested CES, CNCES), and with bilateral spatial correlation, respectively.[23] The figure shows not only that differences in (log) levels are large across the three models for all countries, but, also and more importantly, differences are large across countries with similar self-trade shares. For instance, Mexico, Germany, and Poland have a similar self-trade share of around 60 percent. However, once we account for either symmetric

[23] Results for years prior to 2007 are relegated to the Online Appendix.

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.

spatial correlation or bilateral spatial correlation, their gains are significantly different. In contrast, the CES model would predict equal gains from trade for these three countries. In turn, the CNCES model delivers gains from trade that are more heterogeneous than the ones from CES, but much less heterogenous than the ones coming from the spatial model.

Perhaps not surprisingly, Figure 2 shows that countries that are on average further away from their trading partners have higher gains from trade in the spatial model, while the CNCES model—with sectoral elasticities that are constant over space—fails to do that. Through the lens of our spatial model, countries further away from each other have more dissimilar technologies—i.e., lower sectoral elasticities—and, hence, higher gains from trade. This result captures the idea that more dissimilar countries have higher gains from trade, as proposed by Ricardo.

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

Next, we consider the implications of our spatial model regarding various counterfactual scenarios, and compare them with the implications from the CNCES model. We use the procedure outlined in Section 4.

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 spatial model and the CNCES model. The difference between the two models shows the effect of accounting for bilateral correlation in addition to within-sector correlation. Differences in the predicted real wages can be large, particularly for large changes in trade costs.

Why does correlation matter—and even more for large changes in trade costs? The gains from trade come from two potentially offsetting effects: 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,

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

In contrast, the wage effect is indirect and operates through general equilibrium effects. The market clearing condition for the United States is

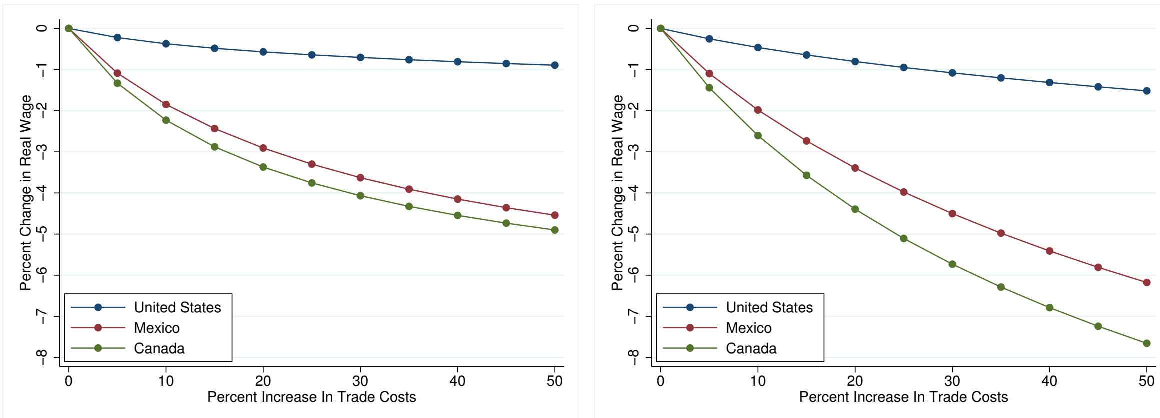
$$W_{USA} L_{USA} = \sum_{d=1}^{N} \frac{P_{USA,d}^{-\theta} G_{USA}^{d}(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta})}{G^{d}(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta})} X_d.$$

If the change in trade costs with Mexico induces U.S. consumers to substitute expenditure away from Mexican goods and towards U.S. goods, the right-hand side of this condition—that captures 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 (i.e., high correlation), then U.S. and Mexican goods would be substitutable and labor demand in the United States would be sensitive to trade costs for imports from Mexico. In this case, 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 of their goods. As a result, the offsetting wage effect would be smaller in the spatial model. As we see in Figure 3, accounting for spatial correlation in technology leads to an increase in the losses from increasing U.S. trade costs with NAFTA partners. Notice, also, that 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

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.

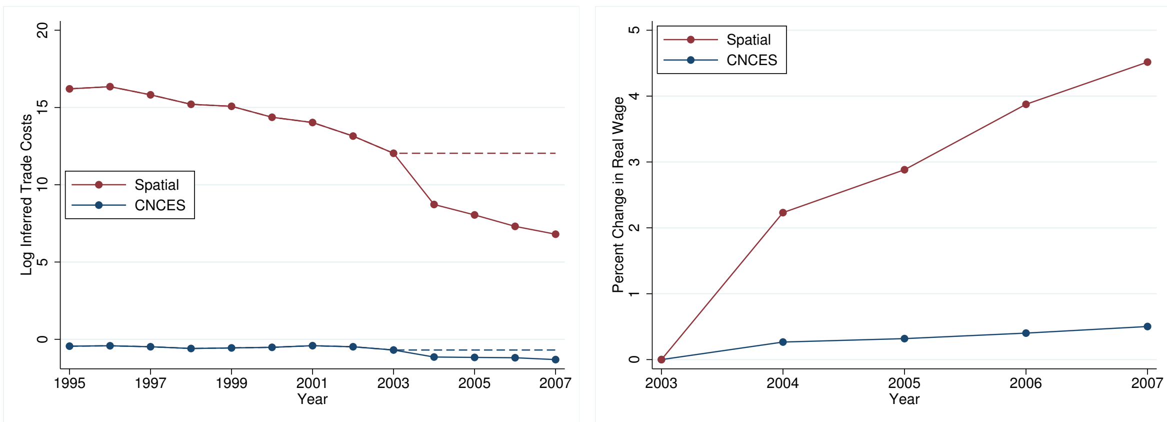
price effect—with an elasticity equal to the expenditure share—does not change between the CNCES and spatial models.

Second, we consider the implications of the rise of Chinese manufacturing imports for the real wage in the United States. To proceed, 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 (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).[24] 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 both 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.

[24] 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.

Finally, we consider 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 case of the spatial model where we account for bilateral correlation in technology and the case of 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. Another example of a ranking reversal between the two models occurs with Korea and Great Britain: While the spatial model puts Korea in seventh place and Great Britain in eighth place, the CNCES model predicts the opposite. The intuition for these unilateral protectionism exercises is similar to the other counterfactuals: the direct price effect is measured by observed expenditure shares, which are constant across models, while the indirect wage effect depends on substitutability of goods across trading partners and, therefore, depends on patterns of spatial correlation.

All these counterfactual exercises illustrate that the welfare implications of changes in trade costs can change substantially once we account for spatial correlation in productivity. These results reflect Ricardo's insight that differences in technological similarity across trade partners matter for the gains from trade.

8 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 arbitrary patterns of correlation in technology between countries. We start from technology primitives that generate multivariate Fréchet productivity with a general correlation structure. Even so, we retain all the tractability of EK-type tools. Importantly, our structure generates the class of GEV factor demand systems and, as such, approximates *any* Ricardian model—not only the one with Fréchet-distributed productivity.

The gains from trade coming from a GEV factor demand system can be written as a simple correction to self-trade shares. Moreover, the theory, by relating macro substitutability patterns to underlying micro structures, provides guidance on incorporating standard micro estimates into macro counterfactual exercises.

Our quantitative application to a multi-sector model of trade reveals that accounting for correlation matters: 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. Technical Report 4.

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.

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

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.

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.

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.

Redding, S. J. and D. E. Weinstein (2017). Aggregating from micro to macro patterns of trade. *Mimeo, Princeton University*.

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.

Simonovska, I. and M. Waugh (2013). The elasticity of trade: Estimates and evidence. *Journal of International Economics 92*, 34–50.

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 location $T > 0$ and shape $\alpha > 0$. Then if $\alpha > 1$, $\mathbb{E}[X] = \Gamma(1 - 1/\alpha)T^{1/\alpha}$. Also, for any $S > 0$ and $\beta > 0$, $(S^{1/\alpha}X)^\beta$ is Fréchet with location ST and shape α/β.*

Proof.

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

and

$$\mathbb{P}[(S^{1/\alpha}X)^\beta \le z] = \mathbb{P}[X \le S^{-1/\alpha}z^{1/\beta}] = e^{-T(S^{-1/\alpha}z^{1/\beta})^{-\alpha}} = e^{-STz^{-\alpha/\beta}}.$$

□

Lemma A.2. *Let $\{X_i\}_{i=1,\dots,N}$ be α-Fréchet with scale parameters $\{T_i\}_{i=1}^N$ and correlation function $G : \mathbb{R}_+^N \to \mathbb{R}_+$. Then, for any $S_i \ge 0$ $i = 1, \dots, N$ and $\beta > 0$, the random vector $\{(S_i^{1/\alpha}X_i)^\beta\}_{i=1}^N$ is α/β-Fréchet with location parameters of $\{S_iT_i\}_{i=1}^N$ and correlation function G.*

Proof.

$$\begin{aligned}
\mathbb{P}[(S_i^{1/\alpha}X_i)^\beta \le y_i, i = 1,\dots,N] &= \mathbb{P}[X_i \le S_i^\alpha y_i^{1/\beta}, i = 1,\dots,N] \\
&= \exp\left[-G(T_1S_1y_1^{-\alpha/\beta},\dots,T_NS_Nx_N^{-\alpha/\beta})\right].
\end{aligned}$$

□

Lemma A.3. *Let $\{X_i\}_{i=1,\dots,N}$ 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 $\max_{i=1,\dots,N} X_i$ is θ-Fréchet with location $G(T_1,\dots,T_N)$. Moreover, let $\{\mathcal{I}_j\}_{j=1}^M$ be any partition of $\{1,\dots,N\}$ and define the random variable $\{Y_1,\dots,Y_M\}$ as*

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

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

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

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

2.
$$\mathbb{P}\left[Y_j = \max_i X_i\right] = \frac{\sum_{i \in \mathcal{I}_j} T_i G_i(T_1, \dots, T_N)}{G(T_1, \dots, T_N)},$$
where $G_i(x_1, \dots, x_N) \equiv \partial G(x_1, \dots, x_N)/\partial x_i$;

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

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

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

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

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

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

and its correlation function must then be

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

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

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

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

Next we prove part (2). We have

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

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:

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

□

B Proof of Lemma 1

Proof. 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 $(x_1, \ldots, x_N) \in \mathbb{R}_+^N$ and let $\lambda > 0$. We have

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

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

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

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

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

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

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

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

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

Therefore, $\lim_{x_o \to \infty} G^d(x_1, \ldots, x_N) = \infty$ 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 $o = 1, \dots, N$, and let $G^d : \mathbb{R}_+^N \to \mathbb{R}_+$ be a correlation function. Suppose that $\{A_{od}(v)\}_{o=1}^N$ satisfies

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

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

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

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

C Proof of Theorem 1

Proof. The proof follows Theorem 2 of De Haan (1984) and Theorem 3 of Penrose (1992). First, we prove sufficiency. Using Assumption 1,

$$\begin{aligned}
\mathbb{P}\left[A_{od}(v) \le a_o \quad \forall o = 1, \dots, N\right] &= \mathbb{P}\left[\max_{i=1,2,\dots} (Z_i(v)A_{iod}(v))^{1/\theta} \le a_o \quad \forall o = 1, \dots, N\right] \\
&= \mathbb{P}\left[\max_{i=1,2,\dots} Z_i(v)A_{iod}(v) \le a_o^{\theta} \quad \forall o = 1, \dots, N\right] \\
&= \mathbb{P}\left[Z_i(v) \le \min_{o=1,\dots,N} \frac{a_o^{\theta}}{A_{iod}(v)} \quad \forall i = 1, 2, \dots\right]
\end{aligned}$$

Using Assumption 2 and 3, the last expression is equal to

$$
\begin{aligned}
&= \exp\left[-\mathbb{E}\int_{\min_o a_o^\theta/A_{iod}(v)}^{\infty} z^{-2}\mathrm{d}z\right] \\
&= \exp\left[-\mathbb{E}\max_{o=1,\dots,N} A_{iod}(v)a_o^{-\theta}\right] \equiv \exp\left[-G^d(T_{1d}a_1^{-\theta},\dots,T_{Nd}a_N^{-\theta})\right]. \qquad (\star)
\end{aligned}
$$

Necessity is proved by showing that for any G^d, $\{T_{od}\}_{o=1}^N$, and θ there exists $A_{iod}(v)$ with $T_{od} = \mathbb{E}A_{iod}(v)$ such that ($\star$) holds for some independent Poisson point process $\{Z_i(v)\}_{i=1,2,\dots}$ on $[0,\infty)$ with intensity measure $z^{-2}dz$. Theorem 2 of De Haan (1984) implies that if $\tilde{A}_{od}(v)$ is multi-variate 1-Fréchet with unit scale parameters and correlation function G^d, then there exists a random variable $\tilde{A}_{iod}(v)$ with $\mathbb{E}\tilde{A}_{iod}(v) < \infty$ and an independent Poisson process $\{\tilde{Z}_i(v)\}_{i=1,2,\dots}$ on $[0,\infty)$ with intensity measure $z^{-2}dz$ such that

$$
\mathbb{P}\left[\max_{i=1,2,\dots} \tilde{Z}_i(v)\tilde{A}_{iod}(v) \leq \tilde{a}_o\right] = \exp\left[G^d(\tilde{a}_1^{-1},\dots,\tilde{a}_N^{-1}\right]
$$

Moreover, we must have $\mathbb{E}\tilde{A}_{iod}(v) = 1$. As a result, ($\star$) holds after a change of variables to $A_{iod}(v) = T_{od}\tilde{A}_{iod}(v)$, and $a_o = \tilde{a}_o^{1/\theta}$.

□

D Proof of Proposition 1

Proof. Perfect competition implies that potential import prices are

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

Then,

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

□

E Proof of Proposition 2

Proof. 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

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

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

F Proof of Proposition 3

Proof. 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

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

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

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

so that we can write the expenditure share as

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

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 using the following correlation function, for some $\theta > \sigma$,

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

This choice of G^d will give the result because it implies a price level,

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

that approximates the true price level. In particular,

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

That is, the price level implied by this correlation function converges pointwise to the price level associated with the true productivity distribution.

The implied GEV factor demand system is

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

That is, the implied GEV factor demand system converges pointwise 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. □

G Proof of Proposition 4

Proof. Micro productivity is distributed multivariate θ-Fréchet, with scale T_{mod} and micro correlation function $F^d(\mathbf{X}_1, \ldots, \mathbf{X}_N)$. Aggregate productivity is $A_{od}(v) \equiv \max_{m=1,\ldots,M_o} A_{mod}(v)$. Using Lemma A.2 and Lemma A.3, $A_{od}(v)$ is distributed as a multivariate θ-Fréchet with scale parameters

$$T_{od} = F^d\left(\mathbf{0}_1, \ldots, \mathbf{0}_{o-1}, \mathbf{T}_{od}, \mathbf{0}_{o+1}, \ldots, \mathbf{0}_N\right) \equiv F^{od}\left(\mathbf{T}_{od}\right),$$

for $\mathbf{T}_{od} \equiv (T_{1od}, \ldots, T_{M_o od})$, and correlation function

$$G^d(x_1, \ldots, x_N) \equiv F^d(\mathbf{\Omega}_{1d} x_1, \ldots, \mathbf{\Omega}_{Nd} x_N),$$

for $\omega_{mod} \equiv T_{mod}/T_{od}$ and $\mathbf{\Omega}_{od} \equiv (\omega_{1od}, \ldots, \omega_{M_o od})$.

□

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).[25] 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.

[25] 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.

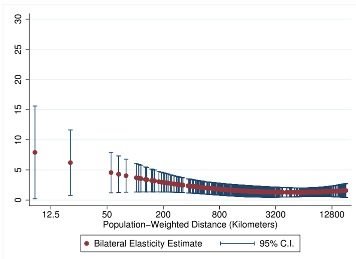
I Additional Tables and Figures

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

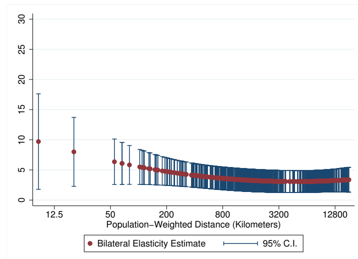
Dep variable	CES $\ln \pi_{sodt}$	Cross-nested CES $\ln \sum_o \pi_{sodt}$	Spatial $\ln \sum_o \pi_{sodt}$
	(1)	(2)	(3)
$\ln(1+t_{sodt})$	5.523 (0.020)***	0.607 (0.025)***	0.489 (0.024)***
$\hat{x}^{cnces}_{sodt}$		✓	
$\hat{x}^{spatial}_{sodt}$			✓
Sector-Covariate Interactions	✓	✓	✓
Sector-Origin-Year 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 (48) by OLS, assuming that $\rho_{sod} = 0$ so that $\sigma_{sod} = \theta$. Results in columns (2) and (3) are from estimating (49) by OLS, with $\hat{x}^l_{sodt} = \ln\widehat{(P_{sod}/P_{sd})}$, $l = spatial, cnces$, from the first-stage gravity equation in (48). In column (2), (47) 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.

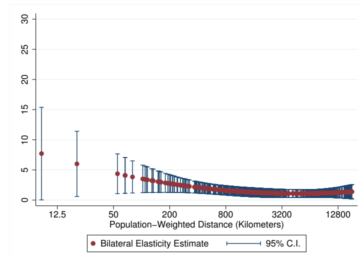
Figure I.1: Elasticities of Substitution and Distance, by sector.



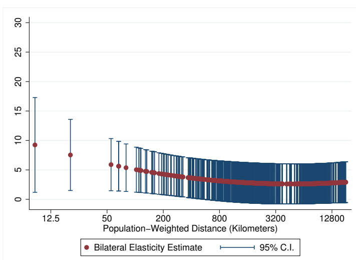
Agriculture, hunting, forestry, and fishing



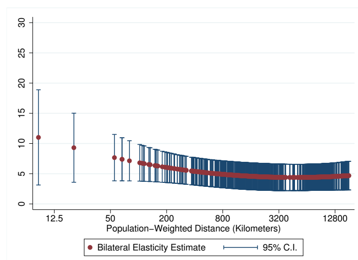
Mining and quarrying



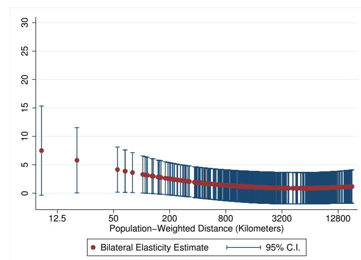
Food, beverages, and tobacco



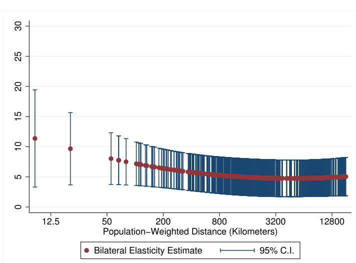
Textiles and leather



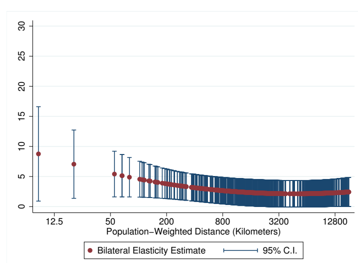
Wood and products of wood and cork



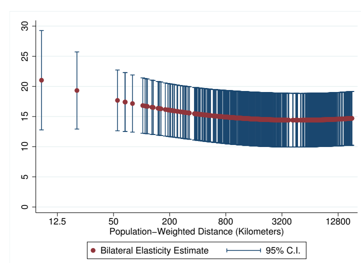
Pulp, paper, publishing and printing



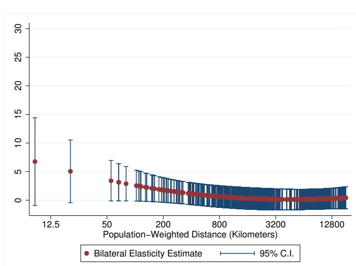
Fuel, chemicals, rubber, and plastics



Other non-metallic mineral products



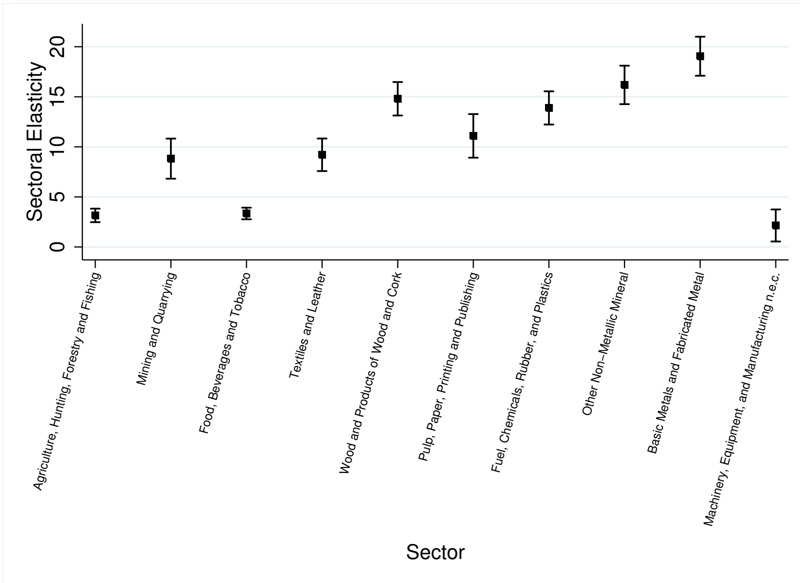
Basic metals and fabricated metal products



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

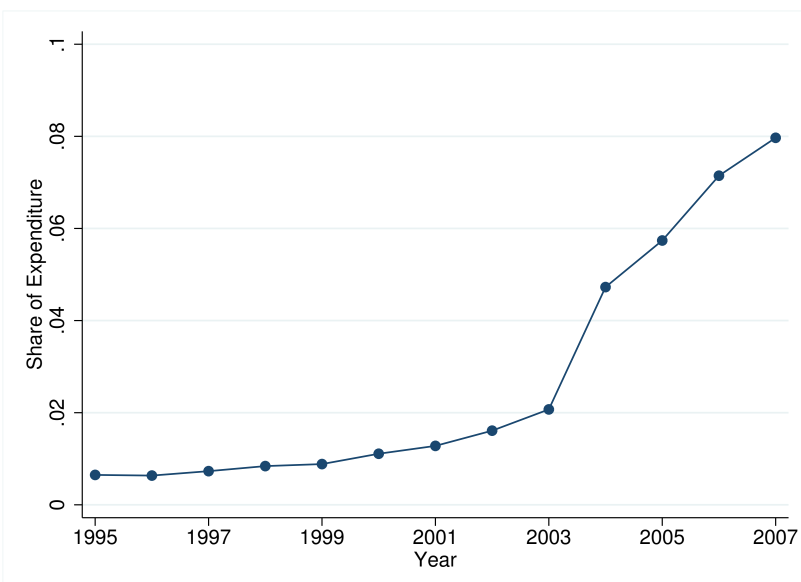
Notes: Results from estimating (48) by OLS.

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



Notes: Results from estimating (48) 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.