Trade with Correlation*

Nelson Lind
Emory University

Natalia Ramondo
Boston University and NBER

March 24, 2021

Abstract

We develop a trade model where correlation in productivity between countries arises from technological similarity. The model spans the full class of generalized extreme value (GEV) import demand systems and formalizes Ricardo's insight that countries with relatively dissimilar technology gain more from trade. Our characterization of productivity links the GEV structure to technological primitives, providing the basis for an estimation procedure for correlation. We estimate significant differences in technological sharing across sectors and countries. These estimates imply larger and more heterogenous gains from trade relative to models that assume independent productivity across sectors.

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

*We thank our discussants Costas Arkolakis and Arnaud Costinot for their very helpful comments. We benefit from comments from Rodrigo Adao, Roy Allen, Dave Donaldson, Jonathan Eaton, Pablo Fajgelbaum, Teresa Fort, Stefania Garetto, and Fernando Parro, as well as from participants at various seminars and conferences. This research is supported by NSF grant #1919372. 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 led to the following insight: Two countries gain more from trade when they have dissimilar production possibilities. If so, different degrees of correlation in productivity lead to heterogeneity in the gains from trade.

While the Ricardian trade model in Eaton and Kortum (2002, henceforth, EK) does not account for correlation in productivity, the rich quantitative literature inspired by EK captures the Ricardian insight by incorporating sectors, multinational production, or global value chains. Yet, the way this literature incorporates correlation is restrictive, potentially removing empirically relevant sources of heterogeneity. Concretely, although multi-sector models create correlation because each sector is present in many countries, productivity is independent across sectors, implying that sectors do not share technologies.

In this paper, we develop a Ricardian theory that allows for rich patterns of correlation in productivity, generates import demand systems spanning the entire generalized extreme value (GEV) class (McFadden, 1978, 1981), and preserves the tractability properties central to the EK model. The novelty of our approach stems from linking the GEV structure to a common foundation based on technological primitives. This foundation sheds light on the properties and limitations of existing models, provides tools to build new models with rich patterns of correlation, and enables us to develop an estimation procedure based on disaggregate data to uncover the technological factors that underly correlation in productivity. In the context of a multi-sector model of trade, our estimates reveal significant technological sharing across sectors and countries, which changes the answers to standard counterfactuals.

The Ricardian trade model in EK assumes that productivity is a random variable drawn from an independent max-stable Fréchet distribution. While the max-stability property leads to tractability, the independence assumption entails that bilateral trade flows follow a constant elasticity of substitution (CES) structure. By preserving the max-stability property central to the EK model, but allowing for a general dependence structure in productivity across countries, our model implies expenditure shares that belong to the GEV class, a class that admits rich substitution patterns. We show that, as a result, any trade model that generates a GEV import demand system is equivalent to a Ricardian model with a max-stable multivariate Fréchet distribution. In this way, our approach brings many existing trade models into a unifying framework, including: multi-sector models (Costinot et al., 2012; Costinot and Rodrìguez-Clare, 2014; Levchenko and Zhang, 2014; Di-Giovanni 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 (Alvarez and Lucas, 2007; Antràs and de Gortari, 2017); and models of trade with domestic geography (Fajgelbaum and Redding, 2014; Ramondo et al., 2016; Redding, 2016).

Despite its generality, our theory leads to intuitive and tractable counterfactual analysis. We can calculate the gains from trade for the GEV class as a simple adjustment to the CES case: The results of Arkolakis et al. (2012) (henceforth, ACR) generalize, after a simple correction, to the GEV class. In the Ricardian context, this correction adjusts a country's self-trade share for correlation in technology with the rest of the world, formalizing Ricardo's insight that more dissimilar countries have higher gains from trade.

We go a step further and characterize the GEV class by providing a structure for technology that is *necessary and sufficient* for productivity to be distributed max-stable multivariate Fréchet. Our approach leverages the spectral representation theorem for max-stable processes (De Haan, 1984; Kabluchko, 2009), which generates max-stable distributions from Poisson processes. We refer to this structure as the *global innovation representation* because productivity is interpreted as the result of countries adopting globally available innovations according to their individual ability to apply them. When countries adopt similar technologies—those with similar characteristics, "latent factors"—they have correlated productivity. This representation enables us to link existing GEV Ricardian models (e.g., many sectors) to technological primitives, as well as build new models with rich patterns of correlation.

This characterization sets the foundation to estimate correlation in productivity. We show that the global innovation representation implies a linear latent-factor structure such that expenditure is approximately cross-nested CES (CNCES). The advantage is that we do not need to impose strong distributional and parametric assumptions, so that we can estimate import demand systems with rich substitution patterns. We show how to estimate this latent-factor structure using an estimator based on non-negative matrix factorization (Lee and Seung, 1999, 2001; Fu et al., 2019) and pseudo Poisson maximum likelihood (Silva and Tenreyro, 2006; Fally, 2015). We apply our procedure to the estimation of a multi-sector model of trade where we fully relax the common assumption of independent sectoral productivity.[1] Our global innovation representation makes clear that this independence assumption corresponds to a structure where sectors do not share technologies and independence of irrelevant alternatives (IIA) holds at the sector level. We find significant departures from IIA for aggregate sectoral categories, suggesting that sectors do share technology.

We estimate that seven latent technological factors explain over 90 percent of the variation in 4-digit SITC bilateral trade flows. These factors are broadly shared across sectors, yet are also used intensively for the production of certain goods. Factors related primarily to the production of simple manufactured goods are highly correlated across countries, while the factors associated with the production of complex manufactured products, such as electronics, and with natural-resource extraction are almost uncorrelated across countries.

The aggregate substitution elasticities generated by these factors differ significantly from those implied by models with independent sectoral productivity, and lead to larger and more heterogenous estimates of the gains from trade. For example, the welfare cost to the United States of a 50-percent tariff on China doubles. This result arises because we estimate that China has a comparative advantage in the technological factor used intensely in the production of complex manufactured goods and this factor is weakly correlated across countries. As a consequence, US consumers cannot easily find an alternative supplier as tariffs increase.

Our paper relates to several strands of the literature. First, we naturally relate to the large trade literature using the Ricardian-EK framework (Eaton and Kortum, 2012). More generally, our approach can be applied to any environment that requires Fréchet tools, such as selection models used in the macro development literature (Lagakos and Waugh, 2013; Hsieh et al., 2013; Bryan and Morten, 2018),

[1] While we relax the independence assumption of multi-sector EK models, we do not incorporate an input-output structure into the model, as in e.g. Caliendo and Parro (2015). We show in Online Appendix O.1 that a trade model with an input-output loop belongs to the GEV class. For a generalization of the input-output network to non-linear environments and its consequences for the gains from trade see Baqaee and Farhi (2019).

as well as trade models used in the urban literature (Ahlfeldt et al., 2015; Monte et al., 2015; Caliendo et al., 2017; Redding and Rossi-Hansberg, 2017).

Our paper is closely related to Adao et al. (2017), who show how to calculate macro counterfactual exercises in neoclassical trade models with invertible import demand systems. They provide sufficient conditions for non-parametric identification using aggregate trade data. Their approach departs from CES demand, but does not necessarily lead to closed-form results. By focusing on the subclass of GEV import demand systems, we operationalize a tractable model of Ricardian comparative advantage where IIA does not need to hold, and relate various disaggregate structures to the macro demand systems studied by Adao et al. (2017). All in all, our distinct contribution is to provide a foundation for the entire GEV class of import demand systems. This foundation allows us to develop an estimation procedure based on latent factors and disaggregate data, and presents an alternative to the Berry et al. (1995) procedure used in Adao et al. (2017).[2]

Relatedly, 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, abandon homothetic demand systems, which we do not, and aim, as we do, to incorporate detailed micro data to estimate elasticities. In contrast with this literature, in our supply-side framework, substitution patterns come from the degree of technological similarity between countries. As a result, we can incorporate heterogeneous elasticities without relying on demand-side factors, which links seemingly different micro structures to common primitives of technology, and ties the micro estimates in this literature to macro counterfactual exercises.

Finally, our global representation result relates to the literature on dynamic innovation and knowledge diffusion processes that generate Fréchet productivity (Kortum, 1997; Eaton and Kortum, 1999, 2001; Buera and Oberfield, 2016). This literature uses extreme-value theory to generate independent max-stable random variables. We introduce a new tool—the spectral representation theorem for max-stable processes—and use it for estimation. In contrast with methods from extreme-value theory, our approach accommodates statistical dependence, and delivers closed-form and exact, not limiting, results.

[2] Our model also generates mixed CES as used in Adao et al. (2017) (see Online Appendix O.1.6).

2 The Ricardian Model of Trade from Primitives

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 $\eta > 0$, $C_d = \left(\int_0^1 C_d(v)^{\frac{\eta-1}{\eta}} \mathrm{d}v\right)^{\frac{\eta}{\eta-1}}$. Expenditure on variety v is $X_d(v) \equiv P_d(v)C_d(v) = (P_d(v)/P_d)^{1-\eta} X_d$ where $P_d(v)$ is the price of the variety, $P_d = \left(\int_0^1 P_d(v)^{1-\eta} \mathrm{d}v\right)^{\frac{1}{1-\eta}}$ is the price level in country d, and X_d is total expenditure in country d.

Each variety, v, is produced with an only-labor constant returns to scale technology. The marginal product of labor, $Z_{od}(v)$, referred to as *productivity*, depends on both the origin country o where the good gets produced and the destination market d where it gets delivered. This variable captures both efficiency of production in the origin and inefficiencies associated with delivery to the destination—trade costs.

Productivity is the result of adopting the best innovation available in country o to serve market d, for variety v. We assume that for each $v \in [0,1]$, there exists an infinite, but countable, set of technological innovations, $i = 1, 2, \ldots$, that represent physical techniques (i.e., blueprints) for producing a good. Each innovation is characterized by two components.

Quality, $Q_i(v)$, measures the fundamental efficiency of the technique, and is identical across all origins and destinations.

Characteristics, $\chi_i(v)$, represent anything specific to the innovation i—for example, which sectors and countries can use the innovation. Combined with bilateral factors (e.g. proximity), characteristics generate heterogeneity in productivity across origins and destinations. This spatial heterogeneity is captured by the function $A_{od}(\chi_i(v))$, which we refer to as *spatial applicability*. For instance, if a characteristic is whether an innovation is known in each country, applicability is zero in countries with no knowledge of the innovation; if the country knows the innovation, applicability is positive and can depend on the proximity between production location o and destination d. In general, these characteristics are not observable and we refer to them as *latent factors*. We formalize our assumption on the determinants of productivity next.

Assumption 1 (Innovation Decomposition). *There exists a measurable space of characteristics* $(\mathcal{X}, \mathbb{X})$ *and for each* $v \in [0,1]$ *an infinite, but countable, set of global innova-*

tions, $i = 1, 2, \ldots$, with quality, $Q_i(v) > 0$, and characteristics, $\chi_i(v) \in \mathcal{X}$, such that

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

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

This assumption specifies that quality and spatial applicability combine multiplicatively to determine patterns of technology adoption. Countries adopt those innovations with the highest applicability-adjusted quality.

Given the productivities arising from technology adoption, the marginal cost to deliver a variety v to destination d from origin o is $W_o/Z_{od}(v)$ where W_o is the nominal wage in country o.[3] Under perfect competition, prices are equal to unit costs and good v is provided to country d by the lowest-cost supplier,

$$P_d(v) = \min_{o=1,\ldots,N} W_o/Z_{od}(v). \tag{2}$$

Next, as in EK, we model productivity as a random draw. We focus on multivariate random variables that satisfy a property known as max stability; that is, we focus on multivariate max-stable Fréchet distributions. The EK model, which is built on independent Fréchet random variables, gets its tractability from this property. We first adapt the tools developed originally for random utility models to Ricardian models of trade. In this way, we are able to relax the independence assumption in EK, and get a flexible, yet tractable, model of trade with an import demand system in the generalized extreme value (GEV) class.

Our contribution with respect to McFadden (1978, 1981) and Eaton and Kortum (2002) is not only to fully characterize the GEV class in the Ricardian context, but more importantly to provide technological primitives—quite natural in the Ricardian context—for the origins of the entire GEV class. Our result in Theorem 1 provides economic interpretation for—and facilitates the estimation of—*any* GEV model.

[3] Notice that our framework does not require the standard assumption on iceberg trade costs (Samuelson, 1954). That assumption entails that $Z_{od}(v) = Z_o(v)\tau_{od}$ and that, for each origin o, the productivity to serve any destinations d with variety v is the same. Relaxing this assumption allows us to have a productivity distribution over *both* origins o and destinations d.

2.1 Max-Stable Multivariate Fréchet Productivity

We want the vector of productivities across countries to be drawn from a max-stable multivariate Fréchet distribution. Max stability, the central property in EK, ensures closed-form solutions for equilibrium prices and expenditure shares.

The vector of productivities $(Z_{1d}(v), \ldots, Z_{Nd}(v))$ is max-stable multivariate Fréchet if for any $\alpha_o \geq 0$ with $o = 1, \ldots, N$ the random variable $\max_{o=1,\ldots,N} \alpha_n Z_{od}(v)$ has a Fréchet distribution with shape parameter θ. In this case, the marginal distributions, $\mathbb{P}\left[Z_{od}(v) \leq z_o\right] = \exp\left[-T_{od} z_o^{-\theta}\right]$, are Fréchet with a (common) shape parameter θ and scale parameter T_{od}. This definition directly implies the max-stability property: The maximum of the components of the vector has the same distribution (to scale) as the marginal distributions. Because a common shape parameter is necessary, we refer to a multivariate max-stable Fréchet distribution as a *multivariate θ-Fréchet* distribution.[4]

These multivariate distributions are parameterized by the *scale parameters* of the marginal distributions, T_{od}, and by a *correlation function*, G^d, also called tail dependence function in probability and statistics (see Appendix Lemma A.2),[5]

$$\mathbb{P}\left[Z_{1d}(v) \leq z_1, \cdots Z_{Nd}(v) \leq z_N\right] = \exp\left[-G^d(T_{1d} z_1^{-\theta}, \ldots, T_{Nd} z_N^{-\theta})\right].$$

As in EK, the scale parameters capture the absolute advantage of countries, while the shape parameter regulates the heterogeneity of productivity draws, which are independent and identically distributed across the continuum of varieties—as in all models based on EK.

Appendix Lemma A.3 presents properties of the correlation function. The key property ensuring max-stability is homogeneity, which implies that the scale of the maximum is $G^d(T_{1d}, \ldots, T_{Nd})$. Additionally, a correlation function also presents the regularity properties of the social surplus function in GEV discrete choice models (McFadden, 1981; Train, 2009): unboundedness ($G^d(x_1, \ldots, x_N) \to \infty$ as $x_o \to \infty$ for any $o = 1, \ldots, N$); and (if sufficiently differentiable) a sign pattern for the cross-partial derivatives. Finally, a correlation function must satisfy a normalization

[4]Multivariate Fréchet distributions may have marginal distributions with different shape parameters, in which case the maximum, even under independence, is not distributed Fréchet.

[5]A function $G : \mathbb{R}_+^N \to \mathbb{R}_+$ is a *correlation function* if $C(u_1, \ldots, u_N) \equiv \exp[-G(-\ln u_1, \ldots, u_N)]$ is a max-stable copula—that is, $C(u_1, \ldots, u_N) = C(u_1^{1/m}, \ldots, u_N^{1/m})^m$ for any $m > 0$ and all $(u_1, \ldots, u_N) \in [0,1]^N$. For details see Gudendorf and Segers (2010).

restriction, $G(0, \ldots, 0, 1, 0, \ldots, 0) = 1$. This property separates the scales, which parameterize the marginal distributions, from the correlation function, which determines the joint distribution of productivity.

In EK, productivity draws are independent across origin countries,

$$\mathbb{P}[Z_{1d}(v) \le z_1, \ldots, Z_{Nd}(v) \le z_N] = \prod_{o=1,\ldots,N} \mathbb{P}[Z_{od}(v) \le z_o] = \exp\left(-\sum_{o=1}^{N} T_{od} z_o^{-\theta}\right). \tag{3}$$

This assumption implies an additive correlation function,

$$G^d(x_1, \ldots, x_N) = \sum_{o=1}^{N} x_o. \tag{4}$$

In this case, the scale of the maximum is simply the sum of marginal distribution scales, $\sum_{o=1}^{N} T_{od}$. Under independence, the shape parameter also captures comparative advantage across countries. Our approach drops the independence assumption and introduces the correlation function, which controls comparative advantage by determining relative productivity levels across origin countries within the same destination market.

The correlation function allows for a flexible dependence structure across countries. But, as Train (2009, page 107) warns: "There is little economic intuition to motivate [the] properties [of G] ... However, it is easy to verify whether a function exhibits these properties. [...] The disadvantage is that the researcher has little guidance on how to specify a G that provides a model that meets the needs of his research. The advantage is that the purely mathematical approach allows the researcher to generate models that he might not have developed while relying only on his economic intuition." In the next section, we attempt to fill that gap. We establish a structure for technology that is *necessary and sufficient* for productivity to be multivariate θ-Fréchet. This structure not only provides an economic justification for any choice of the correlation function, but it also facilitates its estimation.

2.2 The Origins of Max-Stable Multivariate Fréchet Distributions

Next, we establish the conditions under which productivity—the result of adopting the best available innovation, as specified in (1)—has a multivariate θ-Fréchet distribution. We first specify the assumption on the stochastic properties of quality and applicability.

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

Assumption 2 states that innovations follow a Poisson process over qualities and characteristics. The key is that the measure μ, which determines the joint distribution of spatial applicability across origin countries, is relatively unrestricted and allows for rich patterns of correlation—through the inclusion of characteristics.[6]

This assumption has two main implications. First, the measure μ specifies the expected number of innovations with characteristics in any set $B \in \mathbb{X}$ and quality above 1. For example, if $\chi_i(v)$ is a list of sectors that can use the innovation, $\mu(\{s_1, s_2\})$ is the expected number of innovations that can be used by both sectors s_1 and s_2. Second, conditional on being above any cutoff $\underline{q}$, quality is independent of characteristics and distributed Pareto with lower bound $\underline{q}$ and shape θ.[7] As a consequence, spatial applicabilities are independent of global productivities.

The following theorem characterizes when productivity is multivariate θ-Fréchet, and is a consequence of the spectral representation theorem for max-stable processes (De Haan, 1984; Kabluchko, 2009).[8]

[6] Assumption 2 can be interpreted as arising from a random discovery process as in Eaton and Kortum (1999, 2010). In our static framework, we interpret i as indexing all innovations up until the present. The difference between our setup and the Poisson process in Eaton and Kortum (1999) is the inclusion of innovation characteristics. Rather than assuming that innovations are country specific, innovations are global and countries adopt the most efficient innovation for them depending on how characteristics determine the spatial applicability.

[7] The expected number of innovations with $Q_i(v) > \underline{q} > 0$ and $\chi_i(v) \in B \in \mathbb{X}$ is

$$\mathbb{E} \sum_{i=1}^{\infty} \mathbf{1}\{Q_i(v) > \underline{q}, \chi_i(v) \in B\} = \int_B \int_{\underline{q}}^{\infty} \theta q^{-\theta-1} \mathrm{d}q \mathrm{d}\mu(\chi) = \underline{q}^{-\theta} \mu(B).$$

Conditional on $Q_i(v) > \underline{q}$ and $\chi_i(v) \in B$, the likelihood that $Q_i(v) = q$ is

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

[8] Resnick and Roy (1991) characterize upper-semi-continuous max-stable random utility processes on complete metric spaces. Dagsvik (1994) uses the spectral representation theorem to establish behavioral assumptions that imply max-stable random utility processes on countable sets and subsets of Euclidean space. Our global innovation representation establishes necessary and sufficient conditions for max-stable productivity on a finite set (of countries) and puts restrictions on cardinal—and potentially observable—outcomes, rather than on ordinal outcomes such as latent utility. Also, due to Kabluchko (2009), Theorem 1 extends directly to the case with an arbitrary index set.

Theorem 1 (Global Innovation Representation). *We say that productivity has a global innovation representation if it satisfies Assumptions 1 and 2. In that case, for each destination d, productivity is distributed multivariate θ-Fréchet,*

$$\mathbb{P}\left[Z_{1d}(v) \leq z_1, \ldots, Z_{Nd}(v) \leq z_N\right] = \exp\left[-G^d\left(T_{1d} z_1^{-\theta}, \ldots, T_{Nd} z_N^{-\theta}\right)\right], \tag{5}$$

with scale

$$T_{od} \equiv \int_{\mathcal{X}} A_{od}(\chi)^{\theta} d\mu(\chi), \text{ for } o = 1, \ldots, N, \tag{6}$$

and correlation function

$$G^d(x_1, \ldots, x_N) \equiv \int_{\mathcal{X}} \max_{o=1,\ldots,N} \frac{A_{od}(\chi)^{\theta}}{T_{od}} x_o d\mu(\chi). \tag{7}$$

Conversely, if productivity is distributed multivariate θ-Fréchet, it admits a global innovation representation.

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

Theorem 1 establishes that θ-Fréchet productivity can always be interpreted as arising from the spatial applicability of global technologies. In turn, a global innovation representation leads to θ-Fréchet productivity. Correlation in productivity reflects the extent to which countries adopt similar innovations—with characteristics that lead to similar applicability across countries—while scale parameters reflect the average applicability of the innovations adopted by each country. The key aspect of Theorem 1 is that the correlation function has an integral representation over the characteristics of innovations. As a direct consequence, we can approximate the correlation function using the cross-nested CES (CNCES) functional form in the following corollary.

Corollary 1. *Suppose Assumptions 1 and 2 hold and, additionally, there exists a partition of characteristics, $\{\mathcal{X}_k\}_{k=1}^K$, such that, for each k, applicability restricted to $\mathcal{X}_k$ is independent σ_k-Fréchet across origins. That is, the measure μ satisfies*

$$\int_{\mathcal{X}_k} \mathbf{1}\{A_{od}(\chi) \leq a_o \;\; \forall o\} \frac{d\mu(\chi)}{\mu(\mathcal{X}_k)} = \exp\left[-\sum_{o=1}^{N} A_{kod}^{\sigma_k} a_o^{-\sigma_k}\right], \tag{8}$$

for some $\{A_{kod}\}_{o=1}^{N}$, *for each* $k = 1, \ldots, K$.

Then, the distribution of productivity is multivariate θ*-Fréchet with scale*

$$T_{od} = \sum_{k=1}^{K} \Gamma(\rho_k) A_{kod}^{\theta} \mu(\mathcal{X}_k)$$

and a cross-nested CES (CNCES) correlation function,

$$G^d(x_1, \ldots, x_N) = \sum_{k=1}^{K} \left[\sum_{o=1,\ldots,N} (\omega_{kod} x_o)^{\frac{1}{1-\rho_k}} \right]^{1-\rho_k}, \tag{9}$$

where $\omega_{kod} \equiv \Gamma(\rho_k) A_{kod}^{\theta} \mu(\mathcal{X}_k) / \sum_{k'=1}^{K} \Gamma(\rho_{k'}) A_{k'od}^{\theta} \mu(\mathcal{X}_{k'})$ *and* $\rho_k \equiv 1 - \theta/\sigma_k$.

Proof. See Appendix B.2. □

This corollary restricts the measure μ to derive a useful and interpretable form for the correlation function. Each nest of this CNCES correlation function corresponds to a group of characteristics with similar applicability across origins and destinations, which we refer to as a *factor*. The parameter σ_k controls dispersion in applicability within factor k. The Poisson i.i.d. structure in Assumption 2 implies independence across factors. Within factor k, however, dispersion in spatial applicability determines correlation in productivity across origins, measured by the correlation coefficient, ρ_k. As $\sigma_k \to \theta$, dispersion in applicability is high, and $\rho_k \to 0$. Intuitively, when applicability becomes very fat tailed, it dominates the contribution of the common quality component of productivity. In this case, productivity is independent and the k'th nest of the correlation function is additive due to the assumption that applicability is independent across countries. In contrast, as $\sigma_k \to \infty$, dispersion in applicability becomes negligible and $\rho_k \to 1$. In this case, applicability becomes deterministic and heterogeneity in productivity is entirely determined by quality. Since quality is common across countries, productivity becomes perfectly correlated within factor k.

The weights, ω_{kod}, sum to one across factors and indicate the relative importance of each group of characteristics for a given trading pair. If the share of factor k is high for a given country pair, it means that those innovations are particularly productive for production in country o and delivery to d.

The corollary is important for two reasons. First, it establishes that, under some assumptions for spatial applicability, the correlation function has a CNCES form. This case is the building block for multivariate θ-Fréchet distributions commonly used in the literature (e.g. multi-sector EK models). In the examples in Section 2.3, we show how specific EK-type trade models arise as special cases from these primitive assumptions on technology.

Second, as $K \to \infty$, the CNCES correlation function can uniformly approximate any correlation function—the function in (7). The next proposition provides a bridge between the result in Theorem 1 and Corollary 1, setting the base for our CNCES estimation procedure in Section 4.

Proposition 1 (Cross-Nested CES Approximation)**.** *Any correlation function can be uniformly approximated on compact sets by a CNCES correlation function as in* (9)*.*

Proof. The proof constructs partitions of the space of characteristics (i.e. the groups indexed by k in Corollary 1), and for each partition an approximating CNCES correlation function, as in (9). As we let the partitions become increasingly fine, the approximating correlation functions uniformly converge on compact sets to the true correlation function. See Appendix B.3. □

The result in Proposition 1 implies that not only can we uniformly approximate any correlation function with some CNCES correlation function, but also that we can always interpret the nests of that function as corresponding to underlying distinct groups of innovations, representing latent technological factors. It means that the interpretation of Theorem 1 directly transfers to CNCES correlation functions, which are useful for both closed-form theoretical results as well as empirical analysis.[9]

Specifically, this result allows us to estimate the latent factors that underly the distribution of productivity. Section 4 presents our proposed estimator in the context of a multi-sector Ricardian model of trade. Guided by Theorem 1, we uncover from observable sectoral expenditure shares the latent dimensions that indicate that some innovations are more useful in certain sectors, but different sectors can potentially use the same innovations. As we explain next—and in detail in Section 4—the standard multi-sector EK model imposes independent sectoral productivity, which implicitly assumes that technologies are sector specific; that is,

[9]This result is related to the result in Fosgerau et al. (2013) where the choice probabilities of GEV models can be uniformly approximated on compact sets by the choice probabilities of CNCES models. Here, we prove that the underlying distribution can be uniformly approximated.

each sector uses a distinct group of innovations. In this way, those models pair the unobserved factors with observed sectors.

2.3 Examples

We now provide examples of multivariate θ-Fréchet distributions commonly used in the literature, and link them to Theorem 1. All the cases we present are special cases of the CNCES correlation function in Corollary 1. As such, they are derived imposing some additional restrictions on the distribution of spatial applicability.

First, consider the case where each innovation can only be applied in a single country. This restriction corresponds to the model in Eaton and Kortum (2001) where a country's stock of knowledge only depends on their own history of research. It means that each factor is paired with a country ($K = N$), and imposes that A_{kod} is zero for $k \neq o$. The correlation function in (9) collapses to (4), corresponding to independent productivity across countries. This example shows that sharing innovations across countries is necessary for correlation in productivity.

Second, consider an extreme case where all innovations are of the same type, $K = 1$. This restriction means that the distribution of applicability is identical across countries (up to scale) and the relative productivity of any two innovations is the same, on average, across origins. In this case, (9) collapses to

$$G^d(x_1, \ldots, x_N) = \left[\sum_{o=1,\ldots,N} x_o^{\frac{1}{1-\rho}} \right]^{1-\rho}. \tag{10}$$

Corollary 1 directly links the parameter ρ to dispersion in applicability, controlled by σ. High dispersion (low σ), dampens the importance of the global component (quality) and reduces correlation in productivity across countries. Note that for $\rho = 0$, we get the productivity distribution in (3), as in EK. But even for $\rho > 0$, correlation has no impact on trade patterns.[10] This example shows that for correlation to be empirically relevant, countries must be heterogeneous in their ability to use innovations—so we need $K > 1$.

Now consider the case where each factor $k = 1, \ldots, K$ corresponds to a sector $s = 1, \ldots, S$. Corollary 1 implies a CNCES correlation function for aggregate pro-

[10] Eaton and Kortum (2002) point out that $\rho > 0$ implies CES expenditure shares and is equivalent to the case with $\rho = 0$.

ductivity,

$$G^d(x_1,\dots,x_N) = \sum_{s=1}^{S}\left(\sum_{o=1}^{N}(\omega_{sod}x_o)^{\frac{1}{1-\rho_s}}\right)^{1-\rho_s}. \quad (11)$$

This correlation function is commonly used in multi-sector EK-type models (e.g. Costinot and Rodrìguez-Clare, 2014; Caliendo and Parro, 2015). The implicit assumption is that innovations can only be applied in a single sector. Given that the primitive assumption on characteristics is independence (i.e. the Poisson structure in Assumption 2), the lack of common innovations implies independence across sectors. Within each sector, however, productivity draws can be correlated ($\rho_s \geq 0$), and heterogeneous ($\rho_s \neq \rho_{s'}$), with higher sectoral correlation due to more similar applications of innovations across countries (high σ_s).

Our estimation procedure moves away from the assumption of independence across sectors by allowing for technological factors to be shared across sectors, and the number of latent factors, K, to differ from the number of sectors, S. When sectors share factors, within-factor correlation induces across-sector correlation, breaking the independence assumption implicit in pairing factors with sectors.

Finally, it is worth mentioning that the correlation function in (11) is commonly used in applications of the EK model other than the sectoral model. For instance, in Ramondo and Rodríguez-Clare (2013), this specification is used in the context of multinational firms whose home country may differ from their production location (see Online Appendix O.1.2). In that context, factors are paired with home countries and productivity is independent across home countries. Another application is the mixed CES specification used in Adao et al. (2017). A mixed-CES correlation function can be generated by letting $K \to \infty$ and drawing each $\{\omega_{kod}\}_{o=1}^{N}$ and ρ_k from some distribution (see Online Appendix O.1.6).

3 Trade Flows, Prices, and Welfare

Our theory generates macro substitution patterns belonging to the generalized extreme value (GEV) class. This is a large sub-class in the class of invertible demand systems with the gross substitute property, allows for rich patterns of substitution, and leads to closed-form expenditure shares.

In what follows: We derive the import demand system for the Ricardian model presented in Section 2 and establish that it has a one-to-one mapping with expenditure shares in the GEV class; we present properties of the GEV class; and we characterize macro counterfactuals under GEV—a direct generalization of the ACR formula.

3.1 GEV Import Demand

We first present the import demand system implied by θ-Fréchet productivity.

Proposition 2 (Price Levels and Trade Shares). *If productivity is multivariate θ-Fréchet with $\theta > \eta - 1$ and a continuously differentiable correlation function, then:*

1. *The price index in country d is*

$$P_d = G^d \left(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}\right)^{-\frac{1}{\theta}}, \tag{12}$$

for $P_{od} \equiv \gamma T_{od}^{-1/\theta} W_o$ where $\gamma \equiv \Gamma\left(\frac{\theta+1-\eta}{\theta}\right)^{\frac{1}{1-\eta}}$ and $\Gamma(\cdot)$ is the gamma function; and

2. *Country d's expenditure share on country o equals the share of varieties imported,*

$$\pi_{od} \equiv \frac{X_{od}}{X_d} = \frac{P_{od}^{-\theta} G_o^d \left(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}\right)}{G^d \left(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}\right)}, \tag{13}$$

where $G_o^d \left(x_1, \ldots, x_N\right) \equiv \partial G^d \left(x_1, \ldots, x_N\right) / \partial x_o$.

Proof. See Appendix B.4. □

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

Second, the share of expenditure of country d on goods from o equals the probability that o is the lowest cost producer, as in EK, thanks to max-stability.[11]

Finally, the share of varieties imported from o into d has the same form as choice probabilities in GEV discrete choice models, with $P_{od}^{-\theta} = T_{od} W_o^{-\theta}$ replacing choice-

[11] Due to max-stability (see Appendix Lemma A.5), the conditional and unconditional distributions of the maximum are identical: $\mathbb{P}\left[\max_{o'=1,\ldots,N} Z_{o'd}(v) \leq z_{o'} \mid Z_{od}(v) = \max_{o'=1,\ldots,N} Z_{o'd}(v)\right] = \mathbb{P}\left[\max_{o'=1,\ldots,N} Z_{o'd}(v) \leq z_o\right]$. Then $\pi_{od} = \mathbb{E}\left[(P_d(v)/P_d)^{1-\eta} \mathbf{1}\{W_o/Z_{od}(v) = P_d(v)\}\right] = \mathbb{E}\left[(P_d(v)/P_d)^{1-\eta}\right] \mathbb{P}\left[W_o/Z_{od}(v) = P_d(v)\right]$. Since the first term integrates to one, $\pi_{od} = \mathbb{P}\left[W_o/Z_{od} = P_d(v)\right]$. This result does not rely on CES preferences (see Online Appendix O.4).

specific utility.[12] Accordingly, we refer to the import demand system in (13) as a *GEV import demand system*. These import demand systems are uniquely characterized by the shape parameter, θ, and the correlation function, G^d.

An important class of import demand systems within the GEV class is CES. This class is generated by an additive correlation function, so that

$$\pi_{od} = \frac{P_{od}^{-\theta}}{\sum_{o'} P_{o'd}^{-\theta}}. \tag{14}$$

This specification includes most of the workhorse models of trade, such as Armington, Melitz, and EK (Arkolakis et al., 2012). The GEV class, however, is much larger than the CES class, including nested CES, cross-nested CES, and mixed CES, among many others. All these cases can generate patterns of substitution that are richer than CES, the difference coming from the correlation function.

To clearly see this result, compute the elasticity of (13) with respect to the real import price for goods from o' sold in d,

$$\frac{\partial \ln \pi_{od}}{\partial \ln P_{o'd}/P_d} = -\theta \times \mathbf{1}\{o = o'\} - \theta \frac{P_{o'd}^{-\theta} G_{oo'}^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta})}{G_o^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta})}, \tag{15}$$

where $G_{oo'}^d(x_1, \ldots, x_N) \equiv \partial G_o^d(x_1, \ldots, x_N)/\partial x_{o'}$.[13] When the correlation function is additive, the second term is zero and the own-price elasticity is $-\theta$, while the cross-price elasticity is zero. That is, CES entails independence of irrelevant alternatives (IIA). When the correlation function is not additive, the second term in (15) is not zero, generating departures from IIA. The cross-price elasticities are non-negative—GEV import demand entails gross substitutes—because the second cross-partial derivative of the correlation function is non-positive due to the sign-switching property. The own-price elasticity is non-positive because the sum of (15) across o' equals $-\theta$.[14] Since linearity is associated with independence, more

[12]Even though we have not imposed an iceberg structure for trade costs, the scale parameters T_{od} can be mapped into standard variables in the trade literature: an origin-country *productivity index*, $Z_o \equiv T_{oo}^{1/\theta}$, which measures a country's ability to produce goods in their domestic market; and an *iceberg trade cost index*, $\tau_{od} \equiv (T_{oo}/T_{od})^{1/\theta}$, which measures efficiency losses associated with delivering goods to market d. In this way, $T_{od} = (\tau_{od}/Z_o)^{-\theta}$.

[13]Since $G^d(x_1, \ldots, x_N)$ is homogenous of degree one, $G_o^d(x_1, \ldots, x_N)$ is homogenous of degree zero so that we can normalize import prices by the price level before differentiating, and $G_{oo'}^d(x_1, \ldots, x_N)$ is homogenous of degree -1 which allows us to eliminate the price level from the resulting expression.

[14]Since $G_o^d(x_1, \ldots, x_N)$ is homogenous of degree zero, $\sum_{o'=1}^N x_{o'} G_{oo'}^d(x_1, \ldots, x_N) = 0$ so $G_{oo}^d(x_1, \ldots, x_N) = -\sum_{o' \neq o} x_{o'} G_{oo'}^d(x_1, \ldots, x_N) \geq 0$.

curvature is associated with more correlation and stronger departures from IIA.

How is the import demand system in Proposition 2 linked to the technological primitives in Theorem 1? We focus on the cross-nested CES (CNCES) case of Corollary 1, which constitutes the basis of our estimation procedure. In that case, expenditure shares can be decomposed into the sum of expenditure shares on the underlying technological factors,

$$\pi_{od} = \sum_{k=1}^{K} \frac{X^*_{kod}}{X_d}, \quad \text{with} \quad \frac{X^*_{kod}}{X_d} \equiv \frac{\left(\omega_{kod} P_{od}^{-\theta}\right)^{\frac{1}{1-\rho_k}}}{\sum_{o'=1}^{N} \left(\omega_{ko'd} P_{o'd}^{-\theta}\right)^{\frac{1}{1-\rho_k}}} \frac{\left[\sum_{o'=1}^{N} \left(\omega_{ko'd} P_{o'd}^{-\theta}\right)^{\frac{1}{1-\rho_k}}\right]^{1-\rho_k}}{\sum_{k'=1}^{K} \left[\sum_{o'=1}^{N} \left(\omega_{k'o'd} P_{o'd}^{-\theta}\right)^{\frac{1}{1-\rho_{k'}}}\right]^{1-\rho_{k'}}}. \tag{16}$$

The first term in X^*_{kod}/X_d is the share of expenditure on goods made using latent factor k that destination d sources from origin o, while the last term equals the overall share of destination d's expenditure in goods produced using k.

In the special case where each group of characteristics k is paired with a sector, as in (11), we simply replace k by s in (16). Factor-level expenditure equals sectoral expenditure, $X^*_{kod} = X_{sod}$, and within-sector correlation coefficients (and elasticities) are constant across origins and destinations, $\rho_k = \rho_s$.

In the CNCES case, the aggregate elasticity is a weighted average of factor-level elasticities (see Appendix C.2 for derivations),

$$\varepsilon_{oo'd} \equiv \frac{\partial \ln \pi_{od}}{\partial \ln P_{o'd}/P_d} = \sum_{k=1}^{K} \frac{X^*_{kod}}{X_{od}} \varepsilon^*_{koo'd}, \quad \text{with} \quad \varepsilon^*_{koo'd} \equiv \frac{\theta}{1-\rho_k} \left[\rho_k \frac{X^*_{ko'd}}{\sum_{n=1}^{N} X^*_{knd}} - \mathbf{1}\{o = o'\}\right]. \tag{17}$$

Factor-level elasticities, $\varepsilon^*_{koo'd}$, depend on the degree of correlation across origins within each factor, ρ_k, and on the factor expenditure share of o'. The factor-level elasticity is non-positive for $o = o'$ and non-negative for $o \neq o'$. Higher correlation (higher ρ_k) increases within-factor substitution because more similarity in applicability among innovations of type k increases head-to-head competition. If a country tends to export goods using innovations that have very similar applicability across competitors, its aggregate trade flows will be more sensitive to changes in import costs. In contrast, aggregate elasticities will be low for those countries whose exports are specialized in goods that use technologically distinct factors—made with innovations that have very different applicability across countries. Additionally, the aggregate elasticity depends on the composition of expenditure in market d across factors embedded in goods from both o and o'. When o' has a larger share of the market for goods made with high correlation factors, expenditure shifts more rapidly from o' to o when $P_{o'd}$ increases.

Ultimately, elasticities are the result of each country's access to global innovations. When a country has access to innovations that have dissimilar applicability across countries, their productivity will be relatively uncorrelated with other countries, and their goods will not be very substitutable. In contrast, when a country only has access to innovations whose applications are very similar across countries, their productivity will be highly correlated with the rest of the world, and their goods will be easy to substitute.

3.2 GEV Equivalence

Proposition 2 establishes that the Ricardian model with θ-Fréchet productivity maps into the class of GEV import demand systems. However, any trade model that generates trade shares in the GEV class can be cast in terms of a Ricardian model with multivariate θ-Fréchet productivity, and, consequently, has a global innovation representation as established by Theorem 1.

Corollary 2 (GEV Equivalence). *For any trade model that generates a GEV import demand system, there exists a Ricardian model with a global innovation representation that generates the same import demand system.*

By pairing any model with expenditure in the GEV class to a multivariate θ-Fréchet Ricardian model, Corollary 2 ties together many existing non-CES and disaggregate trade models. For example, Ricardian trade models based on θ-Fréchet productivity at the disaggregate level, such as models with many sectors, many regions, multinational firms, or global value chains, have aggregate import demand systems belonging to the GEV class.[15] That is, all these models are equivalent to a model where aggregate productivity is multivariate θ-Fréchet —- a result rooted in the max-stability property due to the fact that aggregate productivity arises from maximizing over disaggregate θ-Fréchet productivity (e.g. sectoral productivity). More generally, the disaggregate outcomes of these models can be linked to the technological primitives in Theorem 1 and can be estimated using the procedure described in Section 4.

[15] Online Appendix O.1 shows the equivalence for the standard multi-sector EK model of trade (without an input-output loop) and other disaggregate EK trade models.

Corollary 2 provides an "umbrella" for a large class of models in the trade literature. Despite their distinct microfoundations, all models in the GEV class can be tied to a common Ricardian interpretation, and can be estimated using the procedure in Section 4.[16] Moreover, macro counterfactuals are also identical, as we explain next.

3.3 Macro Counterfactuals with GEV

We next show the form taken by macro counterfactuals in the GEV class.[17] Heterogeneity in correlation leads to heterogeneity in the gains from trade, but calculations including correlation only require data on expenditure shares across countries, preserving the simplicity of the ACR calculation for the gains from trade.

Specializing (13) to $o = d$ and using the expression for the price index in (12), we can write the real wage in country d as

$$\frac{W_d}{P_d} = \gamma^{-1} T_{dd}^{\frac{1}{\theta}} \left(\widetilde{\pi}_{dd}\right)^{-\frac{1}{\theta}}, \tag{18}$$

where $\widetilde{\pi}_{dd} \equiv \pi_{dd}/G_d^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}) = (P_{dd}/P_d)^{-\theta}$ reflects the real price of domestically produced goods. Using (18), the change in real wages between two equilibria (with the same T_{dd}) is

$$\frac{W_d'/P_d'}{W_d/P_d} = \left(\frac{\widetilde{\pi}_{dd}'}{\widetilde{\pi}_{dd}}\right)^{-\frac{1}{\theta}}. \tag{19}$$

In autarky, country d purchases only its own goods, $\pi_{dd} = 1$, and the price of domestic output is equal to the domestic price level, $P_{dd} = P_d$. The expression in (19) collapses to

$$\frac{W_d/P_d}{W_d^A/P_d^A} = \left(\widetilde{\pi}_{dd}\right)^{-\frac{1}{\theta}}. \tag{20}$$

This expression generalizes the results of ACR to the class of models with GEV import demand systems. With a CES import demand system, the correlation function is additive, and the gains from trade in (20) simplify to the ones in ACR where $\widetilde{\pi}_{dd} = \pi_{dd}$ and two countries with the same self-trade share have the same gains from trade.

[16] By adapting results from the discrete choice literature (Dagsvik, 1995), we can show that GEV import demand systems are dense in the space of import demand system generated by Ricardian models with arbitrary productivity distributions (see Online Appendix O.2).

[17] Online Appendix O.3 presents the equilibrium formally and shows how to apply exact hat-algebra methods.

Importantly, the gains from trade, and other counterfactuals, can be calculated directly from expenditure data, as in ACR. The procedure requires solving a system of equations in $\widetilde{\pi}_{od}$ given data on π_{od}, and estimates of the parameters defining the correlation function (see Appendix C.1 for derivations).

For a given shape parameter θ, the gains from trade starting from autarky under CES are the largest. Once some correlation in productivity exists, these gains are lower. This occurs because the correlation function is convex and linear functions bound convex functions from below. Because an additive correlation function corresponds to the CES case, the right-hand side of (20) is bounded above by the ACR gains from trade. In the extreme, when productivity draws across countries are perfectly correlated, there would be no scope for trade and the gains would be zero. Additionally, the expression in (20) admits the possibility that if two countries have the same self trade share, but one country has very similar technology to other countries—high correlation—their gains from trade will be smaller. In contrast, if that country have dissimilar technology to other countries—low correlation—their gains from trade will be larger. In this way, our framework captures Ricardo's insight on the heterogeneity of the gains from trade across countries.

As for the import demand system, we can write the gains from trade in terms of the technological primitives of Theorem 1. To such end, we focus on the cross-nested CES (CNCES) case of Corollary 1 and use the expenditure shares derived in (16).

Proposition 3. *Under the assumptions of Corollary 1, the gains from trade relative to autarky for destination d in terms of the latent-factor expenditure in* (16) *are*

$$\frac{W_d/P_d}{W_d^A/P_d^A} = \pi_{dd}^{-\frac{1}{\theta}} \left[\sum_{k=1}^{K} \left(\frac{X^*_{kdd}}{\sum_{k'=1}^{K} X^*_{k'dd}} \right) \left(\frac{X^*_{kdd}}{\sum_o X^*_{kod}} \right)^{-\rho_k} \right]^{-\frac{1}{\theta}}. \tag{21}$$

Proof. The result follows from inverting the factor-level demand system. See Appendix B.5. □

For $\rho_k = 0$, the gains from trade reduce to the ACR formula; for $\rho_k = 1$, they collapse to 1. The second term on the right-hand side of (21) measures how much correlation reduces the gains from trade relative to CES. Conditional on expenditure, more correlation within any factor reduces the gains from trade. Additionally, given the distribution of a country's self-trade across factors, the higher factor imports, the lower the gains from trade—and more so if imports occur in high-correlation factors. Intuitively, there are diminishing returns to importing goods made using innovations that have very similar applications across countries.

It is easy to see that when factors are identified with sectors, the gains from trade in (21) can be calculated directly from observed sectoral expenditure.

Before turning to our estimation procedure, in which factors and sectors do not coincide, we provide further intuition on the effects of correlation on the gains from trade using a three-country example.

3.3.1 A three-country example

Consider a three-country world with a correlation function given by

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

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. Using Proposition 3, we can calculate the gains from trade starting from autarky as

$$\frac{W_d/P_d}{W_d^A/P_d^A} = \left[\pi_{dd}^{1-\rho}\left(\pi_{1d} + \pi_{2d}\right)^{\rho}\right]^{-\frac{1}{\theta}} \quad \text{for} \quad d = 1, 2 \quad \text{and} \quad \frac{W_3/P_3}{W_3^A/P_3^A} = \pi_{33}^{-\frac{1}{\theta}}.$$

The gains from trade for countries 1 and 2 depend on the degree of correlation in technology, while the gains for country 3 simply reflect their self-trade share. The gains for trade for country 1 and 2 are given by a Cobb-Douglas combination between each country's expenditure share on its own goods and on the aggregation of its own goods with its peer's goods—the self-trade share if countries 1 and 2 were combined into a single country. *Conditional on observed expenditure*, when $\rho = 0$, we get the ACR formula; for $\rho > 0$, correlation increases effective self trade and implies lower gains from trade; for $\rho \to 1$, the two countries are effectively a single country and the gains from trade depend on their combined self trade.

This intuition carries over to counterfactuals where *equilibrium* expenditure shares change. To see these effects, assume that $T_{od} = T = 1$, for all o, d, and that countries have the same size. Heterogeneity in correlation precludes wage equalization between countries 1 and 2, and country 3, with the (real) wage in the technological-peer countries lower than in the more dissimilar country.[18] In all three countries, wages (and the gains from trade) are decreasing in the parameter ρ, being their highest for $\rho = 0$ and their lowest for $\rho \to 1$. Country 3 gains the most from trade, however, while countries 1 and 2 have lower gains from trade. In turn, trade shares from country 3 increase with ρ, while trade shares from countries 1 and 2 decrease with ρ. In an otherwise identical world, heterogenous correlation impact real wages, expenditure shares, and the gains from trade across countries.

4 Quantitative Application

In this section, we estimate the model and use it for counterfactual exercises. We rely on the results from Theorem 1 and the approximation result in Proposition 1 to estimate the aggregate correlation function. Due to the "latent" nature of technological factors the procedure requires high-dimensional data. For such reason, we estimate a multi-sector version of our model that allows us to incorporate the rich information present in disaggregate trade flow and tariff data. In this way, we infer the latent dimensions of the correlation function from the sectoral data. The model boils down to a *Latent Factor Model* (LFM), which does not pair a latent nest to a sector. It is precisely this generalization that permits departures from the standard multi-sector EK model, which assumes independence across sectors and hence leads to a gravity structure. We refer to this model as the *Sectoral Gravity Model* (SGM).

4.1 Multi-Sector Ricardian Model from Primitives

The economy consists of S sectors. Each sector s in origin o has productivity $Z_{sod}(v)$ for variety $v \in [0,1]$ when delivering to d. Additionally, all varieties shipped from o to d in sector s pay the tariff t_{sod}, so that aggregate productivity is the result of choosing the sector with the highest tariff-adjusted productivity, $Z_{od}(v) \equiv \max_{s=1,\dots,S} Z_{sod}(v)/t_{sod}$. In turn, each destination d sources variety v from the origin with the lowest unit cost, $W_o/Z_{od}(v)$.

Like in our baseline model, sectoral productivity arises from the adoption of glob-

[18] The wage in country 3 is $W_3 = \left(1 + 2^{\frac{1+\theta-\rho}{1+\theta}}\right)^{1/\theta}$, while the wage in countries 1 and 2 is $W = 2^{-\frac{\rho}{1+\theta}} W_3$. Trade shares are: $\pi_{od} = 2^{-\rho} W^{-\theta} = 2^{-\frac{\rho}{1+\theta}} W_3^{-\theta}, o = 1, 2 \quad$ and $\quad \pi_{3d} = W_3^{-\theta}$.

ally available technologies. Sectors are defined by their ability to use innovations, and we assume differences across sectors are independent of geography. That is, each innovation has the same relative applicability across sectors regardless of where the innovation gets applied. The following assumption formalizes this restriction, which boils down to a sectoral extension of Assumption 1.

Assumption 3 (Sectoral Innovation Decomposition). *There exists a measurable space of characteristics $(\mathcal{X}, \mathbb{X})$ and, for each $v \in [0,1]$, an infinite, but countable, set of global innovations, $i = 1, 2, \ldots$, with quality $Q_i(v) > 0$ and characteristics $\chi_i(v) \in \mathcal{X}$, such that*

$$Z_{sod}(v) = \max_{i=1,2,\ldots} Q_i(v) A_{sod}(\chi_i(v)), \quad \text{with} \quad A_{sod}(\chi_i(v)) = A_{od}(\chi_i(v)) B_s(\chi_i(v)),$$

for some measurable functions $\chi \mapsto A_{od}(\chi)$ for each $o, d = 1, \ldots, N$ and $\chi \mapsto B_s(\chi)$ for each $s = 1, \ldots, S$.

The separability of applicability across sectors and countries means that sectors are comparable across all origins and destinations; the sectoral comparative advantage of countries only arises from a country's *access* to innovations. Assumption 3, however, does not impose any restriction on the aggregate correlation function because it implies aggregate productivity satisfying Assumption 1.

We next apply Corollary 1 and Proposition 1 in Section 2.2 to the sectoral model. We partition the innovation characteristics set into K groups and approximate the distribution of applicability on the kth partition with an independent σ_k-Fréchet distribution across origins with scales $A^\theta_{kod} B^\theta_{sk}$. The sectoral productivity distribution implied by this approximation is

$$\mathbb{P}[Z_{sod}(v) \leq z_{so}, \forall s, o] = \exp\left[-\sum_{k=1}^{K} \left(\sum_{s,o} (A^\theta_{kod} B^\theta_{sk} z_{so}^{-\theta})^{\frac{1}{1-\rho_k}} \right)^{1-\rho_k} \mu(\mathcal{X}_k) \right], \tag{22}$$

where $\rho_k = 1 - \theta/\sigma_k$. Aggregate (tariff-adjusted) productivity is also multivariate θ-Fréchet with scale $T_{od} = \sum_{k=1}^{K} T_{kod}$ where $T_{kod} \equiv \Gamma(\rho_k) [\sum_{s=1}^{S} (B_{sk} A_{kod}/t_{sod})^{\frac{\theta}{1-\rho_k}}]^{1-\rho_k} \mu(\mathcal{X}_k)$, and a CNCES correlation function with correlation coefficients ρ_k and weights $\omega_{kod} \equiv T_{kod}/T_{od}$. Latent factors and sectors do not necessarily coincide because sectors can share technologies. The distribution of aggregate productivity, with correlation function given by (9), reflects sectoral primitives, with between-sector correlation arising from the adoption of common technologies across sectors.

Bilateral sectoral expenditure is given by (see Appendix C.3 for derivations)

$$X_{sod} = \sum_{k=1}^{K} \left(\frac{t_{sod}}{t^*_{kod}} \right)^{-\sigma_k} \lambda_{sk} X^*_{kod}, \tag{23}$$

where the variable t^*_{kod} is a factor-level CES tariff index defined by

$$t^*_{kod} \equiv \left(\sum_{s=1}^{S} t_{sod}^{-\sigma_k} \lambda_{sk} \right)^{-\frac{1}{\sigma_k}} \quad \text{with} \quad \lambda_{sk} \equiv \frac{B_{sk}^{\sigma_k}}{\sum_{s'=1}^{S} B_{s'k}^{\sigma_k}}, \tag{24}$$

and the variable X^*_{kod} is latent factor-level expenditure defined in (16). Notice that the share of each factor that gets applied to each sector, λ_{sk}, is exogenous and identical across origins and destinations. This result reflects the assumption that sectoral comparative advantage only arises from a country's ability to use global innovations. Factor-level expenditure and tariff indices reflect each country's access to innovations.

Recall that the case of independence across sectors implies that each sector corresponds to a distinct set of latent factors (formally, $B_{sk} = 0$ for $s \neq k$). In that case: Factor-level expenditure in (16) coincides with observed sector-level expenditure, $X^*_{kod} = X_{sod}$; factor weights are one for $k = s$ and zero otherwise; tariff indices are observed sectoral tariffs, $t^*_{sod} = t_{sod}$; and sectoral elasticities are constant across origins and destinations, $\sigma_k = \sigma_s$. As a consequence, the sectoral independent model can be estimated using gravity regressions, which we present in Section 4.3.1.

4.2 Latent-Factor Model Estimation

When sectors do not correspond to factors, we need an alternative to gravity estimation. We propose an estimator that infers latent expenditure and tariff indices from observed sectoral expenditure and tariffs. We first write (23) as a share of aggregate expenditure

$$\frac{X_{sod}}{X_{od}} = \sum_{k=1}^{K} t_{sod}^{-\sigma_k} \lambda_{sk} \phi_{kod}, \tag{25}$$

where $\phi_{kod} \equiv (t^*_{kod})^{\sigma_k} X^*_{kod}/X_{od}$. This is a (tariff-weighted) factor model where ϕ_{kod} are latent factors. The key result—which is a direct consequence of Theorem 1 and Assumption 3—is that observed sectoral shares are linear in factor weights λ_{sk} and latent factors ϕ_{kod}.

We estimate the factor structure in (25) using 4-digit SITC sectoral tariff and trade flow data from COMTRADE for 1999-2007. Appendix D presents a detail description of the data construction.

Motivated by pseudo Poisson maximum likelihood (PPML) methods used in the gravity literature (Silva and Tenreyro, 2006; Fally, 2015), we choose a Poisson criterion for deviations of observed from predicted sectoral expenditure shares. For a given choice of K, we choose $\Sigma = \{\sigma_k\}_k$, $\Lambda = \{\lambda_{sk}\}_{s,k}$, and $\Phi = \{\phi_{kod}\}_{k,o,d}$ to solve

$$\hat{\Sigma}, \hat{\Lambda}, \hat{\Phi} = \arg\min_{\sigma \geq 0, \Lambda \geq 0, \Phi \geq 0} \sum_{s,o,d} \ell \left(\frac{X_{sod}}{X_{od}}, \sum_{k=1}^{K} t_{sod}^{-\sigma_k} \lambda_{sk} \phi_{kod} \right), \tag{26}$$

where $\ell(x, \hat{x}) \equiv 2(x \ln(x/\hat{x}) - (x - \hat{x}))$.[19]

We adapt techniques from the literature on non-negative matrix factorization to solve the problem in (26).[20] We extend the multiplicative update algorithm in Lee and Seung (1999, 2001) to accommodate both missing data and simultaneous estimation of σ_k. Although latent factor models typically have many solutions, the presence of non-negativity constraints in (26) means that the latent factors can be identified under relatively general conditions (Fu et al., 2019), which, in our context, translate into assumptions related to the sparsity of factors across sectors and countries. Appendix E provides details of the algorithm.

We choose the number of latent factors by estimating (26) for $K = 1, 2, \ldots$, and perform likelihood ratio tests until we fail to reject that the number of latent factors is K versus the alternative of $K + 1$.[21]

Lastly, we need to estimate the parameter θ, which, as shown in Section 3.3, controls the magnitude of the gains from trade. In particular, smaller values of θ imply

[19]The Poisson deviance is the unique likelihood-based criterion that ensures that our estimates of factor-level latent expenditure exactly aggregate to observed trade flows and therefore they are consistent with the structure of the model. Fally (2015) establish that PPML is the unique likelihood-based criterion that preserves the restriction that predicted aggregate expenditure matches observed expenditure.

[20]In principle, the model could be estimated using principal-component analysis (PCA) to infer the factor structure. However, our theory implies that all the factor weights and latent factors are non-negative. PCA estimates typically include negative entries.

[21]The Poisson deviance function is homogenous of degree one and therefore its value depends on scaling of the data. The scaling does not impact the parameters' estimation, but it does matter for likelihood ratio tests. To address this scaling issue, we scale the Poisson deviance by the mean-variance ratio in the data. Effectively, we are relaxing the strict Poisson assumption to assuming that there exists some scaling of the data such that it is approximated by a Poisson distribution, and using the moment condition that the mean equals the variance to standardize the units of the data to be consistent with a Poisson distribution.

larger gains from trade. To get a lower bound on those estimates, we set θ to the minimum estimated σ_k, $\theta = \min_{k=1,\dots,K} \sigma_k$. This is the largest possible value for θ that satisfies that $\theta \leq \sigma_k$ for all k.

With estimates of σ_k, λ_{sk}, ϕ_{kod}, θ, and K, we infer factor-level tariff indices, t^*_{kod}, using (24), and factor-level expenditure, X^*_{kod}, using (25). In turn, we get the weights ω_{kod} from our estimates of latent-factor expenditure.[22]

4.3 Results

Before turning to the estimates of our latent-factor model (LFM), we first estimate the sectoral gravity model (SGM). Because productivity is independent across sectors with symmetric correlation within sectors, sectoral expenditure is CES, and we can estimate the model using gravity-type equations. With those estimates, we reject that elasticities are the same across origins within a sector. We show that these violations of IIA are associated with geographical distance and other characteristics of trading partners. We take this finding as reduced-form evidence against the independence assumption of the SGM, and as motivation for our LFM estimation.

4.3.1 Sectoral gravity model (SGM): estimates and IIA tests.

We estimate the SGM for 14 aggregate sectoral categories from WIOD, denoted by j, and we add a time subscript t to denote the year. Expenditure shares for each $j = 1, \dots, J$ are given by the expression in (16) for $k = j$. To get an estimating equation, we first re-express expenditure as

$$X_{jodt} \equiv \frac{(t^*_{jodt} W_{ot}/A^*_{jodt})^{-\sigma_j}}{\sum_{o'} (t^*_{jo'dt} W_{o't}/A^*_{jo'dt})^{-\sigma_j}} \left(\frac{\sum_{o'} (t^*_{jo'dt} W_{o't}/A^*_{jo'dt})^{-\sigma_j}}{P_{dt}^{\sigma_j}} \right)^{-\theta/\sigma_j} X_d, \tag{27}$$

where $t^*_{jodt} W_{ot}/A^*_{jodt} = \omega_{jodt}^{-1/\theta} P_{jodt}$, the tariff index t^*_{jodt} is the result of aggregating 4-digit SITC tariffs into the aggregate WIOD sectoral category, and A^*_{jodt} is a productivity index (see Appendix C.3 for derivations and index definitions).[23] We assume that $A^*_{jodt} = \exp(a_{jot} + b_{jdt} + c_{jod}) \nu_{jodt}$, where a_{jot} is a sector-origin-time effect, b_{jdt} is a sector-destination-time effect, c_{jod} is a sector-origin-destination effect,

[22] Inverting the expression for X^*_{kod} in (16) yields $\omega_{kod} \equiv \frac{(X^*_{kod})^{1-\rho_k} \left(\sum_{o'=1}^N X^*_{ko'd}\right)^{\rho_k}}{\sum_{k'=1}^K (X^*_{k'od})^{1-\rho_{k'}} \left(\sum_{o'=1}^N X^*_{k'o'd}\right)^{\rho_{k'}}}$.

[23] The aggregation of tariffs follows a model-consistent procedure. See Appendix D.4.

and ν_{jodt} is the residual. Replacing A^*_{jodt} in (27) and re-grouping yield

$$\frac{X_{jodt}}{\sum_{o'=1}^{N} X_{jo'dt}} = \exp\left(\alpha_{jot} + \gamma_{jdt} + \delta_{jod} - \sigma_j \ln t^*_{jodt}\right)\nu_{jodt}, \tag{28}$$

where $\alpha_{jot} \equiv \sigma_j(a_{jot} - \ln W_{ot})$, $\gamma_{jdt} \equiv \sigma_j b_{jdt} + \sigma_j \ln(\sum_{o'}(t^*_{jo'dt} W_{o'}/A^*_{jodt})^{-\sigma_j})^{-1/\sigma_j}$, and $\delta_{jod} \equiv \sigma_j c_{jod}$. Under the no-technology-sharing assumption, expenditure across origins within each WIOD aggregate has a constant elasticity of substitution, $\sigma_j = \theta/(1-\rho_j)$. Further, if $\sigma_j = \sigma$, for all j, we get the independence case of ACR where $\sigma = \theta$ and $\rho = 0$—correlation in productivity is zero not only across sectors but also across origin countries.

Table 1 presents PPML estimates of (28). We estimate specifications with a common tariff coefficient across sectors, with sectoral tariff coefficients, and with interactions of log tariffs with distance and income of the destination and origin countries. The SGM implies that the coefficients on the interaction terms must be zero. The last rows of Table 1 present Wald tests of the null hypothesis that these interaction terms are jointly insignificant.

The first column in Table 1 corresponds to structural estimates of the ACR model and implies $\theta = 2.63$. The estimates in the second column indicate that the tariff elasticity decreases with distance and with income per capita of the destination.[24] The third column includes interactions between origin dummies and log tariffs. The Wald tests indicate that the interaction terms in columns 2 and 3 are jointly significant, suggesting that departures from independence are related to distance, destination income, and also characteristics of the origin country.

The final three specifications in Table 1 allow for the elasticity to be heterogenous across sectors—that is, we move from ACR to SGM. The estimates in column 4 correspond to structural estimates of sectoral elasticities in standard multi-sector Ricardian models. The Wald test rejects the restriction in ACR that elasticities are equal across sectors. Additionally, our Wald tests strongly reject that the tariff interactions in columns 5 and 6 are zero. This result indicates that allowing for heterogenous sectoral elasticities is not enough to capture the substitution patterns in the data since we observe significant violations of IIA in exactly the same patterns

[24]Distance and income are relative to the United States, which means that the elasticity estimates when these covariates are included are interpreted as the elasticity when the United States is both the origin and destination. Hence, in column 2 the USA-to-USA elasticity falls because the United States are both close to itself and have high income relative to most countries.

Table 1: Sectoral Gravity Model (SGM): IIA Tests. PPML.

Dep. variable	$X_{jodt} / \sum_{o'} X_{jo'dt}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\sigma \times \ln t_{jodt}$	2.626***	0.785**	1.654***			
	(0.107)	(0.296)	(0.42)			
$\sigma_1 \times \ln t_{jodt}$				4.150***	2.464***	2.161***
				(0.227)	(0.347)	(0.431)
$\sigma_2 \times \ln t_{jodt}$				4.374***	2.373*	0.326
				(1.145)	(1.207)	(1.193)
$\sigma_3 \times \ln t_{jodt}$				2.211***	0.292	1.623***
				(0.109)	(0.305)	(0.453)
$\sigma_4 \times \ln t_{jodt}$				1.815***	0.202	0.957
				(0.287)	(0.382)	(0.523)
$\sigma_5 \times \ln t_{jodt}$				1.128***	-0.931*	-1.374*
				(0.31)	(0.449)	(0.616)
$\sigma_6 \times \ln t_{jodt}$				1.249***	-0.642	-0.202
				(0.24)	(0.387)	(0.546)
$\sigma_7 \times \ln t_{jodt}$				4.005***	2.108*	0.115
				(0.928)	(0.982)	(0.876)
$\sigma_8 \times \ln t_{jodt}$				2.395***	0.564	0.795
				(0.233)	(0.375)	(0.54)
$\sigma_9 \times \ln t_{jodt}$				0.659**	-1.34***	-0.585
				(0.253)	(0.401)	(0.513)
$\sigma_{10} \times \ln t_{jodt}$				3.256***	1.38***	1.645**
				(0.229)	(0.374)	(0.523)
$\sigma_{11} \times \ln t_{jodt}$				2.827***	1.191**	1.814***
				(0.361)	(0.445)	(0.548)
$\sigma_{12} \times \ln t_{jodt}$				5.174***	3.575***	4.582***
				(0.756)	(0.767)	(0.655)
$\sigma_{13} \times \ln t_{jodt}$				2.360***	0.640	1.245*
				(0.38)	(0.464)	(0.500)
$\sigma_{14} \times \ln t_{jodt}$				2.196***	0.338	0.978*
				(0.278)	(0.398)	(0.499)
$-\Delta \ln Dist_{od} \times \ln t_{jodt}$		-0.184*			-0.316***	
		(0.076)			(0.082)	
$-\Delta \ln Y_{ot} \times \ln t_{jodt}$		-0.123			-0.14	
		(0.082)			(0.083)	
$-\Delta \ln Y_{dt} \times \ln t_{jodt}$		-0.840***			-0.768***	
		(0.154)			(0.148)	
$o \times \ln t_{jodt}$	No	No	Yes	No	No	Yes
Obs.	121,086	121,086	121,086	121,086	121,086	121,086
Deviance	7.025	7.013	4.449	7.003	6.989	4.427
Null Hypothesis		ACR	ACR	ACR	SGM	SGM
χ^2		41.139	224.472	186.254	45.288	235.008
P-Value		0.0	0.0	0.0	0.0	0.0

Notes: Estimates of (28). $Dist_{od}$ = population-weighted bilateral distance between origin o and destination d, Y_{ot} = income per capita in o at time t, $\Delta \ln Dist_{od} = \ln Dist_{od} - \ln Dist_{\text{USA,USA}}$, and $\Delta \ln Y_{ot} = \ln Y_{ot} - \ln Y_{\text{USA},t}$. j refers to an aggregate WIOD sectoral category. All specifications include $j \times o \times t$, $j \times d \times t$, and $j \times o \times d$ fixed effects. Last three rows show results of Wald tests for the null that interaction terms are jointly insignificant. Robust standard errors in parenthesis with levels of significance denoted by *** p < 0.01, and ** p < 0.05 and * p<0.1.

Table 2: Model Order Selection: Likelihood Ratio Test.

	Number of latent factors K				
	4	5	6	7	8
R^2 for sectoral trade flows, X_{sodt}	0.826	0.835	0.937	0.937	0.936
R^2 for sectoral trade shares, X_{sodt}/X_{odt}	0.219	0.245	0.285	0.313	0.341
Observations	5,528,764	5,528,764	5,528,764	5,528,764	5,528,764
Poisson Deviance	292,161	278,379	266,955	256,823	248,288
Degrees of Freedom	37,744	47,180	56,616	66,052	75,488
Log Likelihood Ratio	13,782	11,424	10,132	8,535	
P-Value	0.0	0.0	0.0	1.0	

Notes: Results from estimating LFM with $K = 4, \ldots, 8$. P-values refer to the test with null hypothesis of K latent factors against the alternative of $K + 1$ latent factors.

as for the ACR specification in the first three columns. Our LFM estimation will not impose any structure on sectoral elasticities; instead, we will check ex-post if they correlate with observable variables.

4.3.2 Latent factor model (LFM): estimates

We next present estimates of our LFM directly using the disaggregate 4-digit sectoral data. Our reduced-form evidence suggests that departures from IIA remain after aggregating 4-digit SITC data to the aggregate sectoral categories in WIOD. The implication is that those sectoral aggregates may not correspond to latent technological factors, and that the standard procedure of estimating sectoral gravity regressions may not recover the true import demand system. Our LFM estimation procedure relaxes the assumptions of the SGM, and allow for shared technology across sectors. Heterogenous elasticities of substitution across origins, destinations, and sectors are generated through the latent factors.

First, Table 2 shows that we fail to reject that a model with $K = 8$ is significantly different from one with $K = 7$. Although each additional factor adds 9,436 parameters, only seven factors are necessary to get a reasonable fit to the sectoral trade flow data: they are able to explain 93.7 percent of the variation in the sectoral trade flow data, and 31.3 percent of the variation in expenditure shares within an origin-destination-time pair. This LFM is a relatively parsimonious model; the SGM estimated in column 4 of Table 1 requires 14 factors—one for each WIOD sector.

Next, we examine how technology gets shared across sectors and how much dependence in sectoral productivity is implied by the trade data. Table 3 shows statistics related to estimates of factor elasticities, σ_k and the factor-weight matrix Λ. We rank the latent factors from largest ($F1$) to lowest ($F7$) elasticity. With $\theta = \min_k \sigma_k = 0.375$, this first panel of Table 3 shows that the highest within-factor cross-origin correlation is $\rho_1 = 0.927$ and the lowest is $\rho_7 = 0$. The second panel presents statistics for the matrix Λ. First, the fraction of factor weights that are zero ranges from 6.2 to 23.3 percent–e.g. 23.3 percent of 4-digit SITC sectors do not use technologies related to $F6$. Second, each factor is concentrated in a few 4-digit SITC sectors. The largest weight for each factor ranges from 0.045 to 0.281, with 90 percent of the weights below 0.003 for all factors. Since $\sum_s \lambda_{sk} = 1$, this indicates a very high level of sectoral concentration within each factor. Despite this concentration, the third panel of Table 3 shows that each factor has some weight on the majority of 1-digit SITC sectors (see Appendix Figures F.1 and F.2 for more details).

To measure how these estimates violate the no-shared technology assumption of the SGM, we examine how many 4-digit SITC sectors use each factor (Figure 1a) and how many factors each 4-digit SITC sector uses (Figure 1b). 75 to 95 percent of sectors use each factor. Less than 15 percent of sectors use less than 4 factors, while about 75 percent of sectors use at least six out of the seven factors. Clearly, factors are not unique to sectors, as SGM assumes.

Additionally, we examine how intensively sectors use pairs of factors as well as factors are used by pairs of 2-digit sectors. Figure 1c shows a histogram of a similarity measure that captures the intensity factors weigh on the same sectors, ranging from completely orthogonal (0) to identical weights (1). Similarity is concentrated close to zero for all pairs of factors, consistent with the structural interpretation that factors are groups of distinct technologies, so that they weigh on sectors in distinct ways. Figure 1d shows the distribution of the similarity measure across sector pairs. Even at a 2-digit level of aggregation, many sector pairs weigh on factors identically (high similarity), with some pairs never loading on the same factors (low similarity). Appendix Table F.1 presents summary statistics using 4-digit sectors.

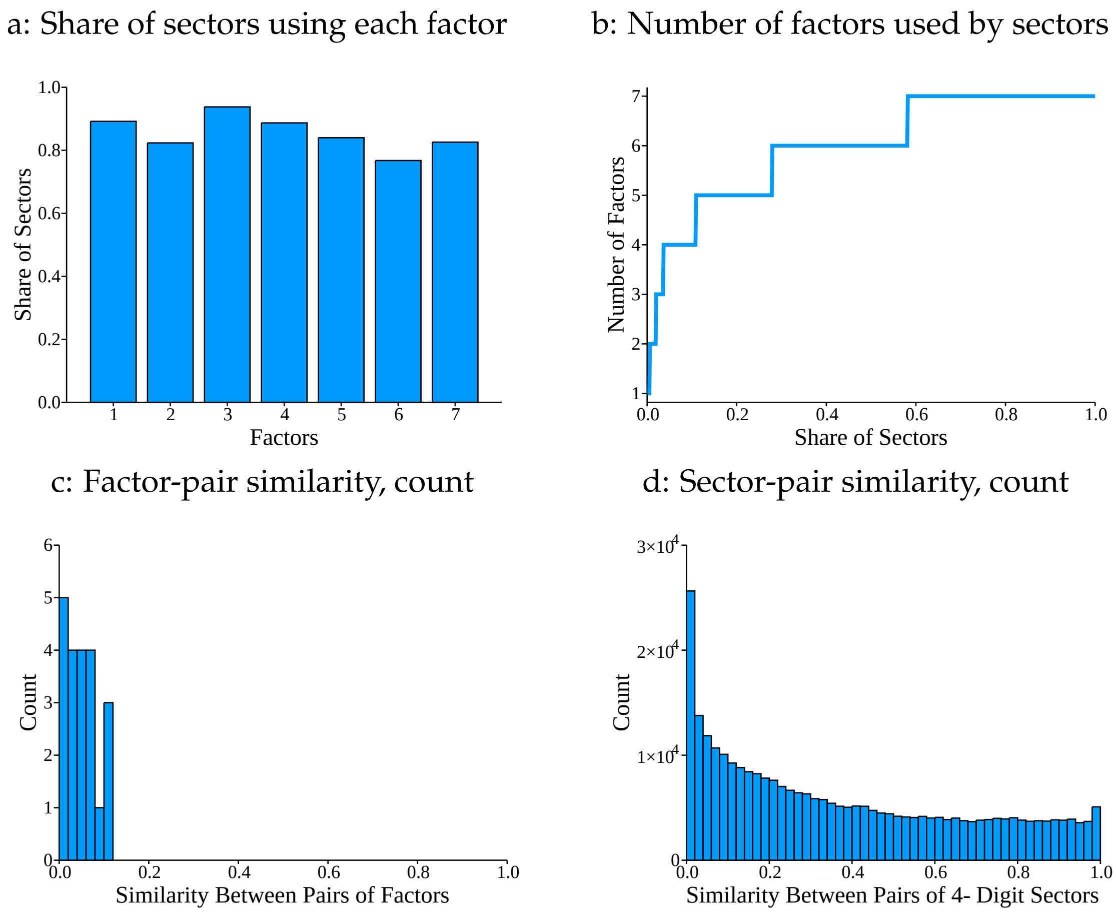
To get some interpretation for each factor, we examine how factors load on sectors (using our estimates of λ_{sk}), as well as patterns of expenditure, export intensity, and domestic absorption by factor (using our estimates of X^*_{kod}). First, the bottom

Table 3: Estimates of factor elasticities and factor weights. Summary Statistics.

	Factor						
	F1	F2	F3	F4	F5	F6	F7
σ_k	5.175	4.869	4.625	1.482	0.671	0.390	0.375
$\rho_k = 1 - \theta/\sigma_k$	0.927	0.923	0.919	0.747	0.44	0.038	0.00
	Factor Weights: Summary Statistics						
Zero Share	0.108	0.177	0.062	0.113	0.16	0.233	0.174
90th Percentile	0.003	0.002	0.003	0.003	0.002	0.001	0.001
99th Percentile	0.016	0.011	0.013	0.019	0.029	0.026	0.018
Maximum	0.045	0.281	0.113	0.036	0.059	0.105	0.277
	Factor Weights: 1-Digit Sectoral Shares						
Food and live animals	0.120	0.008	0.097	0.048	0.323	0.006	0.032
Beverages and tobacco	0.041	0.001	0.023	0.012	0.049	0.004	0.014
Crude materials, except fuels	0.030	0.005	0.029	0.075	0.312	0.017	0.126
Mineral fuels and lubricants	0.006	0.002	0.003	0.106	0.002	0.001	0.492
Animal/veg. oils and waxes	0.010	0.000	0.004	0.003	0.011	0.000	0.001
Chemicals and related, n.e.s.	0.040	0.046	0.305	0.132	0.027	0.036	0.067
Manufactures by material	0.245	0.138	0.065	0.399	0.107	0.026	0.203
Machinery, transport equip.	0.103	0.74	0.294	0.153	0.155	0.807	0.025
Misc. manufactures	0.378	0.05	0.103	0.048	0.013	0.103	0.006
Commodities/trans. n.e.s.	0.000	0.001	0.007	0.002	0.001	0.000	0.020
Residual	0.027	0.010	0.070	0.021	0.000	0.000	0.013
	Factor Statistics						
Share of Expenditure	0.063	0.123	0.117	0.333	0.258	0.071	0.034
Self-Trade Share	0.514	0.455	0.492	0.900	0.962	0.408	0.438
Share of Total Self-Trade	0.044	0.076	0.078	0.406	0.336	0.039	0.02
Share of Total Exports	0.118	0.255	0.228	0.127	0.037	0.161	0.073
Rank 1 Exporter in 1999	CHN	DEU	USA	CAN	USA	USA	RUS
Rank 2 Exporter in 1999	ITA	JPN	DEU	DEU	BRA	JPN	CAN
Rank 3 Exporter in 1999	IND	USA	FRA	USA	CAN	CHN	GBR
Rank 1 Exporter in 2007	CHN	DEU	USA	DEU	BRA	CHN	CAN
Rank 2 Exporter in 2007	ITA	JPN	DEU	NLD	USA	KOR	RUS
Rank 3 Exporter in 2007	IND	USA	FRA	USA	AUS	JPN	AUS

Notes: Residual sectors capture residual expenditure in the WIOD aggregate data attributable to missing observations in the SITC data. See Appendix E.

Figure 1: Factor Weights: Extensive and Intensive Margins.



Notes: Sectors refer to 4-digit SITC sectors. Similarity is cosine similarity, $\sum_i x_i y_i / \sqrt{(\sum_i x_i^2 \sum_i y_i^2)}$.

panel of Table 3 shows that $F4$ and $F5$ make up for the majority of global expenditure, with $F4$ representing ten percent of global trade and $F5$ barely traded. In contrast, the remaining factors are heavily traded, with self-trade shares ranging from around 40 to 50 percent. Turning to the use of factors by sectors, we supplement the 1-digit sectoral shares in Table 3 by reporting the top-three 2-digit shares for each factor in Table 4. These shares reveal the technological identity of each factor. For example, the 2-digit weights reveal that $F1$ is most useful in manufacturing clothing, textiles, and fabrics, but it also captures technologies that are used to produce vegetables and fruits. Table 3 shows that China, Italy, and India are the exporters that use this factor. In contrast, $F2$ is used in the production of "machinery and transport equipment"—in particular road vehicles—as indicated by the rankings in Table 4. Germany, Japan, and the United States are the countries who use this technology the most, as measured by each country's share of total exports

Table 4: Factor Weights: Top-Three 2-Digit Sectors.

	Rank	Code	Description	Weight
F1	1	84	Articles of apparel and clothing accessories	0.231
	2	65	Textile yarn, fabrics, made-up articles, nes, and related products	0.107
	3	05	Vegetables and fruit	0.076
F2	1	78	Road vehicles	0.422
	2	77	Electric machinery, apparatus and appliances, nes, and parts, nes	0.092
	3	74	General industrial machinery and equipment, nes, and parts of, nes	0.063
F3	1	54	Medicinal and pharmaceutical products	0.142
	2	74	General industrial machinery and equipment, nes, and parts of, nes	0.068
	3	51	Organic chemicals	0.063
F4	1	67	Iron and steel	0.136
	2	64	Paper, paperboard, and articles of pulp, of paper or of paperboard	0.105
	3	33	Petroleum, petroleum products and related materials	0.098
F5	1	79	Other transport equipment	0.115
	2	28	Metalliferous ores and metal scrap	0.111
	3	01	Meat and preparations	0.071
F6	1	75	Office machines and automatic data processing equipment	0.283
	2	76	Telecommunications, sound recording and reproducing equipment	0.259
	3	77	Electric machinery, apparatus and appliances, nes, and parts, nes	0.193
F7	1	33	Petroleum, petroleum products and related materials	0.291
	2	32	Coal, coke and briquettes	0.115
	3	68	Non-ferrous metals	0.092

Notes: Factor weights λ_{sk} from estimating LFM in (26).

that rely on this factor. Although $F2$ captures technologies central to the production of cars and is the most concentrated factor, it is also used in many other production processes involving sectors that produce electrical machinery, industrial machinery, power generating machinery, and telecommunications equipment.

For the remaining factors, we see that: $F3$ relates to medical products, chemicals, and industrial equipment; $F4$ and $F5$ are used for making less complex products, related to basic materials, and are the two factors most related to self trade; $F6$ relates to highly specialized manufactured goods such as electronics and scientific instruments; and $F7$, the factor with the lowest cross-country correlation, is related to extraction of energy and minerals, and its major exporters are Russia and Canada.

Finally, our estimates of factor-level expenditure imply correlation function weights, ω_{kod}, that are used to calculate counterfactual exercises. Appendix Figure F.3 shows the distribution of these weights across factors and origins when the United States is the destination market, for 1999 and 2007.

Table 5: Aggregate Elasticities and Country Observables. OLS.

	$\ln -\varepsilon_{oodt}$		$\ln \varepsilon_{oo'dt}$
	(1)		(2)
$\ln \frac{X_{odt}}{X_{dt}}$	-0.206***		
	(0.007)		
$\ln \text{Distance}_{od}$	-0.054***	$\ln \text{Distance}_{oo'}$	-0.173***
	(0.011)		(0.005)
$\lvert \ln Y_{ot} - \ln Y_{dt} \rvert$	0.038***	$\lvert \ln Y_{ot} - \ln Y_{o't} \rvert$	-0.471***
	(0.006)		(0.007)
$o \times t$	Yes	$o \times d \times t$	Yes
$d \times t$	Yes	$o' \times d \times t$	Yes
Obs	8,649	Obs	259,470
R^2	0.724	R^2	0.626
Within-R^2	0.369	Within-R^2	0.032

Notes: Aggregate expenditure elasticities calculated using (17) from latent-factor model (LFM) estimates. Distance_{od} = population-weighted bilateral distance between o and d, and Y_{ot} = income per capita in o at time t. Subscript t refers to years 1999 to 2007.

4.3.3 Implied Expenditure Elasticities

We next compute the aggregate expenditure elasticities, $\varepsilon_{oo'dt}$, using the expressions in Section 3.1 and our LFM estimates of the correlation function. The reduced-form estimates in Table 1 indicate that violations of IIA are related to the distance between the origin and destination, destination income, and other origin-specific factors. Our LFM estimates of the implied own-price elasticities should capture these patterns. The first column of Table 5 presents OLS estimates of the log of (negative) own-price elasticity on country observables. Since the theory implies that the own-price elasticity is inversely proportional to expenditure and country observables are correlated with expenditure, we include the log of the bilateral expenditure share as a control. We also include origin-year and destination-year fixed effects. Own-price elasticities are negatively and significantly related to distance and positively related to income differences between the origin and destination, consistent with the reduced-form evidence in Table 1.[25]

The second column of Table 5 correlates cross-price elasticities with country observables. The reduced-form estimates in Table 1 do not directly speak to these elasticities. However, if similar countries tend to adopt similar technologies—

[25]The inclusion of fixed effects absorbs any correlation between the elasticities and each country's income, so that the positive estimate on the income difference is not inconsistent with the negative estimate on the level of destination income in Table 1.

Table 6: Aggregate Expenditure Elasticities: China serving the United States.

	SGM	LFM		SGM	LFM		SGM	LFM		SGM	LFM
AUS	0.152	0.348	IND	0.271	1.528	HUN	0.572	0.113	DNK	0.366	0.075
AUT	0.146	0.04	IRL	0.226	0.001	GRC	0.119	2.209	RUS	0.052	0.004
BEL	0.125	0.008	ITA	0.224	1.559	GBR	0.191	0.034	SVK	0.151	0.244
BGR	0.242	2.027	JPN	0.272	0.038	TUR	0.186	2.54	DEU	0.215	0.024
BRA	0.116	0.367	KOR	0.432	0.139	FRA	0.169	0.283	PRT	0.251	0.987
CAN	0.089	0.047	MEX	0.385	0.336	SWE	0.186	0.017	ESP	0.151	0.715
CHN	-2.866	-1.301	NLD	0.181	0.003	FIN	0.234	0.007	CZE	0.292	0.184
POL	0.254	0.747	SVN	0.250	1.550	USA	0.102	0.005			

Notes: Expenditure elasticities, $\varepsilon_{o,CHN,USA,t}$, calculated using (17) from latent-factor model (LFM) estimates, and sectoral gravity model (SGM) estimates. Subscript t refers to years 1999 to 2007.

consistent with faster diffusion between nearby countries (Keller, 2002; Bottazzi and Peri, 2003; Comin et al., 2013; Keller and Yeaple, 2013)—then the theory suggests that they will also have higher elasticities. We regress the log of the cross-price elasticity between o and o' within each destination d on the log of distance and the difference in income between o and o'. We include origin-destination-year fixed effects for both origins (which controls for destination expenditure on both countries). We find that more distance and that larger income differences between origins are both associated with lower substitution elasticities, implying that these countries do not trade goods that are close substitutes in export markets.

These broad patterns suggest that the LFM helps to match departures from IIA that SGM does not capture, and implies elasticities consistent with the intuition that similar countries should be more substitutable. To get a sense of the difference between LFM and SGM, Table 6 compares the values for aggregate elasticities for the two models focusing on China as competition for other origins serving the United States, $\varepsilon_{o,CHN,USA}$.[26] LFM estimates indicate that Chinese goods are close substitutes of goods from Turkey, Bulgaria, and Greece, for US consumers, while they are very poor substitutes for goods from Ireland, Netherland, Russia, and the United States itself. These rankings are different when the elasticities are calculated using estimates from the SGM. Relative to LFM, the SGM estimates imply a larger own-price elasticity for China and more similar cross-price elasticities across alternative origins serving the US market. These differences in elasticities can create very different answers to counterfactual exercises, as we show next.

[26] For the SGM, we calculate aggregate elasticities using the estimates in column 4 of Table 1. We use the same θ as LFM to ensure that differences solely come from the correlation function.

4.4 Counterfactual Exercises

Armed with our LFM estimates, we perform two counterfactual exercises. First, we compute the gains from trade starting from autarky. Second, we examine how US protectionism impacts real wages.

4.4.1 The gains from trade.

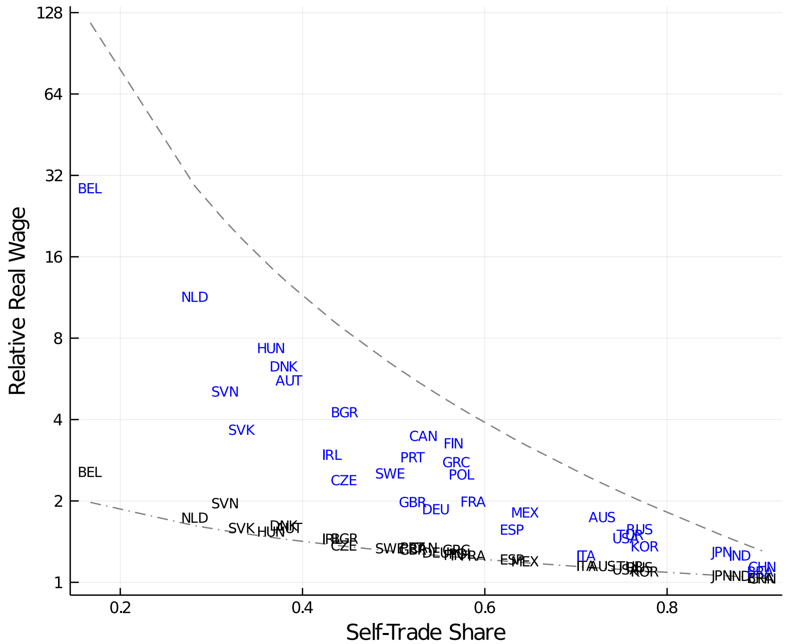
Figure 2 shows the gains from trade against self-trade shares, using the LFM, the SGM, and the ACR model. For the ACR model, we use $\theta = 2.626$ as estimated in column 1 of Table 1. For comparison, we also use the estimate for θ implied by our LFM procedure of 0.375. LFM correlation (blue dots) implies much higher gains from trade than the ACR model (dash-dot line), but as explained in Section 3.3, conditional on the same shape parameter θ, correlation always decreases the gains from trade (blue dots vs dash line). Gains from trade under LFM are much more heterogeneous than under ACR and SGM. For instance, the ACR model delivers the same gains for Bulgaria, Ireland, Czech Republic —because they have very similar self-trade shares. Incorporating LFM correlation entails that gains are different among these three countries depending on the degree of similarity with trading partners. Incorporating correlation through sectors assuming that they do not share technology results in gains from trade that are not very different from the ACR model. This result implies that the way correlation in productivity is introduced into the model matters for counterfactuals. In particular, it is important to let the data reveal correlation patterns rather than assume that those patterns are sector specific.

4.4.2 The cost of protectionism.

Consider the scenario where destination d raises tariffs on origin o'. The impact on the real wage in d can be decomposed as

$$\mathrm{d}\ln\frac{W_d}{P_d} = \underbrace{(1-\pi_{dd})\mathrm{d}\ln\frac{W_d}{W_{o'}}}_{\text{Domestic Wage Effect}} - \underbrace{\sum_{o\neq o' \text{ and } o\neq d}\pi_{od}\mathrm{d}\ln\frac{W_o}{W_{o'}}}_{\text{Third Party Effect}} - \underbrace{\pi_{o'd}\mathrm{d}\ln t_{o'd}}_{\text{Direct Tariff Effect}}, \tag{29}$$

Figure 2: Gains From Trade: ACR, SGM, LFM. Year 2007.



Notes: Real wages in the observed equilibrium relative to autarky real wages. Calculations using estimates from latent factor model (LFM, blue dots), sectoral gravity model (SGM, black dots), ACR model with $\theta = 2.626$ (dash-dot line), and ACR model with $\theta = 0.375$ (dash line).

where $t_{o'd} \equiv [\sum_{k=1}^{K} (t^*_{ko'd})^{-\theta}]^{-\frac{1}{\theta}}$ is the tariff component of $P_{o'd}$.[27] The first term is the effect on real wages in d of changing $W_d/W_{o'}$, while the second term is the effect on countries other than d and o'. The third term is the direct effect of increasing tariffs on real wages in d.

We calculate the elasticity in (29) and its components for each possible pair of countries, in 2007, and compare between the LFM and SGM model. Figure 3 shows density plots and Table 7 shows moments of the percent difference in the components of the real wage elasticity. On average, the direct wage effect is smaller in the LFM by more than 4 percent, and the third-party effect is larger by almost 16 percent. The plot does not show the direct tariff effect because it only depends on observed expenditure and is identical between the two models. The weaker increase in domestic labor demand and stronger increase in third party labor demand increases the impact of raising tariff in the LFM relative to the SGM by almost 18 percent, on average. These results show that the difference in substitution elasticities between the two models leads to first-order differences in counterfactuals, typically making

Figure 3: Diff. in tariff effects, densities.

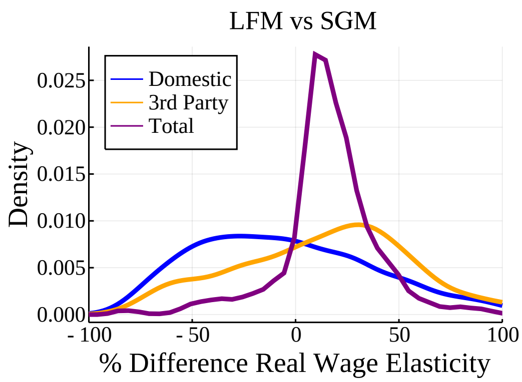


Table 7: Diff. in tariff effects, moments.

%Δ Real Wage Elasticity (LFM vs SGM)			
	Domestic	3rd Party	Total
Mean	-4.08	15.86	17.88
Std.	44.02	45.89	40.72
Skewness	0.54	0.33	7.39
10th Pctl.	-57.79	-47.72	-9.22
25th Pctl.	-38.72	-16.98	7.63
50th Pctl.	-8.72	18.87	15.34
75th Pctl.	25.59	43.48	26.75
90th Pctl.	56.01	68.29	43.70

Notes: Figure 3 shows density plots of the percent difference in the components of (29) between the latent factor model (LFM) and sectoral gravity model (SGM). Blue corresponds to the domestic wage effect, orange corresponds to the third party effect, and purple shows the full effect. The direct tariff effect is identical between the two models. Table 7 shows moments of these densities.

the cost of increasing tariffs larger in the LFM than in the SGM.

Figure 4 focuses on the effect of the United States increasing tariffs on China from 0 to 100 percent, in the LFM and the the SGM, for 2007.[28] For instance, the welfare cost to the United States of imposing a 50-percent tariff on China doubles in the LFM. Mimicking the results in Table 7, the cumulative effect of rising domestic wages is smaller, while the cumulative (negative) effect of rising third-party wages is larger for LFM. This is because US consumers substitute less towards their own goods and more towards third parties in the LFM relative to the SGM (second row of Figure 4). Additionally, the cumulative direct effect of higher tariffs is larger in the LFM because the local direct-tariff effect is proportional to the share of US expenditure on Chinese goods, and as tariffs rise, US consumers shift expenditure away from China, dampening the cumulative direct tariff effect. However, US consumers substitute less away from China in the LFM, which means this dampening effect is weakened, and the cumulative direct cost on US consumers from rising tariffs is larger in the LFM than in the SGM.

The difference in substitution patterns between the two models comes from differences in expenditure shares across technological factors, which correspond to sectors in the SGM. The final row of the figure shows that, for all factors, US expenditure shifts away from China when tariffs rise. However, it does so much

[27]The result follows from $\frac{\partial \ln P_d}{\partial \ln P_{od}} = -\theta^{-1} \frac{\partial \ln G^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta})}{\partial \ln P_{od}} = \pi_{od}$.

[28]We compute each component of the change in US real wages by integrating each term of (29) from 0 to Δt where Δt is the total change in tariffs on the x-axis of Figure 4.

more rapidly for factors with a higher correlation across countries; US consumers are able to find alternative suppliers for products made using those factors. Factors that are not similar across countries are harder to substitute. For instance, $F6$, which corresponds to technologies mostly used in complex manufactured goods, such as electronics, has a very low correlation across countries and US consumers do not rapidly shift their expenditure away from China. Because the SGM pairs latent factors with sectors, it mixes together substitution patterns across technologies and tends to estimate sectoral elasticities that are higher.[29] Consequently, shifts in US expenditure away from Chinese goods occur more rapidly.

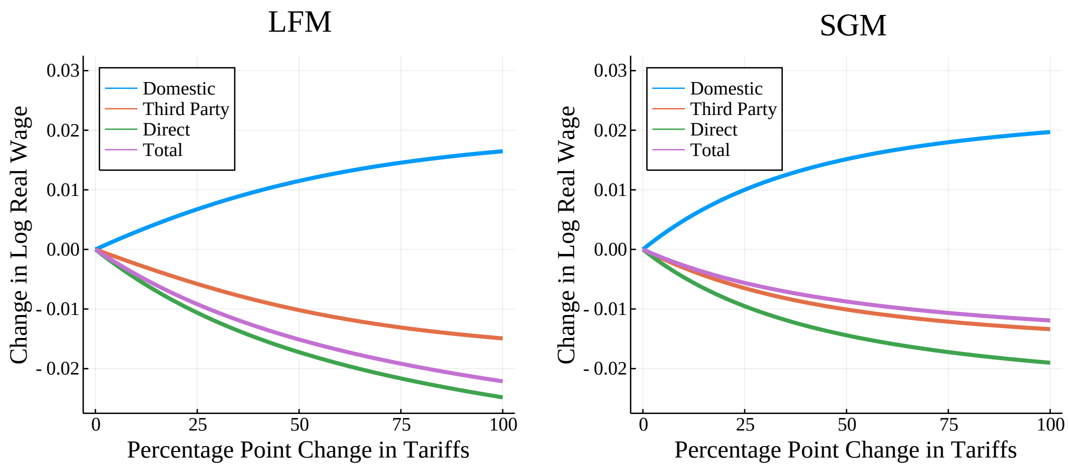
5 Conclusions

This paper is motivated by the old Ricardian idea that a country gains from trading with those countries who are technologically dissimilar. We develop a Ricardian theory of trade that allows for rich patterns of correlation in technology between countries, retains all the tractability of EK-type tools, and spans the entire class of GEV import demand systems. Our key contribution is to provide a structure for technology, based on the adoption of innovations, that is necessary and sufficient to generate max-stable multivariate Fréchet productivity with a general dependence structure. The theory, by relating macro substitutability patterns to underlying technological factors, allows us to develop a procedure to estimate a rich correlation structure based on disaggregate sectoral data. Our quantitative application to a multi-sector trade model reveals that differences in correlation across countries matter: Gains are much more heterogeneous across countries than in the case of independent sectoral productivity.

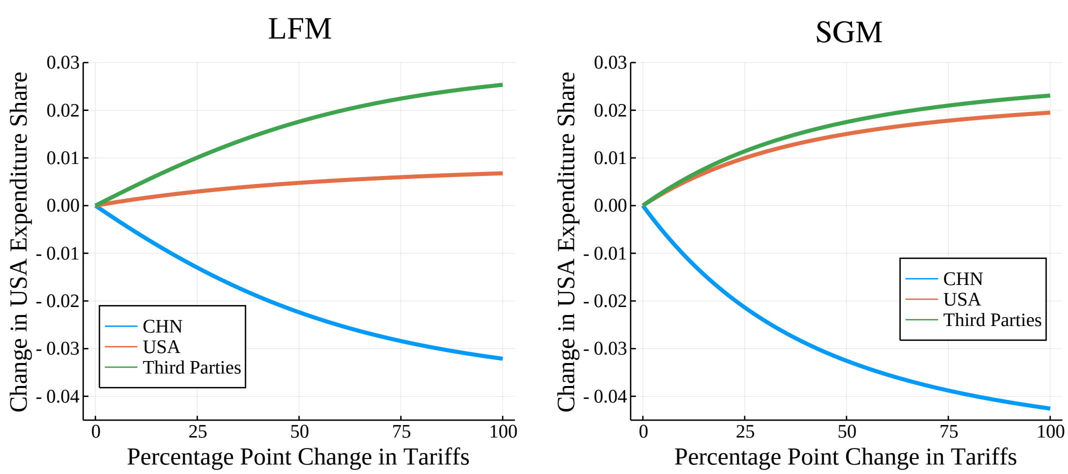
[29] The first factors in LFM absorb most of the covariance between sectoral tariffs and trade flows, implying high elasticities. The remaining covariance is associated with other factors, which end up with low elasticities. The SGM estimates an elasticity for each WIOD aggregate. If weight is put on underlying sectors with high expenditure-tariff covariance, elasticities tend to be higher.

Figure 4: Effect of the United States raising tariffs on China. Year 2007.

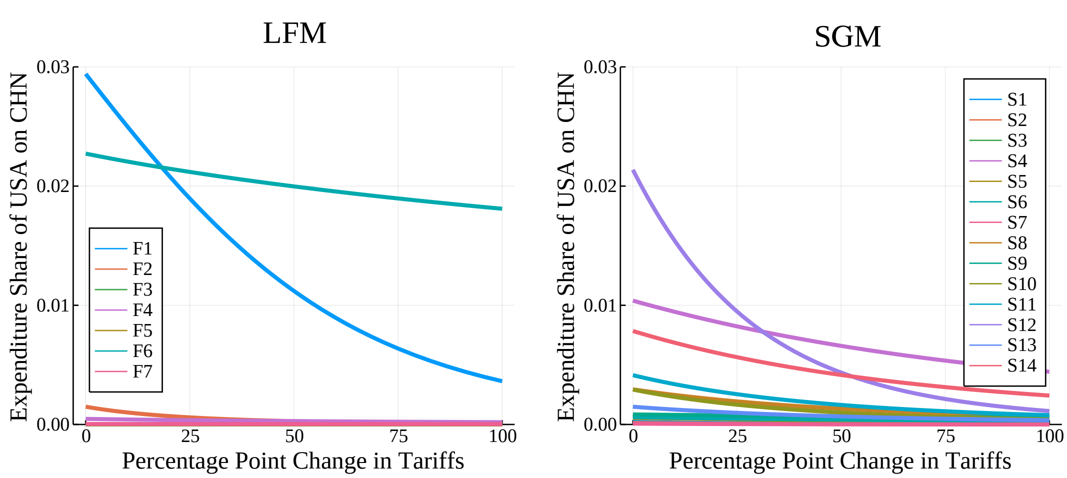
A. US Real Wage.



B. US Expenditure Shares.



C. US Factor-level Expenditure Shares from China.



Notes: LFM = Latent Factor Model; SGM = Sectoral Gravity Model. Changes in log US real wage are decomposed using (29).

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*.

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

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

Baqaee, D. R. and E. Farhi (2019). Networks, barriers, and trade. *NBER Working Paper*.

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.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–90.

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

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

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

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

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

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

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

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

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

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

Dagsvik, J. K. (1994). Discrete and continuous choice, max-stable processes, and independence from irrelevant attributes. *Econometrica*, 1179–1205.

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.

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.

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. (2015). Structural gravity and fixed effects. *Journal of International Economics 97*(1), 76–85.

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.

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.

Fu, X., K. Huang, N. D. Sidiropoulos, , and W.-K. Ma (2019). Nonnegative matrix factorization for signal and data analytics. *IEEE Signal Processing Magazine*, 59–80.

Gudendorf, G. and J. Segers (2010). Extreme-value copulas. In *Copula theory and its applications*, pp. 127–145. Springer.

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

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

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

Keller, W. and S. R. Yeaple (2013). The gravity of knowledge. *American Economic Review 103*(4), 1414–44.

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

Kortum, S. (1997). Research, patenting, and technological change. *Econometrica 65*(6), 1389–1419.

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*.

Lee, D. D. and H. S. Seung (1999). Learning the parts of objects by non-negative matrix factorization. *Nature 401*(6755), 788–791.

Lee, D. D. and H. S. Seung (2001). Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pp. 556–562.

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).

McFadden, D. (1981). *Econometric Models of Probabilistic Choice*. MIT Press: Cambridge.

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.

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.

Resnick, S. I. and R. Roy (1991). Random usc functions, max-stable processes and continuous choice. *The Annals of Applied Probability*, 267–292.

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

Samuelson, P. (1954). The transfer problem and transport costs, ii: Analysis of effects of trade impediments. *Economic Journal 64*(254), 264–289.

Silva, J. S. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and statistics 88*(4), 641–658.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

A Properties of Fréchet Random Variables

Let $\Gamma(x) \equiv \int_0^\infty t^{x-1}e^{-t}\mathrm{d}t$ denote the Gamma function.

Lemma A.1. *If X is Fréchet with scale A and shape $\alpha > 1$, then $\mathbb{E}[X] = \Gamma(1-1/\alpha)A^{1/\alpha}$.*

Proof. $\mathbb{E}[X] = \int_0^\infty z\alpha A z^{-\alpha-1}e^{-Az^{-\alpha}}\mathrm{d}z = \int_0^\infty t^{-1/\alpha}e^{-t}\mathrm{d}tA^{1/\alpha} = \Gamma(1-1/\alpha)A^{1/\alpha}$. $\square$

Lemma A.2. *A vector $(Z_{1d}(v), \ldots, Z_{Nd}(v))$ is max-stable multivariate Fréchet if and only if there exists $\theta > 0$, scale parameters $\{T_{od}\}_{o=1}^N$, and a correlation function, G^d such that*

$$\mathbb{P}\left[Z_{1d}(v) \leq z_1, \ldots, Z_{Nd}(v) \leq z_N\right] = \exp\left[-G^d(T_{1d}z_1^{-\theta}, \ldots, T_{Nd}z_N^{-\theta})\right]. \tag{A.1}$$

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 (A.1) is the joint distribution of productivity across origins.

Suppose productivity is θ-Fréchet. Then its marginal distributions are

$$\mathbb{P}[Z_{od}(v) \leq z] = \exp[-T_{od}z^{-\theta}] \equiv F_{od}(z).$$

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 $\max_{o=1,\ldots,N} x_o^{1/\theta}Z_{od}(v)$ must be distributed as a θ-Fréchet random variable. That is, there exists some $T > 0$ such that

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

Since this holds for any $(x_1, \ldots, x_N) \in \mathbb{R}_+^N$, we have

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

for $T^d : \mathbb{R}_+^N \to \mathbb{R}_+$ defined as

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

Note that the joint distribution of productivity can be written as

$$\begin{aligned}\mathbb{P}[Z_{1d}(v) \leq z_1, \ldots, Z_{Nd}(v) \leq z_N] &= \mathbb{P}[Z_{1d}(v)/z_1 \leq 1, \ldots, Z_{Nd}(v)/z_N \leq 1] \\ &= \mathbb{P}\left[\max_{o=1,\ldots,N} Z_{od}(v)/z_o \leq 1\right].\end{aligned}$$

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

$$\mathbb{P}\left[\max_{o=1,\ldots,N} Z_{od}(v)/z_o \leq 1\right] = \exp\left[-T^d(z_1^{-\theta}, \ldots, z_N^{-\theta})\right].$$

Therefore, the joint distribution of productivity satisfies

$$\mathbb{P}[Z_{1d}(v) \leq z_1, \ldots, Z_{Nd}(v) \leq z_N] = e^{-G^d(T_{1d}z_1^{-\theta}, \ldots, T_{Nd}z_N^{-\theta})},$$

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 G^d is a correlation function. To do so, we show that the copula of productivity,

$$\begin{aligned}C^d(u_1, \ldots, u_N) &\equiv \mathbb{P}\left[F_{1d}(Z_{1d}(v)) \leq u_1, \ldots, F_{Nd}(Z_{Nd}(v)) \leq u_N\right] \\ &= \mathbb{P}\left[Z_{1d}(v) \leq F_{1d}^{-1}(u_1), \ldots, Z_{Nd}(v) \leq F_{Nd}^{-1}(u_N)\right] \\ &\equiv \exp\left[-G^d(T_{1d}F_{1d}^{-1}(u_1)^{-\theta}, \ldots, T_{Nd}F_{Nd}^{-1}(u_N)^{-\theta})\right] \\ &\equiv \exp\left[-G^d(-\ln u_1, \ldots, -\ln u_N)\right],\end{aligned}$$

is an max-stable copula.

To do so, we first show that G^d is homogenous. Fix $(x_1, \ldots, x_N) \in \mathbb{R}_+^N$ and let $\lambda > 0$. We have

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

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

Now, we show that C^d is an max-stable copula. Let $m > 0$ and fix a $(u_1, \ldots, u_N) \in$

$[0,1]^N$. Using the homogeneity of G^d, we have

$$\begin{aligned}
C^d(u_1^{1/m},\ldots,u_N^{1/m})^m &= \exp[-mG^d(-m^{-1}\ln u_1,\ldots,-m^{-1}\ln u_N)] \\
&= \exp[-G^d(-\ln u_1,\ldots,-\ln u_N)] \\
&= C^d(u_1,\ldots,u_N).
\end{aligned}$$

Therefore, C^d is an max-stable copula and G^d is a correlation function.

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

$$\mathbb{P}[Z_{1d}(v) \leq z_1,\ldots,Z_{Nd}(v) \leq z_N] = \exp\left[-G^d(T_{od}z_1^{-\theta},\ldots,T_{Nd}z_N^{-\theta})\right].$$

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

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

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

□

Lemma A.3 (Properties of the Correlation Function). *Let $G : \mathbb{R}_+^N \to \mathbb{R}_+$ be a correlation function. Then:*

1. *G is homogenous of degree one.*
2. *G is unbounded, $G(x_1,\ldots,x_N) \to \infty$ as $x_o \to \infty$ for any $o = 1,\ldots,N$.*
3. *If the mixed partial derivatives of G exist and are continuous up to order N, then the o'th partial derivative of G with respect to o distinct arguments is non-negative if o is odd and non-positive if o is even.*
4. *$G(0,\ldots,0,1,0,\ldots,0) = 1$.*

Proof. Let $G : \mathbb{R}_+^N \to \mathbb{R}_+$ be a correlation function. Then there exists max-stable copula, $C : [0,1]^N \to [0,1]$, such that

$$G(x_1, \ldots, x_N) \equiv -\ln C(e^{-x_1}, \ldots, e^{-x_N}).$$

Recall that for an max-stable copula,

$$C(u_1, \ldots, u_N) = C(u_1^{1/m}, \ldots, u_N^{1/m})^m$$

for all $m > 0$ and $(u_1, \ldots, u_N) \in [0,1]^N$.

We first show that G is homogenous of degree one. Fix $(x_1, \ldots, x_N) \in \mathbb{R}_+^N$ and $\lambda > 0$. We have

$$\begin{aligned}
G(\lambda x_1, \ldots, \lambda x_N) &= -\ln C(e^{-\lambda x_1}, \ldots, e^{-\lambda x_N}) \\
&= -\ln C((e^{-x_1})^\lambda, \ldots, (e^{-x_N})^\lambda) \\
&= -\ln C(e^{-x_1}, \ldots, e^{-x_N})^\lambda \\
&= -\lambda \ln C(e^{-x_1}, \ldots, e^{-x_N}) \\
&= \lambda G(x_1, \ldots, x_N)
\end{aligned}$$

where the third equality uses the fact that C is an max-stable copula. Therefore, G is homogenous of degree one.

The unboundedness property follows from the limiting properties of copulas. Fix o. Then,

$$\lim_{x_o \to \infty} e^{-G^d(x_1, \ldots, x_N)} = \lim_{x_o \to \infty} C(e^{-x_1}, \ldots, e^{-x_N}) = 0$$

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

The sign-switching property simply follows from the non-negativity of joint probability density functions. If the mixed partial derivatives of G exist and are continuous up to order N, then for any integer $M \leq N$ and distinct integers n_m for

$m = 1, \ldots, M$ we have

$$\begin{aligned}
&\frac{\partial^M C(u_1, \ldots, u_N)}{\partial u_{n_1}, \ldots, \partial u_{n_M}} \\
&= \frac{\partial^M \exp\left[-G(-\ln u_1, \ldots, -\ln u_N)\right]}{\partial u_{n_1}, \ldots, \partial u_{n_M}} \\
&= -\exp\left[-G(-\ln u_1, \ldots, -\ln u_N)\right] \frac{\partial^M G(-\ln u_1, \ldots, -\ln u_N)}{\partial u_{n_1}, \ldots, \partial u_{n_M}} \\
&= \exp\left[-G(x_1, \ldots, x_N)\right] \frac{\partial^M G(x_1, \ldots, x_N)}{\partial x_{n_1}, \ldots, \partial x_{n_M}}\Bigg|_{x_1 = -\ln u_1, \ldots, x_N = -\ln u_1} \frac{(-1)^{M-1}}{\prod_{m=1}^M u_{n_m}}
\end{aligned}$$

Since C is a copula, its mixed partial derivatives must be non-negative if they exist. Then the mixed partial derivative of the correlation function is

$$\frac{\partial^M G(x_1, \ldots, x_N)}{\partial u_{n_1}, \ldots, \partial u_{n_M}} = (-1)^{M-1} \prod_{m=1}^M e^{-x_{n_m}} \frac{\partial^M C(u_1, \ldots, u_N)}{\partial u_{n_1}, \ldots, \partial u_{n_M}}\Bigg|_{u_1 = e^{-x_1} \ldots u_N = e^{-x_N}} e^{(x_1, \ldots, x_N)},$$

which is non-negative for odd M and non-positive for even M.

$\square$

Lemma A.4. *Let $\{X_i\}_{i=1}^N$ be α-Fréchet with scale parameters $\{A_i\}_{i=1}^N$ and correlation function $G : \mathbb{R}_+^N \to \mathbb{R}_+$. Then, for any $B_i \geq 0$ $i = 1, \ldots, N$ and $\beta > 0$, $\max_{i=1,\ldots,N} B_i X_i^\beta$ is Fréchet with scale $G(A_1 B_1^{\alpha/\beta}, \ldots, A_N B_N^{\alpha/\beta})$, shape α/β.*

Proof.

$$\begin{aligned}
\mathbb{P}\left[\max_{i=1,\ldots,N} B_i X_i^\beta \leq y\right] &= \mathbb{P}\left[X_1 \leq (y/B_1)^{1/\beta}, \ldots, X_N \leq (y/B_N)^{1/\beta}\right] \\
&= \exp\left[-G(A_1 (y/B_1)^{-\alpha/\beta}, \ldots, A_N (y/B_N)^{-\alpha/\beta})\right] \\
&= \exp\left[-G(A_1 B_1^{\alpha/\beta}, \ldots, A_N B_N^{\alpha/\beta}) y^{-\alpha/\beta}\right].
\end{aligned}$$

$\square$

Lemma A.5. *Let $\{X_i\}_{i=1}^N$ be θ-Fréchet with scale parameters $\{T_i\}_{i=1}^N$ and continuously differentiable correlation function $G : \mathbb{R}_+^N \to \mathbb{R}_+$. Then*

1. $\mathbb{P}\left[X_i = \max_{i'=1,\ldots,N} X_{i'}\right] = \frac{T_i G_i(T_1, \ldots, T_N)}{G(T_1, \ldots, T_N)}$ *where* $G_i(x_1, \ldots, x_N) \equiv \partial G(x_1, \ldots, x_N)/\partial x_i$;
2. $\mathbb{P}\left[X_i \leq x \mid X_i = \max_{i'=1,\ldots,N} X_{i'}\right] = \mathbb{P}\left[\max_{i=1,\ldots,N} X_i \leq x\right].$

Proof. We first prove part 1. We have, for $G_i(x_1, \dots, x_N) = \partial G(x_1, \dots, x_N)/\partial x_i$,

$$
\begin{aligned}
&\mathbb{P}\left[\max_{i'=1,\dots,N} X_{i'} \le x \text{ and } X_i = \max_{i'=1,\dots,N} X_{i'}\right] = \mathbb{P}\left[X_j \le x \text{ and } X_i \le X_{i'}, \forall i' \ne i\right] \\
&= \int_0^x \frac{\partial}{\partial t}\, \mathbb{P}\left[X_{i'} \le z, \forall i' \ne i, \text{ and } X_i \le t\right]\Big|_{z=t} \mathrm{d}t = \int_0^x \frac{\partial}{\partial x_i}\, e^{-G(T_1 x_1^{-\theta}, \dots, T_N x_N^{-\theta})}\Big|_{x_1=t,\dots,x_N=t} \mathrm{d}t \\
&= \int_0^x e^{-G(T_1 x_1^{-\theta}, \dots, T_N x_N^{-\theta})} G_i(T_1 x_1^{-\theta}, \dots, T_N x_N^{-\theta}) T_i \theta x_i^{-\theta-1}\Big|_{x_1=t,\dots,x_N=t} \mathrm{d}t \\
&= \int_0^x e^{-G(T_1,\dots,T_N)t^{-\theta}} G_i(T_1, \dots, T_N) T_i \theta t^{-\theta-1} \mathrm{d}t \\
&= \frac{T_i G_i(T_1, \dots, T_N)}{G(T_1, \dots, T_N)} \int_0^x e^{-G(T_1,\dots,T_N)t^{-\theta}} G(T_1, \dots, T_N) \theta t^{-\theta-1} \mathrm{d}t \\
&= \frac{T_i G_i(T_1, \dots, T_N)}{G(T_1, \dots, T_N)} e^{-G(T_1,\dots,T_N)x^{-\theta}}.
\end{aligned}
$$

Let $x \to \infty$ to get $\mathbb{P}\left[X_i = \max_{i'=1,\dots,N} X_{i'}\right] = \frac{T_i G_i(T_1,\dots,T_N)}{G(T_1,\dots,T_N)}$.

Finally, part 2 follows from the previous results:

$$
\begin{aligned}
\mathbb{P}\left[\max_{i'=1,\dots,N} X_{i'} \le x \mid X_i = \max_{i'=1,\dots,N} X_{i'}\right] &= \frac{\mathbb{P}\left[\max_{i'=1,\dots,N} X_{i'} \le x \text{ and } X_i = \max_{i'=1,\dots,N} X_{i'}\right]}{\mathbb{P}\left[X_i = \max_{i'=1,\dots,N} X_{i'}\right]} \\
&= \frac{\frac{T_i G_i(T_1,\dots,T_N)}{G(T_1,\dots,T_N)} e^{-G(T_1,\dots,T_N)x^{-\theta}}}{\frac{T_i G_i(T_1,\dots,T_N)}{G(T_1,\dots,T_N)}} \\
&= e^{-G(T_1,\dots,T_N)x^{-\theta}} = \mathbb{P}\left[\max_i X_i \le x\right].
\end{aligned}
$$

$\square$

B Proofs

B.1 Proof of Theorem 1

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

$$
\begin{aligned}
&\mathbb{P}\left[Z_{1d}(v) \le z_1, \ldots, Z_{Nd}(v) \le z_N\right] \\
&= \mathbb{P}\left[\max_{i=1,2,\ldots} Q_i(v) A_{od}(\chi_i(v)) \le z_o, \forall o = 1, \ldots, N\right] \\
&= \mathbb{P}\left[Q_i(v) A_{od}(\chi_i(v)) \le z_o, \forall o = 1, \ldots, N, \forall i = 1, 2, \ldots\right] \\
&= \mathbb{P}\left[Q_i(v) \le \min_{o=1,\ldots,N} z_o / A_{od}(\chi_i(v)), \forall i = 1, 2, \ldots\right] \\
&= \mathbb{P}\left[Q_i(v) > \min_{o=1,\ldots,N} z_o / A_{od}(\chi_i(v)), \text{ for no } i = 1, 2, \ldots\right].
\end{aligned}
$$

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

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

for $T_{od} \equiv \int_{\mathcal{X}} A_{od}(\chi)^{\theta} \mathrm{d}\mu(\chi)$. By the monotone convergence theorem,

$$
\begin{aligned}
\mathbb{P}[Z_{od} \le z_o] &= \lim_{z'_o \to \infty \forall o' \neq o} \exp\left[-\int_{\mathcal{X}} \max_{o=1,\ldots,N} A_{od}(\chi)^{\theta} z_o^{-\theta} \mathrm{d}\mu(\chi)\right] \\
&= \exp\left[-\int_{\mathcal{X}} A_{od}(\chi)^{\theta} z_o^{-\theta} \mathrm{d}\mu(\chi)\right] = \exp\left[-T_{od} z_o^{-\theta}\right],
\end{aligned}
$$

for $T_{od} \equiv \int_{\mathcal{X}} A_{od}(\chi)^{\theta} \mathrm{d}\mu(\chi)$. Therefore, the marginal distribution of $Z_{od}(v)$ is Fréchet with scale T_{od} and shape θ.

For each $o = 1, \ldots, N$ let $\alpha_o \geq 0$. Then

$$\begin{aligned}\mathbb{P}\left[\max_{o=1,\ldots,N} \alpha_o Z_{1d}(v) \leq z\right] &= \mathbb{P}\left[Z_{1d}(v) \leq \alpha_o^{-1} z \forall o = 1, \ldots, N\right] \\ &= \exp\left[-\int_{\mathcal{X}} \max_{o=1,\ldots,N} A_{od}(\chi)^{\theta} \alpha_o^{\theta} \mathrm{d}\mu(\chi) z^{-\theta}\right],\end{aligned}$$

which is a Fréchet distribution with scale $\int_{\mathcal{X}} \max_{o=1,\ldots,N} A_{od}(\chi)^{\theta} \alpha_o^{\theta} \mathrm{d}\mu(\chi)$ and shape θ. Therefore, $(Z_{1d}(v), \ldots, Z_{Nd}(v))$ is max-stable multivariate Fréchet. It's correlation function is $G^d(x_1, \ldots, x_N) \equiv \int_{\mathcal{X}} \max_{o=1,\ldots,N} \frac{A_{od}(\chi)^{\theta}}{T_{od}} x_o \mathrm{d}\mu(\chi)$.

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

B.2 Proof of Corollary 1

Suppose Assumptions 1 and 2 hold. Suppose that there exists a partition of characteristics, $\{\mathcal{X}_k\}_{k=1}^K$, such that

$$\int_{\mathcal{X}_k} \mathbf{1}\{A_{od}(\chi) \leq a_o \ \ \forall o\} \frac{\mathrm{d}\mu(\chi)}{\mu(\mathcal{X}_k)} = \exp\left[-\sum_{o=1}^N A_{kod}^{\sigma_k} a_o^{-\sigma_k}\right],$$

for some σ_k and $\{A_{kod}\}_{o=1}^N$, for each $k = 1, \ldots, K$. By Theorem 1, the distribution of productivity is multivariate θ-Fréchet. The scale for o, d is

$$\begin{aligned}T_{od} &= \sum_{k=1}^K \int_{\mathcal{X}_k} A_{od}(\chi)^{\theta} \mathrm{d}\mu(\chi) \\ &= \sum_{k=1}^K \int_0^{\infty} a_o^{\theta} \frac{\partial}{\partial a_o} \exp\left[-A_{kod}^{\sigma_k} a_o^{-\sigma_k}\right] \mathrm{d}a_o \mu(\mathcal{X}_k) \\ &= \sum_{k=1}^K \Gamma(\rho_k) A_{kod}^{\theta} \mu(\mathcal{X}_k),\end{aligned}$$

for $\rho_k = 1 - \theta/\sigma_k$ by Lemma A.4 and Lemma A.1. The correlation function is

$$
\begin{aligned}
G^d(x_1, \ldots, x_N) &= \sum_{k=1}^{K} \int_{\mathcal{X}_k} \max_{o=1,\ldots,N} \frac{A_{od}(\chi)^\theta}{T_{od}} x_o \mathrm{d}\mu(\chi) \\
&= \sum_{k=1}^{K} \int_{\mathbb{R}_+^N} \max_{o=1,\ldots,N} \frac{a_o^\theta}{T_{od}} x_o \frac{\partial^N}{\partial a_1 \ldots \partial a_N} \exp\left[-\sum_{o=1}^{N} A_{kod}^{\sigma_k} a_o^{-\sigma_k}\right] \mathrm{d}(a_1, \ldots, a_N)\mu(\mathcal{X}_k) \\
&= \sum_{k=1}^{K} \int_0^\infty t \frac{\partial}{\partial t} \exp\left[-\sum_{o=1}^{N} A_{kod}^{\sigma_k} \left(\frac{x_o}{T_{od}}\right)^{\sigma_k/\theta} t^{-\sigma_k/\theta}\right] \mu(\mathcal{X}_k) \\
&= \sum_{k=1}^{K} \Gamma(\rho_k) \left(\sum_{o=1}^{N} A_{kod}^{\sigma_k} \left(\frac{x_o}{T_{od}}\right)^{\frac{\sigma_k}{\theta}}\right)^{\frac{\theta}{\sigma_k}} \mu(\mathcal{X}_k) = \sum_{k=1}^{K} \left(\sum_{o=1}^{N} (\omega_{kod} x_o)^{\frac{1}{1-\rho_k}}\right)^{1-\rho_k},
\end{aligned}
$$

for $\omega_{kod} \equiv \Gamma(\rho_k) A_{kod}^\theta \mu(\mathcal{X}_k)/T_{od}$ by Lemma A.4 and Lemma A.1

B.3 Proof of Proposition 1

To simplify notation, we suppress the destination index, d. By Theorem 1, any correlation function can be written as

$$
G(x_1, \ldots, x_N) = \int_{\mathcal{X}} \max_{o=1,\ldots,N} a_o(\chi) x_o \mathrm{d}\mu(\chi),
$$

with $a_o(\chi) \equiv \frac{A_o(\chi)^\theta}{\int_{\mathcal{X}} A_o(\chi)^\theta \mathrm{d}\mu(\chi)}$ for some measurable space $(\mathcal{X}, \mathbb{X})$, measurable functions $\chi \to A_o(\chi)$ for each o, and σ-finite meausre μ such that $\int_{\mathcal{X}} A_{od}(\chi)^\theta \mathrm{d}\mu(\chi) < \infty$ for each o.

Since for each o, $\chi \mapsto a_o(\chi)$ is measurable, there exists a sequence of monotone increasing simple functions that converges pointwise to it. Without loss of generality we can take each function in these sequences to have the form

$$
\tilde{a}_{no}(\chi) = \sum_{m=1}^{M_n} \mathbf{1}\{\mathcal{X}_{mno}\} a_{mno},
$$

where $\{\mathcal{X}_{mno}\}_{m=1}^{M_n}$ is a pairwise disjoint partition of $\mathcal{X}$. Next, let $\{\mathcal{X}_{kn}\}_{k=1}^{K_n}$ be the largest pairwise disjoint partition of $\mathcal{X}$ that is a refinement of $\{\mathcal{X}_{mno}\}_{m=1}^{M_n}$ for every o and define simple functions with respect to this new partition of the form

$$
a_{no}(\chi) = \sum_{k=1}^{K_n} \mathbf{1}\{\mathcal{X}_{kn}\} a_{kno},
$$

such that $\tilde{a}_{no}(\chi) = a_{no}(\chi)$ for all $\chi \in \mathcal{X}$. This construction makes the collection of simple functions defined on a common partition of characteristics.

Define

$$F_n(x_1, \ldots, x_N) \equiv \frac{n}{n+1} \int_{\mathcal{X}} \max_{o=1,\ldots,N} a_{no}(\chi) x_o \mathrm{d}\mu(\chi) = \frac{n}{n+1} \sum_{k=1}^{K_n} \max_{o=1,\ldots,N} a_{kno} x_o \mu(\mathcal{X}_{kn}).$$

Due to the monotone convergence theorem

$$\lim_{n\to\infty} F_n(x_1, \ldots, x_N) = \lim_{n\to\infty} \frac{n}{n+1} \lim_{n\to\infty} \int_{\mathcal{X}} \max_{o=1,\ldots,N} a_{no}(\chi) x_o \mathrm{d}\mu(\chi) = G(x_1, \ldots, x_N).$$

That is, $F_n(x_1, \ldots, x_N)$ converges pointwise to $G(x_1, \ldots, x_N)$.

We now construct an approximating function that is almost a CNCES correlation function and lies between $F_n(x_1, \ldots, x_N)$ and $F_{n+1}(x_1, \ldots, x_N)$. Let $\rho_n \in [0, 1)$. For any $\rho_{kn} \in [\rho_n, 1)$ for $k = 1, \ldots, K_n$ we have

$$\begin{aligned}
&F_n(x_1, \ldots, x_N) = \frac{n}{n+1} \sum_{k=1}^{K_n} \max_{o=1,\ldots,N} a_{kno} x_o \mu(\mathcal{X}_{kn}) \\
&\leq \frac{n}{n+1} \sum_{k=1}^{K_n} \left(\sum_{o=1}^{N} (a_{kno} x_o)^{\frac{1}{1-\rho_{kn}}} \right)^{1-\rho_{kn}} \mu(\mathcal{X}_{kn}) \leq \frac{n}{n+1} \sum_{k=1}^{K_n} N^{1-\rho_{kn}} \max_{o=1,\ldots,N} a_{kno} x_o \mu(\mathcal{X}_{kn}) \\
&\leq \frac{n}{n+1} N^{1-\rho_n} \int_{\mathcal{X}} \max_{o=1,\ldots,N} a_{no}(\chi) x_o \mathrm{d}\mu(\chi) \leq \frac{n}{n+1} N^{1-\rho_n} \int_{\mathcal{X}} \max_{o=1,\ldots,N} a_{n+1,o}(\chi) x_o \mathrm{d}\mu(\chi) \\
&= \frac{n^2+2n}{n^2+2n+1} N^{1-\rho_n} F_{n+1}(x_1, \ldots, x_N).
\end{aligned}$$

If we choose ρ_n sufficient close to 1, we can achieve $\frac{n^2+2n}{n^2+2n+1} N^{1-\rho_n} \leq 1$ since $\frac{n^2+2n}{n^2+2n+1} < 1$ and $\lim_{\rho_n \to 1} N^{1-\rho_n} = 1$. In particular, choose $\rho_n \geq 1 - \frac{\ln \frac{n^2+2n+1}{n^2+2n}}{\ln N}$ and set

$$\hat{F}_n(x_1, \ldots, x_N) \equiv \frac{n}{n+1} \sum_{k=1}^{K_n} \left(\sum_{o=1}^{N} (a_{kno} x_o)^{\frac{1}{1-\rho_{kn}}} \right)^{1-\rho_{kn}} \mu(\mathcal{X}_{kn}).$$

Then

$$F_n(x_1, \ldots, x_N) \leq \hat{F}_n(x_1, \ldots, x_N) \leq F_{n+1}(x_1, \ldots, x_N) \leq G(x_1, \ldots, x_N).$$

Since $F_n(x_1, \ldots, x_N) \to G(x_1, \ldots, x_N)$ pointwise, we also have $\hat{F}_n(x_1, \ldots, x_N) \to$

$G(x_1, \ldots, x_N)$ pointwise. Moreover, since

$$F_n(x_1, \ldots, x_N) \leq \hat{F}_n(x_1, \ldots, x_N) \leq F_{n+1}(x_1, \ldots, x_N) \leq \hat{F}_{n+1}(x_1, \ldots, x_N),$$

the sequence $\hat{F}_n(x_1, \ldots, x_N)$ for $n = 1, 2, \ldots$ is monotone increasing. Since $\hat{F}_n(x_1, \ldots, x_N)$ is continuous, and $G(x_1, \ldots, x_N)$ is also continuous, by Dini's theorem, $\hat{F}_n(x_1, \ldots, x_N) \to G(x_1, \ldots, x_N)$ uniformly on compact sets.

We now construct a CNCES correlation function whose difference with $\hat{F}_n(x_1, \ldots, x_N)$ uniformly converges to zero. Let

$$\delta_{no} \equiv \int_{\mathcal{X}} a_{no}(\chi) \mathrm{d}\mu(\chi) = \sum_{k=1}^{K_n} a_{kno} \mu(\mathcal{X}_{kn}).$$

By the monotone convergence theorem $\lim_{n \to \infty} \delta_{no} = 1$ since $\int_{\mathcal{X}} a_o(\chi) \mathrm{d}\mu(\chi) = 1$.

Define $\omega_{kno} \equiv \delta_{no}^{-1} a_{kno} \mu(\mathcal{X}_{kn})$ and

$$\hat{G}_n(x_1, \ldots, x_N) \equiv \sum_{k=1}^{K_n} \left(\sum_{o=1}^{N} (\omega_{kno} x_o)^{\frac{1}{1-\rho_{kn}}} \right)^{1-\rho_{kn}}.$$

Because $\sum_{k=1}^{K_n} \omega_{kno} = 1$, $\hat{G}_n(x_1, \ldots, x_N)$ is a CNCES correlation function.

Note that $\hat{G}_n(x_1, \ldots, x_N) \geq \hat{F}_n(x_1, \ldots, x_N)$, and hence

$$\begin{aligned}
&|\hat{G}_n(x_1, \ldots, x_N) - \hat{F}_n(x_1, \ldots, x_N)| \\
&= \left| \sum_{k=1}^{K_n} \left(\sum_{o=1}^{N} (\delta_{no}^{-1} a_{kno} x_o)^{\frac{1}{1-\rho_{kn}}} \right)^{1-\rho_{kn}} \mu(\mathcal{X}_{kn}) - \hat{F}_n(x_1, \ldots, x_N) \right| \\
&\leq \left| \left(\max_{o=1,\ldots,N} \delta_{no}^{-1} \right) \sum_{k=1}^{K_n} \left(\sum_{o=1}^{N} (a_{kno} x_o)^{\frac{1}{1-\rho_{kn}}} \right)^{1-\rho_{kn}} \mu(\mathcal{X}_{kn}) - \hat{F}_n(x_1, \ldots, x_N) \right| \\
&= \left| \frac{n+1}{n} \max_{o=1,\ldots,N} \delta_{no}^{-1} - 1 \right| \hat{F}_n(x_1, \ldots, x_N).
\end{aligned}$$

We now show that the CNCES correlation function, $\hat{G}_n(x_1, \ldots, x_N)$, converges uniformly on compact sets to $G(x_1, \ldots, x_N)$. Fix any compact set $X \subset \mathbb{R}_+^N$. We have

$$
\begin{aligned}
& \lim_{n\to\infty} \sup_{(x_1,\ldots,x_N)\in X} |\hat{G}_n(x_1,\ldots,x_N) - G(x_1,\ldots,x_N)| \\
& \le \lim_{n\to\infty} \sup_{(x_1,\ldots,x_N)\in X} |\hat{G}_n(x_1,\ldots,x_N) - \hat{F}_n(x_1,\ldots,x_N)| \\
& \qquad + \lim_{n\to\infty} \sup_{(x_1,\ldots,x_N)\in X} |\hat{F}_n(x_1,\ldots,x_N) - G(x_1,\ldots,x_N)| \\
& \le \lim_{n\to\infty} \left| \frac{n+1}{n} \max_{o=1,\ldots,N} \delta_{no}^{-1} - 1 \right| \lim_{n\to\infty} \sup_{(x_1,\ldots,x_N)\in X} \hat{F}_n(x_1,\ldots,x_N) \\
& \qquad + \lim_{n\to\infty} \sup_{(x_1,\ldots,x_N)\in X} |\hat{F}_n(x_1,\ldots,x_N) - G(x_1,\ldots,x_N)| \\
& \le \lim_{n\to\infty} \left| \frac{n+1}{n} \max_{o=1,\ldots,N} \delta_{no}^{-1} - 1 \right| \sup_{(x_1,\ldots,x_N)\in X} G(x_1,\ldots,x_N) = 0.
\end{aligned}
$$

The last inequality holds since $\hat{F}_n(x_1, \ldots, x_N)$ converges uniformly to $G(x_1, \ldots, x_N)$ on compact sets, which means that the second term is zero, and also means we can interchange the limit and supremum in the first term. Therefore, $\hat{G}_n(x_1, \ldots, x_N)$ converges uniformly to $G(x_1, \ldots, x_N)$ on compact sets.

B.4 Proof of Proposition 2

Since destination prices are given by (2), the price index in destination d is

$$
\begin{aligned}
P_d &= \left[\int_0^1 \min_{o=1,\ldots,N} (W_o/Z_{od}(v))^{1-\eta} \mathrm{d}v \right]^{\frac{1}{1-\eta}} \\
&= \left[\mathbb{E} \max_{o=1,\ldots,N} (Z_{od}(v)/W_o)^{\eta-1} \mathrm{d}v \right]^{\frac{1}{1-\eta}} \\
&= \gamma G^d \left(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta} \right)^{-\frac{1}{\theta}} \\
&= G^d \left(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta} \right)^{-\frac{1}{\theta}},
\end{aligned}
$$

where $P_{od} \equiv \gamma T_{od}^{-1/\theta} W_o$, $\gamma = \Gamma\left(\frac{\theta+1-\eta}{\theta}\right)^{\frac{1}{1-\eta}}$, due to Appendix Lemma A.4 and Appendix Lemma A.1.

The expenditure share of d on o is

$$
\begin{aligned}
\pi_{od} \equiv \frac{X_{od}}{X_d} &= \int_0^1 \left(\frac{P_d(v)}{P_d}\right)^{1-\eta} \mathbf{1}\left\{\frac{W_o}{Z_{od}(v)} = P_d(v)\right\} \mathrm{d}v \\
&= \mathbb{E}\left(P_d \max_{o'=1,\ldots,N} \frac{Z_{o'd}(v)}{W_{o'}}\right)^{\eta-1} \mathbf{1}\left\{\frac{Z_{od}(v)}{W_o} = \max_{o'=1,\ldots,N} \frac{Z_{o'd}(v)}{W_{o'}}\right\} \\
&= \mathbb{E}\left[\left(P_d \max_{o'=1,\ldots,N} \frac{Z_{o'd}(v)}{W_{o'}}\right)^{\eta-1} \mid \frac{Z_{od}(v)}{W_o} = \max_{o'=1,\ldots,N} \frac{Z_{o'd}(v)}{W_{o'}}\right] \mathbb{P}\left[\frac{Z_{od}(v)}{W_o} = \max_{o'=1,\ldots,N} \frac{Z_{o'd}(v)}{W_{o'}}\right] \\
&= \mathbb{E}\left[\left(P_d \max_{o'=1,\ldots,N} \frac{Z_{o'd}(v)}{W_{o'}}\right)^{\eta-1}\right] \mathbb{P}\left[\frac{Z_{od}(v)}{W_o} = \max_{o'=1,\ldots,N} \frac{Z_{o'd}(v)}{W_{o'}}\right] \\
&= \mathbb{E}\left[\left(\frac{P_d(v)}{P_d}\right)^{1-\eta}\right] \mathbb{P}\left[\frac{Z_{od}(v)}{W_o} = \max_{o'=1,\ldots,N} \frac{Z_{o'd}(v)}{W_{o'}}\right], \\
&= \mathbb{P}\left[\frac{W_o}{Z_{od}(v)} = \min_{o'=1,\ldots,N} \frac{W_{o'}}{Z_{o'd}(v)}\right]
\end{aligned}
$$

using part 2 of Appendix Lemma A.5 and the previous result for the price level. By part 1 of Appendix Lemma A.5,

$$
\mathbb{P}\left[\frac{W_o}{Z_{od}(v)} = \min_{o'=1,\ldots,N} \frac{W_{o'}}{Z_{o'd}(v)}\right] = \frac{T_{od} W_o^{-\theta} G_o^d(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta})}{G^d(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta})} = \frac{P_{od}^{-\theta} G_o^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta})}{G^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta})}.
$$

B.5 Proof of Proposition 3

Let Assumptions 1 and 2 hold and suppose there exists a partition of characteristics, $\{\mathcal{X}_k\}_{k=1}^K$, such that, for each k, applicability restricted to $\mathcal{X}_k$ is independent σ_k-Fréchet across origins. Then productivity is θ-Fréchet. The correlation function is CNCES as in (9) and expenditure shares are given by (16).

First, we can recover bilateral import prices up to a factor-destination constant using the within-factor component in (16),

$$
\frac{\omega_{kod}^{-\frac{1-\rho_k}{\theta}} P_{od}/P_d}{\left(\sum_{o'=1}^N \omega_{ko'd} (P_{o'd}/P_d)^{-\frac{\theta}{1-\rho_k}}\right)^{-\frac{1-\rho_k}{\theta}}} = \left(\frac{X_{kod}^*}{\sum_{o'=1}^N X_{ko'd}^*}\right)^{-\frac{1-\rho_k}{\theta}}.
$$

The denominator on the left-hand-side can be recovered from the between-factor component in (16),

$$\left(\sum_{o'=1}^{N}\omega_{ko'd}\left(P_{o'd}/P_{d}\right)^{-\frac{\theta}{1-\rho_k}}\right)^{-\frac{1-\rho_k}{\theta}} = \left(\sum_{o'=1}^{N}\frac{X^*_{ko'd}}{X_d}\right)^{-\frac{1}{\theta}}.$$

Together we have

$$\omega_{kod}^{-\frac{1-\rho_k}{\theta}} P_{od}/P_d = \left(\frac{X^*_{kod}}{\sum_{o'=1}^{N} X^*_{ko'd}}\right)^{-\frac{1-\rho_k}{\theta}}\left(\sum_{o'=1}^{N}\frac{X^*_{ko'd}}{X_d}\right)^{-\frac{1}{\theta}}.$$

Take this result to a power of $-\theta$ and sum across k to get

$$(P_{od}/P_d)^{-\theta} = \sum_{k=1}^{K}\left(\frac{X^*_{kod}}{\sum_{o'=1}^{N} X^*_{ko'd}}\right)^{1-\rho_k}\left(\sum_{o'=1}^{N}\frac{X^*_{ko'd}}{X_d}\right) = \sum_{k=1}^{K}\left(\frac{X^*_{kod}}{X_d}\right)^{1-\rho_k}\left(\sum_{o'=1}^{N}\frac{X^*_{ko'd}}{X_d}\right)^{\rho_k}.$$

The gains from trade relative to autarky are then

$$\begin{aligned}\frac{W_d/P_d}{W_d^{\mathrm{A}}/P_d^{\mathrm{A}}} &= \left(\sum_{k=1}^{K}\left(\frac{X^*_{kdd}}{X_d}\right)^{1-\rho_k}\left(\sum_{o=1}^{N}\frac{X^*_{kod}}{X_d}\right)^{\rho_k}\right)^{-\frac{1}{\theta}}\\ &= \pi_{dd}^{-\frac{1}{\theta}}\left(\sum_{k=1}^{K}\frac{X^*_{kdd}}{X_{dd}}\left(\sum_{o=1}^{N}\frac{X^*_{kod}}{X^*_{kdd}}\right)^{\rho_k}\right)^{-\frac{1}{\theta}}.\end{aligned}$$

C Derivations

C.1 Solving for Correlation-Adjusted Trade Shares

Using the homogeneity of degree zero of G^d_o, expenditure shares in (13) can be written as

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

where $P^{-\theta}_{od} \equiv T_{od}W^{-\theta}_o$. Noting that $\widetilde{\pi}_{od} \equiv \pi_{od}/G^d_o\left(P^{-\theta}_{1d}, \ldots, P^{-\theta}_{Nd}\right) = (P_{od}/P_d)^{-\theta}$ yields the system

$$\pi_{od} = \widetilde{\pi}_{od}G^d_o\left(\widetilde{\pi}_{1d}, \ldots, \widetilde{\pi}_{Nd}\right) \quad \text{for} \quad o = 1, \ldots, N. \tag{C.1}$$

Given expenditure share data and the correlation function for each $d = 1, \ldots, N$, the expression in (C.1) constitutes a system of N equations in the N unknown correlation-adjusted expenditure shares across origins.

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 (C.1), satisfies strict gross substitutability and is homogenous of degree one. As a result, it is injective and there is a unique solution for $\{\widetilde{\pi}_{od}\}^N_{o=1}$, given $\{\pi_{od}\}^N_{o=1}$ (see, for instance, Berry et al., 2013).

C.2 Expenditure Elasticities

The within-k elasticity of substitution is

$$\begin{aligned}
\frac{\partial \ln \frac{X^*_{kod}}{X_d}}{\partial \ln \frac{P_{o'd}}{P_d}} &= \frac{\partial \ln}{\partial \ln \frac{P_{o'd}}{P_d}} \frac{\left(\omega_{kod}(P_{od}/P_d)^{-\theta}\right)^{\frac{1}{1-\rho_k}}}{\sum_{n=1}^N \left(\omega_{knd}(P_{nd}/P_d)^{-\theta}\right)^{\frac{1}{1-\rho_k}}} \left(\sum_{n=1}^N \left(\omega_{knd}(P_{nd}/P_d)^{-\theta}\right)^{\frac{1}{1-\rho_k}}\right)^{1-\rho_k} \\
&= -\rho_k \frac{\partial}{\partial \ln \frac{P_{o'd}}{P_d}} \ln\left(\sum_{n=1}^N \left(\omega_{knd}(P_{nd}/P_d)^{-\theta}\right)^{\frac{1}{1-\rho_k}}\right) - \mathbf{1}\{o = o'\}\frac{\theta}{1-\rho_k} \\
&= \frac{\rho_k\theta}{1-\rho_k} \frac{\left(\omega_{ko'd}(P_{o'd}/P_d)^{-\theta}\right)^{\frac{1}{1-\rho_k}}}{\sum_{n=1}^N \left(\omega_{knd}(P_{nd}/P_d)^{-\theta}\right)^{\frac{1}{1-\rho_k}}} - \mathbf{1}\{o = o'\}\frac{\theta}{1-\rho_k} \\
&= \theta\left[\frac{\rho_k}{1-\rho_k} \frac{X^*_{ko'd}}{\sum_{n=1}^N X^*_{knd}} - \mathbf{1}\{o = o'\}\frac{1}{1-\rho_k}\right].
\end{aligned}$$

The aggregate elasticity of substitution is then

$$
\begin{aligned}
\varepsilon_{oo'd} \equiv \frac{\partial \ln \frac{X_{od}}{X_d}}{\partial \ln \frac{P_{o'd}}{P_d}} = \frac{\partial \ln}{\partial \ln \frac{P_{o'd}}{P_d}} \sum_{k=1}^{K} \frac{X^*_{kod}}{X_d} = \sum_{k=1}^{K} \frac{X^*_{kod}}{X_{od}} \frac{\partial \ln (X^*_{kod}/X_d)}{\partial \ln \frac{P_{o'd}}{P_d}} \\
= \theta \sum_{k=1}^{K} \frac{X^*_{kod}}{X_{od}} \left[\frac{\rho_k}{1-\rho_k} \frac{X^*_{ko'd}}{\sum_{n=1}^{N} X^*_{knd}} - \mathbf{1}\{o=o'\} \frac{1}{1-\rho_k} \right].
\end{aligned}
$$

C.3 Latent Factor Model

Define $P_{sod} \equiv \gamma T_{sod}^{-1/\theta} W_o$, where $T_{sod} \equiv \sum_{k=1}^{K} \Gamma(\rho_k) A^\theta_{kod} B^\theta_{sk} t^{-\theta}_{sod} \mu(\mathcal{X}_k)$. Generalizing (16) to be at the sector-origin level rather than just the origin level yields

$$
X_{sod} = \sum_{k=1}^{K} \frac{\omega_{skod}^{\frac{1}{1-\rho_k}} (P_{sod}/P_d)^{-\frac{\theta}{1-\rho_k}}}{\sum_{s'=1}^{S} \sum_{o'=1}^{N} \omega_{s'ko'd}^{\frac{1}{1-\rho_k}} (P_{s'o'd}/P_d)^{-\frac{\theta}{1-\rho_k}}} \left(\sum_{s'=1}^{S} \sum_{o'=1}^{N} \omega_{s'ko'd}^{\frac{1}{1-\rho_k}} (P_{s'o'd}/P_d)^{-\frac{\theta}{1-\rho_k}} \right)^{1-\rho_k} X_d,
$$

where

$$
\omega_{skod} \equiv \frac{\Gamma(\rho_k) A^\theta_{kod} B^\theta_{sk} t^{-\theta}_{sod} \mu(\mathcal{X}_k)}{\sum_{k'=1}^{K} \Gamma(\rho_{k'}) A^\theta_{k'od} B^\theta_{sk'} t^{-\theta}_{sod} \mu(\mathcal{X}_{k'})}.
$$

Aggregating over sectors yields aggregate expenditure,

$$
X_{od} = \sum_{k=1}^{K} \frac{\sum_{s=1}^{S} \omega_{skod}^{\frac{1}{1-\rho_k}} (P_{sod}/P_d)^{-\frac{\theta}{1-\rho_k}}}{\sum_{s=1}^{S} \sum_{o'=1}^{N} \omega_{sko'd}^{\frac{1}{1-\rho_k}} (P_{so'd}/P_d)^{-\frac{\theta}{1-\rho_k}}} \left(\sum_{s=1}^{S} \sum_{o'=1}^{N} \omega_{sko'd}^{\frac{1}{1-\rho_k}} (P_{so'd}/P_d)^{-\frac{\theta}{1-\rho_k}} \right)^{1-\rho_k} X_d.
$$

Note that

$$
\begin{aligned}
\omega_{skod}^{\frac{1}{1-\rho_k}} (P_{sod})^{-\frac{\theta}{1-\rho_k}} &= (A^\theta_{kod} B^\theta_{sk} t^{-\theta}_{sod} \mu(\mathcal{X}_k) W_o^{-\theta})^{\frac{1}{1-\rho_k}} \\
&= (A^\theta_{kod} \mu(\mathcal{X}_k) W_o^{-\theta})^{\frac{1}{1-\rho_k}} (t^*_{kod})^{-\frac{\theta}{1-\rho_k}} \left(\frac{t_{sod}}{t^*_{kod}} \right)^{-\frac{\theta}{1-\rho_k}} \lambda_{sk} \sum_{s=1}^{S} B^{\sigma_k}_{sk} \\
&= \left(\frac{t_{sod}}{t^*_{kod}} \right)^{-\frac{\theta}{1-\rho_k}} \lambda_{sk} \left(\frac{t^*_{kod} W_o}{A^*_{kod}} \right)^{-\frac{\theta}{1-\rho_k}},
\end{aligned}
$$

where $t^*_{kod} \equiv \left(\sum_{s=1}^{S} t^{-\sigma_k}_{sod} \lambda_{sk} \right)^{-\frac{1}{\sigma_k}}$, $\lambda_{sk} \equiv \frac{B^{\sigma_k}_{sk}}{\sum_{s'=1}^{S} B^{\sigma_k}_{s'k}}$, and $A^*_{kod} \equiv \left(\sum_{s=1}^{S} B^{\sigma_k}_{sk} \right)^{\frac{1}{\sigma_k}} A_{kod} \mu(\mathcal{X}_k)^{1/\theta}$. Then,

$$
\sum_{s=1}^{S} \omega_{skod}^{\frac{1}{1-\rho_k}} (P_{sod})^{-\frac{\theta}{1-\rho_k}} = \sum_{s=1}^{S} (A^\theta_{kod} B^\theta_{sk} t^{-\theta}_{sod} \mu(\mathcal{X}_k) W_o^{-\theta})^{\frac{1}{1-\rho_k}} \quad = (t^*_{kod} W_o / A^*_{kod})^{-\frac{\theta}{1-\rho_k}},
$$

since $\sum_s \lambda_{sk} = 1$.

Replacing in aggregate expenditure yields

$$X_{od} = \sum_{k=1}^{K} \frac{(t^*_{kod} W_o / A^*_{kod})^{-\frac{\theta}{1-\rho_k}}}{\sum_{o'=1}^{N} (t^*_{ko'd} W_{o'} / A^*_{ko'd})^{-\frac{\theta}{1-\rho_k}}} \left(\sum_{o'=1}^{N} \left(\frac{t^*_{ko'd} W_{o'} / A^*_{ko'd}}{P_d} \right)^{-\frac{\theta}{1-\rho_k}} \right)^{1-\rho_k} X_d,$$

Define $T_{od} \equiv \sum_{k=1}^{K} (A^*_{kod}/t^*_{kod})^\theta$, $\omega_{kod} \equiv (A^*_{kod}/t^*_{kod})^\theta / \sum_{k'=1}^{K} (A^*_{k'od}/t^*_{k'od})^\theta$, and $P_{od} \equiv T_{od}^{-1/\theta} W_o$. This expression coincides with the one in (16).

We can further substitute into the sector-level demand system to obtain (23),

$$X_{sod} = \sum_{k=1}^{K} \frac{\left(\frac{t_{sod}}{t^*_{kod}}\right)^{-\frac{\theta}{1-\rho_k}} \lambda_{sk} (t^*_{kod} W_o / A^*_{kod})^{-\frac{\theta}{1-\rho_k}}}{\sum_{s'=1}^{S} \sum_{o'=1}^{N} \omega_{s'ko'd}^{\frac{1}{1-\rho_k}} (P_{s'o'd}/P_d)^{-\frac{\theta}{1-\rho_k}}} \left(\sum_{s'=1}^{S} \sum_{o'=1}^{N} \omega_{s'ko'd}^{\frac{1}{1-\rho_k}} (P_{s'o'd}/P_d)^{-\frac{\theta}{1-\rho_k}} \right)^{1-\rho_k} X_d$$

$$= \sum_{k=1}^{K} \left(\frac{t_{sod}}{t^*_{kod}} \right)^{-\frac{\theta}{1-\rho_k}} \lambda_{sk} \frac{\omega_{kod}^{\frac{1}{1-\rho_k}} (P_{od}/P_d)^{-\frac{\theta}{1-\rho_k}}}{\sum_{o'=1}^{N} \omega_{ko'd}^{\frac{1}{1-\rho_k}} (P_{o'd}/P_d)^{-\frac{\theta}{1-\rho_k}}} \left(\sum_{o'=1}^{N} \omega_{ko'd}^{\frac{1}{1-\rho_k}} (P_{o'd}/P_d)^{-\frac{\theta}{1-\rho_k}} \right)^{1-\rho_k} X_d$$

$$= \sum_{k=1}^{K} \left(\frac{t_{sod}}{t^*_{kod}} \right)^{-\frac{\theta}{1-\rho_k}} \lambda_{sk} X^*_{kod},$$

where $\sigma_k = \frac{\theta}{1-\rho_k}$.

D Data Construction

For our quantitative analysis, we use 4-digit SITC trade flow data and tariff data from the United Nations COMTRADE Database. We also use trade flow data in aggregated sector categories from the World Input-Output Database (WIOD). Gravity covariates are from the Centre D'Études Prospectives et d'Informations Internationales (CEPII).

D.1 Map from SITC Codes to WIOD Sectors

The WIOD data allow us to compute the total value of trade between a sample of 40 countries across 35 sectors from 1995 through 2011. While the sector classification in this dataset comes from aggregating underlying data classified according to the third revision of the International Standard Industrial Classification (ISIC), the COMTRADE tariff data are 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 aggregates of WIOD sectors.

First, we match ISIC and SITC definitions using existing correspondences to Harmonized System (HS) product definitions. These correspondences come from the World Bank's World Integrated Trade Solution (WITS).[30] This merge matches 5,701 products out of 5,705 total HS products, creating 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 25 aggregates of WIOD sectors. This leaves products in the ISIC code 99 ("Goods n.e.c.") without a WIOD sector definition. This results in a HS-product-level dataset with labels for the 25 WIOD aggregates and 764 SITC codes.

At this point, there are two issues left to 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. 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.

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

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 the 12 remaining codes.This results in a map from 925 4-digit SITC codes to 25 WIOD aggregates.

D.2 Reconciling WIOD and COMTRADE Data

We drop those countries in WIOD with completely missing data in COMTRADE, and aggregate the 35 WIOD sectors to the 25 aggregates in our concordance with 4-digit SITC codes, and restrict the sample to 1999 through 2007. These restrictions leave a balanced sample of 25 WIOD aggregates for 31 countries over 9 years.[31] Finally, we keep the 14 WIOD aggregates that correspond to traded goods.

We then turn to the COMTRADE data. First, we drop all countries not in our WIOD sample and drop a few instances of self-trade that only appear in a few countries. We then merge the data with WIOD data, scaling units of both datasets to be in thousands of US dollars, and adding missing observations to fill in all possible pairs of the 925 SITC codes, 31 origin countries, 31 destination countries, and 9 years.

Next, we compare the WIOD aggregate level expenditure implied by the COMTRADE data to the values coming from WIOD in order to infer missing values and zeros in the underlying SITC-level expenditure data. On average, the two data sets match at the WIOD aggregate level. However, there are some instances where WIOD aggregates are larger than WIOD aggregates implied by COMTRADE, and some instances where they are smaller. In the former case, we infer that there are true missing values in the COMTRADE data, while in the later case we infer that

[31]There exist three small negative values in this dataset, which all are instances of self trade for certain sectors and are negligible share of total self-trade. We assume that output is incorrect and replace these value with zero (effectively increasing output in that WIOD aggregate and country).

the WIOD aggregates have missing underlying values and the missing values in COMTRADE are actually zeros.

We adjust the data as follows. Conditional on having a zero in the corresponding WIOD aggregate, 20.6 percent of SITC observations have a value in COMTRADE. The remaining we infer to be true zeros rather than missing observations, so whenever the WIOD aggregate is zero and a SITC value is missing, we set the SITC value to zero. Otherwise, we assume that the WIOD data is incorrect and use the information in the COMTRADE data to fill in the zeros in the WIOD. For observations where WIOD aggregates are positive, we infer zeros and missing values in COMTRADE as follows. First, if the WIOD aggregate value implied by COMTRADE is missing but the WIOD aggregate is positive, we treat all the underlying SITC observations from COMTRADE as missing. Second, if the WIOD aggregate is less than the WIOD aggregate implied by COMTRADE, we infer that the WIOD data is incorrect, replace its value with the value implied by COMTRADE, and treat all the SITC missing values underlying the aggregate as zeros. Finally, if the WIOD aggregate is greater than the WIOD aggregate implied by COMTRADE, we infer that the discrepancy is due to missing values in COMTRADE. As such, we leave all missing SITC-level observations underlying the WIOD aggregate as true missing values. The resulting dataset has 23.3 percent inferred missing SITC values and 25.4 percent inferred zeros, and its WIOD aggregates are always greater than or equal to the aggregate of the underlying SITC expenditure data. We observe no self-trade data in COMTRADE, so conditional on self trade, all SITC values are missing. Among missing values, 13.9 percent are self trade observations.

D.3 Tariff Interpolation

Although our estimation can handle missing expenditure values at the SITC-level, it requires a full sample of tariff observations. We use the tariff measure in COMTRADE which is the minimum of tariffs across underlying products. 49.1 percent of these tariff values are missing including missing values associated with self-trade observations (which make up 3.2 percent of the data). Among those that are missing, 47.2 percent also have a missing value for expenditure, indicating that about half of the missing tariff data comes from no COMTRADE observation. Among observations with a non-missing value for expenditure, 33.8 percent of tariffs are missing. We interpolate SITC tariff data as follows. First, we use the minimum within each 4-digit SITC code (across origins within a destination-year) to fill in missing values, which leaves 18.5 percent of observations missing. Second, we interpolate using the minimum within each 3-digit SITC code (leaving 1.3 percent missing), the minimum within each 2-digit SITC code (leaving 0.33 percent missing), and, finally, the minimum within each 1-digit SITC code (leaving no missing values). Finally, we set self-trade tariffs to zero.

D.4 WIOD Aggregate-Level Tariffs

To estimate the independent sector model, we require WIOD sector-level tariff data. We aggregate the COMTRADE tariff data to the WIOD aggregate sector level as follows. We use our model-based aggregation procedure to compute the aggregate applied tariff and total trade value in the COMTRADE data by SITC code, exporter, importer, and year. In particular, the model implies that when factors correspond to WIOD sectors, the within-WIOD-sector factor weights correspond to global expenditure shares. Then, up to a first order approximation around zero tariffs, factor-level tariff indices, which under SGM correspond to WIOD sectors, are equal to a weighted average of underlying 4-digit SITC tariffs using these global expenditure shares as weights. We use these global expenditure weighted tariff averages for WIOD sector-level tariffs.

E Latent Factor Model Estimation: Algorithm

We do not observe all sectors in (25). Additionally, we need to account for observed tariffs, and simultaneously estimate σ_k for $k = 1, \ldots, K$. The presence of missing data requires to use an adjusted version of (26), which we describe in Section E.1. We solve this adjusted problem using an extension of the multiplicative-update non-negative matrix factorization (NMF) algorithm of Lee and Seung (1999, 2001) to accommodate covariates and missing data, which we present in Section E.2.

E.1 Accounting for Missing Data

The WIOD expenditure data occasionally have more expenditure than the total expenditure across SITC 4-digit sectors within that WIOD aggregate. To model expenditure coming from sources other than those in the SITC 4-digit data, we include a synthetic sector within each SITC 4-digit aggregate. When the SITC 4-digit data match the WIOD data, there is no expenditure on this synthetic sector. We then have 773 4-digit sectors plus 14 WIOD synthetic sectors, where the former may be missing, and the latter are always observed. In the following notation we do not differentiate between these sectors, so that $S = 773 + 14$.

Appending a t subscript to denote year, let $\mathcal{S}_{jodt}$ be the set of observed sectors for origin o delivering to destination d at time t in WIOD aggregate j. We use data from WIOD to construct residual expenditure on unobserved sectors, which is

$$R_{jodt} = \sum_{s \in \mathcal{S} \setminus \mathcal{S}_{jodt}} \sum_{k=1}^{K} t_{sod}^{-\sigma_k} \lambda_{sk} \phi_{kod},$$

where $\mathcal{S} = \{1, \ldots, S\}$.

Since the sum of Poisson variables is also Poisson with scale equal to the sum of underlying scale parameters, we can write the objective function in terms of an observed component and residual component,

$$\mathcal{L} = \sum_{jodt} \left[\sum_{s \in \mathcal{S}_{jodt}} \ell \left(\frac{X_{sod}}{X_{od}}, \sum_{k=1}^{K} t_{sod}^{-\sigma_k} \lambda_{sk} \phi_{kod} \right) + \ell \left(R_{jodt}, \sum_{s \in \mathcal{S} \setminus \mathcal{S}_{jodt}} \sum_{k=1}^{K} t_{sod}^{-\sigma_k} \lambda_{sk} \phi_{kod} \right) \right].$$

The algorithm in the following section provides a method to minimize this function.

E.2 NMF with Covariates and Missing Data

The extensions in the multiplicative-update non-negative matrix factorization (NMF) algorithm of Lee and Seung (1999, 2001) do not change the properties of the algorithm.

The data are (X_{it}, Z_{it}) where $i = 1, \ldots, N$ is a (potential) unit of observation, while $t = 1, \ldots, T$ indexes cross sections. We assume that $X_{it} \mid Z_{it}$ is a Poisson random variable with scale

$$\hat{X}_{it} = \sum_{k=1}^{K} Z_{it}^{-\sigma_k} \lambda_{ik} \phi_{kt},$$

for some unknown parameters $\{\sigma_k, \Lambda_k, \Phi_k\}_{k=1}^{K}$, with $\Lambda_k \equiv (\lambda_{1k}, \ldots, \lambda_{Nk})'$ and $\Phi_k \equiv (\phi_{1k}, \ldots, \phi_{Tk})'$. We assume that all values of Z_{it} are observed, but for each t there are some (but not all) values of X_{it} that are unobserved. However, we also observe some aggregates that are representative of each full cross section. For each i, there is a $j(i)$ such that in every t we observe

$$\bar{X}_{jt} \equiv \sum_{i=1}^{N} \mathbf{1}\{j(i) = j\} X_{it}.$$

Although we do not observe all the data at the i-level, we indirectly observe them via these aggregates.

Let $\mathcal{I}_t$ denote the observations in cross-section t, and define the component of each aggregate that is attributable to missing data—the residual component of the aggregate—as

$$R_{jt} \equiv \bar{X}_{jt} - \sum_{i \in \mathcal{I}_t} \mathbf{1}\{j(i) = j\} X_{it} = \sum_{i \notin \mathcal{I}_t} \mathbf{1}\{j(i) = j\} X_{it}.$$

Since the sum of Poisson random variables is Poisson with scale equal to the sum of the underlying scales, we have that $R_{jt} \mid \hat{X}_{1t}, \ldots, \hat{X}_{Nt}$ is Poisson with scale $\hat{R}_{it} = \sum_{i \notin \mathcal{I}_t} \mathbf{1}\{j(i) = j\} \hat{X}_{it}$.

In this setup, each $\hat{X}_{it}$ contributes to explaining the observed data through a unique observation—either because X_{it} is observed directly, or because it is unobserved and shows up in the residual of a unique j. Define the group of potential observations that i is aggregated with as $\mathcal{I}_{it} = \{i\}$ if $i \in \mathcal{I}_t$ and $\mathcal{I}_{it} = \{i' \in \mathcal{I}_t \mid j(i') = j(i)\}$ if $i \notin \mathcal{I}_t$. Then, define

$$Y_{it} \equiv \sum_{i' \in \mathcal{I}_{it}} X_{i't} = \begin{cases} X_{it} & \text{if } i \in \mathcal{I}_t \\ R_{j(i)t} & \text{if } i \notin \mathcal{I}_t \end{cases} \quad \text{and} \quad \hat{Y}_{it} \equiv \sum_{i' \in \mathcal{I}_{it}} \hat{X}_{i't}.$$

It is useful to define the "filled in" $N \times T$ data matrix, $\mathbf{Y}$, with entries $[\mathbf{Y}]_{it} = Y_{it}$ and a prediction matrix $\hat{\mathbf{Y}}$ with entries $[\hat{\mathbf{Y}}]_{it} = \hat{Y}_{it}$. When there is no missing data, this prediction matrix can be written as

$$\hat{\mathbf{Y}} = \sum_{k=1}^{K} \mathbf{Z}^{-\sigma_k} \odot (\Lambda_k \Phi_k'),$$

where $\mathbf{Z}$ is the matrix of explanatory variables, $[\mathbf{Z}]_{it} = Z_{it}$. In the case without explanatory variables, set $\sigma_k = 0$ for all k), and get

$$\mathbb{E}[\mathbf{Y}] = \hat{\mathbf{Y}} = [\Lambda_1 \dots \Lambda_k][\Phi_1 \dots \Phi_k]'.$$

That is, we have a matrix-factorization problem. Because all the data and parameters are non-negative, it is a non-negative matrix factorization problem. The present model generalizes this problem to incorporate missing data and explanatory variables with factor-specific coefficients.

The Poisson deviance is

$$\mathcal{L} = \sum_{t=1}^{T} \left[\sum_{i \in \mathcal{I}_t} \ell(X_{it}, \hat{X}_{it}) + \sum_{j=1}^{J} \ell\left(R_{jt}, \sum_{i \notin \mathcal{I}_t} \mathbf{1}\{j(i) = j\} \hat{X}_{it} \right) \right].$$

It is useful to re-write this expression as

$$\mathcal{L} = \sum_{t=1}^{T} \left[\sum_{i \in \mathcal{I}_t} \ell(X_{it}, \hat{X}_{it}) + \sum_{i \notin \mathcal{I}_t} \frac{\ell\left(R_{j(i)t}, \sum_{i' \notin \mathcal{I}_t} \mathbf{1}\{j(i') = j\} \hat{X}_{i't} \right)}{\sum_{i' \notin \mathcal{I}_t} \mathbf{1}\{j(i') = j\}} \right].$$

But then

$$\mathcal{L} = \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{\ell(Y_{it}, \hat{Y}_{it})}{N_{it}}, \tag{E.1}$$

where $N_{it} = 1$ if $i \in \mathcal{I}_t$ and $N_{it} = \sum_{i'=1}^{N} \mathbf{1}\{j(i') = j(i)\}$ if $i \notin \mathcal{I}_t$. Recall that $\ell(x, \hat{x}) \equiv 2(x \ln(x/\hat{x}) - (x - \hat{x})) = 2(\hat{x} - x \ln \hat{x} + x \ln x - x)$ so that $\partial \ell(x, \hat{x})/\partial \hat{x} = 2(1 - x/\hat{x})$.

The derivative in $\lambda_{i'k}$ is then

$$\frac{\partial \mathcal{L}}{\partial \lambda_{i'k}} = 2\sum_{i=1}^{N}\sum_{t=1}^{T}\left(1-\frac{Y_{it}}{\hat{Y}_{it}}\right)\frac{\mathbf{1}\{i'\in\mathcal{I}_{it}\}Z_{i't}^{-\sigma_k}\phi_{kt}}{N_{it}} = 2\sum_{t=1}^{T}\left(1-\frac{Y_{it}}{\hat{Y}_{it}}\right)Z_{i't}^{-\sigma_k}\phi_{kt}.$$

We can therefore write the gradient in Λ_k as

$$\frac{\partial \mathcal{L}}{\partial \Lambda_k} = 2\mathbf{Z}^{-\sigma_k}\Phi_k - 2\left(\frac{\mathbf{Y}}{\hat{\mathbf{Y}}}\odot\mathbf{Z}^{-\sigma_k}\right)\Phi_k,$$

where $[\mathbf{Z}]_{it} = Z_{it}$ and $\odot$ denotes element-wise multiplication. The update multiplies the existing value of Λ_k by the ratio of the negative component of the gradient to the positive component,

$$\Lambda_k \leftarrow \Lambda_k \odot \frac{\left(\frac{\mathbf{Y}}{\hat{\mathbf{Y}}}\odot\mathbf{Z}^{-\sigma_k}\right)\Phi_k}{(\mathbf{Z}^{-\sigma_k})\Phi_k}. \tag{E.2}$$

Larger entries of Λ_k increase predicted values. When the current prediction is below the observed value, this update increases Λ_k, thereby increasing the predicted values. Any time we update Λ_k, we follow up by performing $\Phi_k \leftarrow \Phi_k(\mathbf{1}'\Lambda_k)$, and $\Lambda_k \leftarrow \Lambda_k/(\mathbf{1}'\Lambda_k)$, where $\mathbf{1}$ denotes a vector of ones. This update has no effect on predictions and forces the normalization $\sum_{i=1}^{N}\lambda_{ik} = 1$.

Similarly, we get an updating rule for Φ_k given by

$$\Phi_k \leftarrow \Phi_k \odot \frac{\left(\frac{\mathbf{Y}}{\hat{\mathbf{Y}}}\odot\mathbf{Z}^{-\sigma_k}\right)'\Lambda_k}{(\mathbf{Z}^{-\sigma_k})'\Lambda_k}. \tag{E.3}$$

Finally, the derivative in σ_k is

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \sigma_k} &= -2\sum_{i=1}^{N}\sum_{t=1}^{T}\left(1-\frac{Y_{it}}{\hat{Y}_{it}}\right)\sum_{i'\in\mathcal{I}_{it}}\frac{Z_{i't}^{-\sigma_k}\lambda_{i'k}\phi_{kt}\ln Z_{i't}}{N_{it}} \\
&= -2\sum_{i=1}^{N}\sum_{t=1}^{T}\left(1-\frac{Y_{it}}{\hat{Y}_{it}}\right)Z_{it}^{-\sigma_k}\lambda_{ik}\phi_{kt}\ln Z_{it} \\
&= -2\mathbf{1}'\left[\mathbf{Z}^{-\sigma_k}\odot(\Lambda_k\Phi_k')\odot\ln\mathbf{Z} - \frac{\mathbf{Y}}{\hat{\mathbf{Y}}}\odot\mathbf{Z}^{-\sigma_k}\odot(\Lambda_k\Phi_k')\odot\ln\mathbf{Z}\right]\mathbf{1}.
\end{aligned}$$

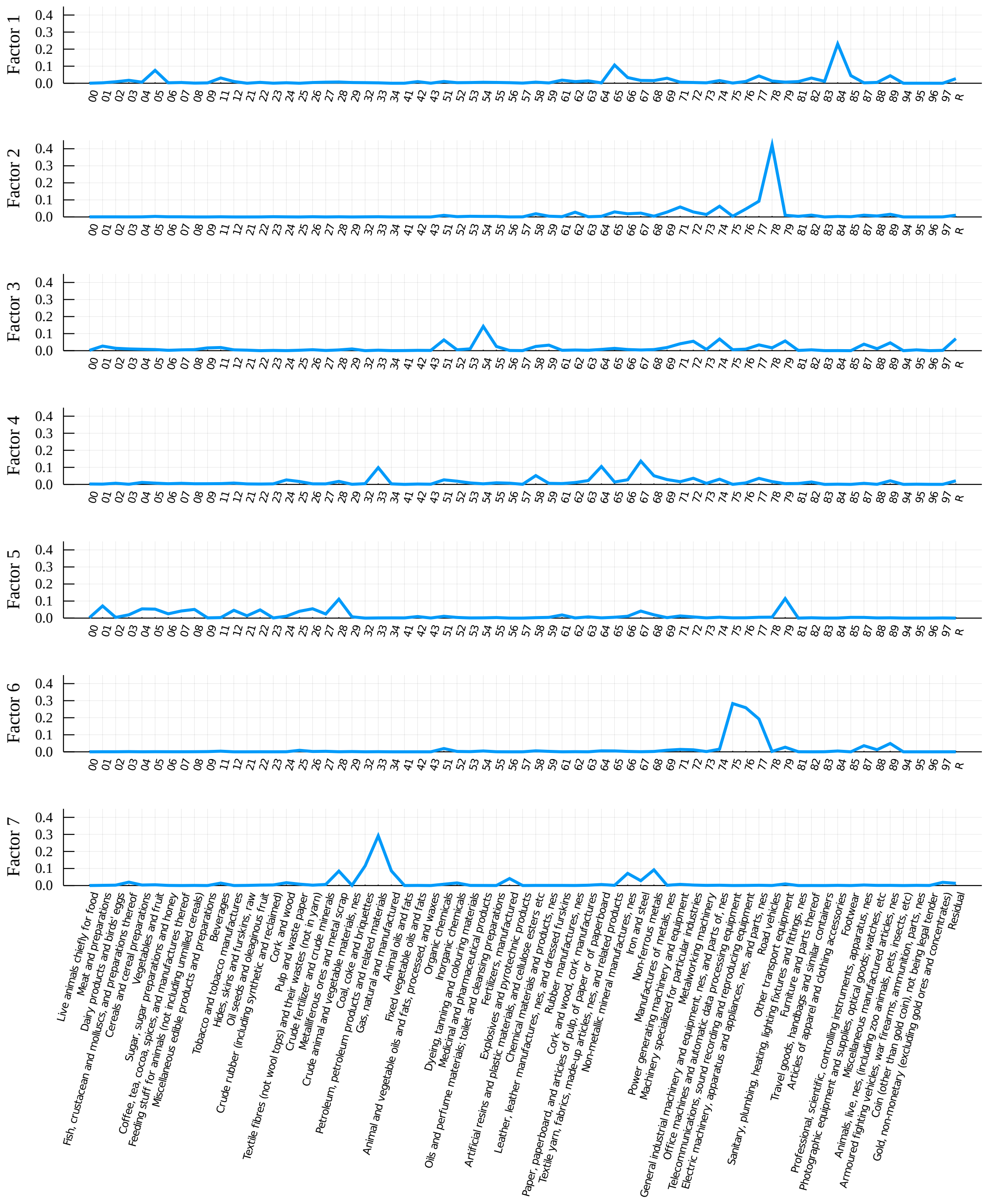
The implied updating rule is

$$\sigma_k \leftarrow \sigma_k \odot \frac{\mathbf{1}'[\mathbf{Z}^{-\sigma_k} \odot (\Lambda_k \Phi_k') \odot \ln \mathbf{Z}]\mathbf{1}}{\mathbf{1}'\left[\frac{\mathbf{Y}}{\hat{\mathbf{Y}}} \odot \mathbf{Z}^{-\sigma_k} \odot (\Lambda_k \Phi_k') \odot \ln \mathbf{Z}\right]\mathbf{1}}. \tag{E.4}$$

Using the proof technique in Lee and Seung (2001), one can show that (E.1) is monotonically decreasing in any of (E.2), (E.3), and (E.4). To estimate the model, we sequentially iterate on these updating rules until convergence. With no guarantee of finding the global optimum, we repeat the algorithm from many random starting values and use the version with the lowest value of (E.1) as our estimate.

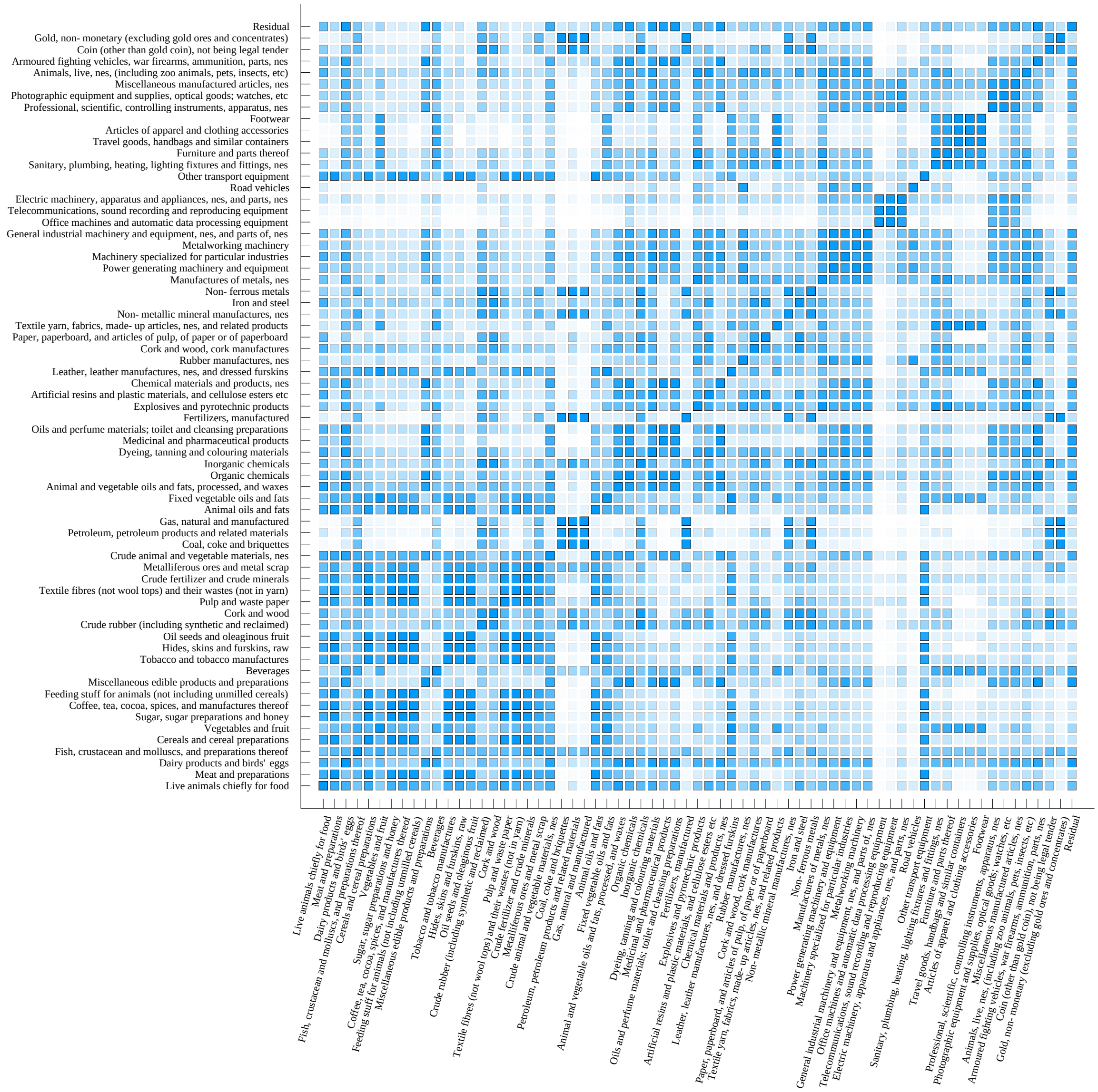
F Additional Quantitative Results

Figure F.1: Factor Weights: 2-Digit Sectors.



Notes: Estimates of factor weights across 4-digit sectors, aggregated to 2-digit sectoral level.

Figure F.2: Similarity of Factor Use Across 2-Digit Sectors.



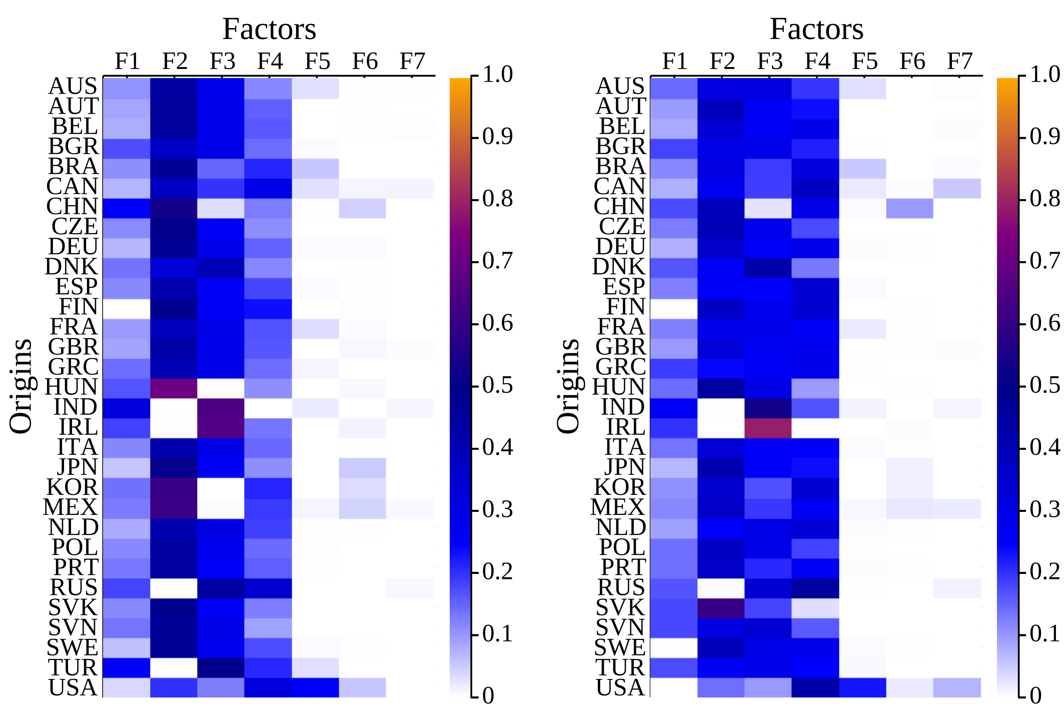
Notes: Similarity is calculated as cosine similarity, $\sum_i x_i y_i / \sqrt{(\sum_i x_i^2 \sum_i y_i^2)}$.

Table F.1: Similarity of Factor Weights.

	Pairs of Factors		Pairs of 4-Digit Sectors	
	Fraction of 4-Digit Sectors Shared	Similarity	Fraction of Factors Shared	Similarity
Mean	0.74	0.05	0.746	0.374
Standard Deviation	0.056	0.035	0.19	0.301
Minimum	0.649	0.001	0.0	0.0
10th Percentile	0.667	0.011	0.429	0.026
Median	0.745	0.046	0.714	0.302
90th Percentile	0.799	0.108	1.0	0.848
Maximum	0.842	0.112	1.0	1.0

Notes: For each column, we computed the measure in the second row for the unit of observation in the first row, then report the moments of the distribution of the measure across that unit of observation. Sectors are 4-digit SITC sectors. Similarity is calculated as cosine similarity, $\sum_i x_i y_i / \sqrt{(\sum_i x_i^2 \sum_i y_i^2)}$.

Figure F.3: Correlation Function Weights.



a: USA, 1999 b: USA, 2007

Notes: Each figure shows ω_{kod} across k and o for $d = USA$. Each row sums to one.