American Economic Review 2023, 113(2): 317–353
https://doi.org/10.1257/aer.20190781

Trade with Correlation†

By Nelson Lind and Natalia Ramondo*

We develop a trade model with correlation in productivity across countries. The model spans the full class of generalized extreme value import demand systems and implies that countries with relatively dissimilar technology gain more from trade. In the context of a multisector trade model, we provide a tractable and flexible estimation procedure for correlation based on compressing highly disaggregate sectoral data into a few latent factors related to technology classes. We estimate significant heterogeneity in correlation across sectors and countries, which leads to quantitative predictions that are significantly different from estimates of models assuming independent productivity across sectors or countries. (*JEL* C38, F11, F13, F14, L16, O30)

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.

The recent quantitative trade literature, building on Eaton and Kortum (2002; henceforth, EK), incorporates Ricardian motives for trade by treating productivity as a random draw across goods, countries, and other observable economic units—such as sectors. In these models, the joint distribution of productivity determines the gains from trade. However, this literature relies on independence assumptions, which, although leading to convenient functional forms for estimation, restrict empirically relevant expenditure substitution patterns and impact inference on the gains from trade.

In this paper, we develop a Ricardian model that allows for rich patterns of correlation in productivity. By relaxing the independence assumptions used in the literature, the model generates import demand systems spanning the entire generalized extreme value (GEV) class (McFadden 1978, 1981). Our approach sheds light on the properties and limitations of existing models, provides tools to build new models, and enables the development of a flexible estimation procedure for correlation.

*Lind: Department of Economics, Emory University (email: nlind@emory.edu); Ramondo: Department of Economics, Boston University and NBER (email: nramondo@bu.edu). Arnaud Costinot was the coeditor for this article. This research is supported by NSF grant 1919372. We thank our discussant Costas Arkolakis. 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. All errors are our own. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

†Go to https://doi.org/10.1257/aer.20190781 to visit the article page for additional materials and author disclosure statements.

Our quantitative application proposes a cross-nested constant-elasticity-of-substitution (CES) structure for productivity where the novelty comes from treating each nest as an unobserved—or, latent—dimension of the data. We apply this structure to a multisector Ricardian trade model where these latent nests have an intuitive interpretation as technology classes. These technologies can be used to produce goods classified in different sectors, allowing sectors to share production methods. Multisector models in the literature typically preclude sectors from sharing technologies, instead pairing each latent nest with an observed sectoral category. This assumption, which implies independent productivity across sectors, is particularly problematic because observed sectoral classifications may not correspond to technology classes, preventing the model from capturing cross-sector substitution patterns, which may matter for counterfactual analysis.

Our estimation procedure uses disaggregate sectoral data to uncover the latent nests of the correlation structure, and reveals significant sharing of technologies across countries and sectors. This sharing manifests in considerable heterogeneity in correlation in productivity, which, in turn, changes the answers to standard counterfactuals.

We start in Section I by presenting a Ricardian trade model with a general dependence structure for productivity, which preserves the max-stability property of Fréchet distributions crucial for tractability in EK.[1] Our model reduces to EK in the special case of independent productivity, a case that implies that bilateral trade flows follow a CES structure. In Section II, we show that correlation in productivity implies expenditure shares that belong to the GEV class. This class admits rich substitution patterns and includes—but it is not restricted to—models in the EK tradition, such as models with many sectors (Costinot, Donaldson, and Komunjer 2012; Costinot and Rodríguez-Clare 2014; Levchenko and Zhang 2014; Caliendo and Parro 2015; French 2016; Lashkaripour and Lugovskyy 2017), multinational production (Ramondo and Rodríguez-Clare 2013; Alviarez 2019), global value chains (Antràs and de Gortari 2020), and domestic geography (Ramondo, Rodríguez-Clare, and Saborío-Rodríguez 2016; Redding 2016), among others.

Despite its generality, our theory leads to intuitive and tractable counterfactual analysis. We can calculate the gains from trade for the GEV class by adjusting the CES case in Arkolakis, Costinot, and Rodríguez-Clare (2012; henceforth, ACR) to account for correlation in technology with the rest of the world: Countries with a similar degree of openness to the rest of the world but more dissimilar technologies will enjoy higher gains from trade. Although the sufficient-statistic approach for the gains from trade in ACR only depends on one parameter (i.e., the shape parameter of the Fréchet distribution), this result hinges on the assumption of independent productivity across countries. Once we abandon this assumption, the sufficient statistic approach requires additional parameters that capture correlation in productivity—and they will need to be estimated.

Section III presents the quantitative application of our framework to a multisector trade model. We use a cross-nested CES structure where productivity is independent

[1] A distribution is max-stable if it is preserved under the max operator (up to location and scale). Thanks to the max-stability property, the probability of a source being the lowest-cost producer in a given destination is equal to that source expenditure share in the destination.

across nests, and each nest is treated as a latent (unobserved) factor.[2] This latent factor model (LFM) allows factors to be shared across sectors and countries, inducing correlation in productivity. When we restrict each factor to be unique to a sector, productivity is independent across sectors, and gravity, as defined by ACR, holds at the sectoral level.

A key advantage of the LFM is that it relaxes the strong restrictions on expenditure substitution patterns that exist in many sectoral models. In particular, we can depart from a gravity structure for expenditure at the sector level—a structure that entails independence of irrelevant alternatives (IIA) within and across sectors. Concretely, this means that we allow for sectoral elasticities that can vary across countries and for cross-sector elasticities that are different from zero.

A second important advantage of LFM is that, by decoupling latent factors from sectors in the data, it avoids imposing that a sector, as defined by some arbitrary choice of aggregation available in the data, corresponds to some fundamental aspect of the production process. Instead, we treat latent factors as fundamental aspects of the production process (e.g., a technology class); sectors are categories designed to apply policies, such as import tariffs.[3]

Our estimation procedure regroups observed sectors into fewer latent factors. Similar to principal-component analysis, it compresses the data from a high to a lower dimension. To perform this compression, we use disaggregate sectoral expenditure and tariff data, and adapt techniques from the literature on nonnegative matrix factorization (Lee and Seung 1999, 2000; Fu et al. 2019) coupled with a pseudo Poisson maximum likelihood criterion for estimation (Santos Silva and Tenreyro 2006; Fally 2015).[4] The identification strategy underlying this procedure comes from a separability assumption—sectors use latent factors in the same proportions everywhere and sectoral comparative advantage arises from an exporter's ability to use each latent factor.

Our LFM estimates show that seven latent factors are enough to explain almost 95 percent of the variation in four-digit Standard International Trade Classification (SITC) bilateral trade flows. These factors are broadly shared across sectors, but they are also used intensively for the production of certain goods. Factors related to the production of simple manufactured goods are highly correlated across countries, while factors associated with the production of complex manufactured goods, such as electronics, and with natural-resource extraction present low correlation across countries.

The substitution expenditure elasticities generated by our LFM estimator differ significantly from those implied by gravity estimates. The difference comes from the zero cross-sector elasticities imposed by the gravity model. In particular, we estimate cross-price elasticities that are much more heterogeneous. These estimates, in turn, shape the amount of correlation across countries and sectors predicted by each model.

[2] The reference to factors is taken in analogy to the macro and finance literatures using principal-component analysis.

[3] This manifests in the phenomenon known as tariff engineering. Firms make marginal adjustments to their product to change their tariff classification to one with a lower duty. In turn, governments use reclassifications to pursue trade-policy goals (e.g., Costa Tavares 2006). Tariff engineering has led to many legal cases (e.g., *The United States v. Citroen*, *The United States v. Heartland By-Products*), starting with the case of US import duties to sugar in 1881 (see Irwin 2017), as well as a flurry of articles in the business press (e.g., Frankel 2018).

[4] Principal-component analysis cannot be applied in our case because, typically, it delivers negative estimates. Expenditure, however, is nonnegative.

By estimating significant heterogeneity in correlation across sectors and countries, we show in Section IV that our quantitative model predicts that, among countries equally open to the rest of the world, the ones with relatively dissimilar technology to their partners gain more from trade. For instance, Canada, which is similar in its degree of openness to Germany, has gains from trade that are almost 90 percent higher. Our LFM estimates reveal that Canada is a top exporter of low-correlation factors, while Germany specializes in factors with high correlation in productivity across countries. The difference in gains between these two countries is only 4 percent if we use estimates from models that assume independent productivity across sectors. In contrast, we find that, conditional on a country's openness, variation in estimated gains is much higher for LFM.

Related Literature.—Our paper naturally relates to the large trade literature using the Ricardian-EK framework; see Eaton and Kortum (2012) for a review. 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. 2019; Bryan and Morten 2019), or trade models used in the urban literature (Ahlfeldt et al. 2015; Monte, Redding, and Rossi-Hansberg 2015).

Our paper is closely related to Adao, Costinot, and Donaldson (2017). Their framework includes the class of GEV import demand systems that we use in our paper and they provide sufficient conditions for nonparametric identification of the large class of invertible import demand systems. 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 model of Ricardian comparative advantage where IIA does not need to hold and leads to closed-form expressions. Our LFM estimation procedure, based on latent factors and disaggregate data, presents a flexible alternative to the Berry, Levinsohn, and Pakes (1995) procedure in Adao, Costinot, and Donaldson (2017), but shares the same broad goal of flexibly allowing for departures from IIA.

Relatedly, papers such as Caron, Thibault, and Markusen (2014); Lashkari and Mestieri (2016); Brooks and Pujolas (2019); Feenstra et al. (2017); and Bas, Mayer, and Thoenig (2017), among others, estimate import demand systems with more flexible substitution patterns than CES. They abandon homothetic demand systems, which we do not, but aim, as we do, to incorporate disaggregate data to estimate elasticities. In contrast with this literature, we link expenditure substitution patterns to the degree of technological similarity across countries and sectors. In this way, we can incorporate heterogeneity in elasticities without relying on demand-side factors.

Finally, the quantitative trade literature typically incorporates an amplification mechanism for trade through input-output networks (e.g., Caliendo and Parro 2015). While our model relaxes the independence assumptions of multisector EK models, it does not incorporate an input-output structure. Even though correlation in productivity is a distinct economic mechanism from input-output linkages, it could lead to similar quantitative predictions. A comparison between the gains from trade implied by the LFM and estimates of the sectoral gravity model augmented by those linkages reveals that LFM gains are not only higher but also much more heterogeneous, suggesting that the correlation structure of this model captures economic forces that are distinct from input-output linkages.

I. The Ricardian Model of Trade

Consider a global economy consisting of N countries. We use the subscript o for origin countries and d for destination countries. Countries produce and trade a continuum of goods, $v \in [0,1]$. Consumers have identical CES preferences over goods with elasticity of substitution $\eta > 1$. Expenditure on v is $X_d(v) = \left(\frac{P_d(v)}{P_d}\right)^{1-\eta} X_d$, where $P_d(v)$ is the price of good v, $P_d = \left(\int_0^1 P_d(v)^{1-\eta} dv\right)^{\frac{1}{1-\eta}}$ is the price level, and X_d is total expenditure in country d.

Each good v is produced with an only-labor constant returns to scale technology,

$$Y_{od}(v) = Z_{od}(v) L_{od}(v).$$

Productivity $Z_{od}(v)$ 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 the efficiency of production in the origin and inefficiencies associated with delivery to the destination. In this way, we do not impose the standard assumption on iceberg trade costs (Samuelson 1954), which would correspond to the special case of $Z_{od}(v) = Z_o(v)/\tau_{od}$. As in EK, we model productivity as a random variable drawn from a max-stable multivariate Fréchet distribution. By applying the tools developed originally for random utility models (McFadden 1978, 1981) to Ricardian models of trade, we get a flexible, yet tractable, model of trade in the GEV class. Models in this class capture Ricardo's insight that the degree of technological similarity determines the gains from trade.

A. *Max-Stable Multivariate Fréchet Productivity*

We assume that the joint distribution of productivity across origin countries is given by

$$\text{(1)} \quad \Pr\left[Z_{1d}(v) \leq z_1, \ldots, Z_{Nd}(v) \leq z_N\right] = \exp\left[-G^d\left(T_{1d} z_1^{-\theta}, \ldots, T_{Nd} z_N^{-\theta}\right)\right],$$

where $T_{od} > 0$ is the scale parameter and $\theta > 0$ the shape parameter characterizing the marginal (Fréchet) distributions, $\Pr\left[Z_{od}(v) \leq z\right] = e^{-T_{od} z^{-\theta}}$. The scale parameters capture the absolute advantage of countries, while the shape parameter regulates the heterogeneity of independent and identically distributed productivity draws across the continuum of goods, as in all models based on EK.

The function G^d is a *correlation function*, also called tail dependence function in probability and statistics. This function allows for a flexible dependence structure across origin countries o serving destination d.[5] It is homogeneous of degree one, ensuring the property of max-stability. Max-stability implies that the

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

distribution of the maximum is Fréchet with shape θ, and that the conditional and unconditional distributions of the maximum are identical. As for EK, this result is crucial for tractability because it implies that expenditure shares equal the probability of a destination importing from a given origin country. Additionally, a correlation function presents the regularity properties of the social surplus function in GEV discrete choice models (McFadden 1981; Train 2009): unboundedness, and a sign pattern for cross-partial derivatives (reflected in expenditure satisfying the gross-substitute property).[6] Finally, we impose a normalization restriction, $G(0,\ldots,0,1,0,\ldots,0) = 1$, so that the scales, which parameterize the marginal distributions, are separated from the correlation function, which determines the joint distribution of productivity.

To fix ideas, assume that productivity is independent across countries, as in EK:

$$(2) \qquad \Pr\left[Z_{1d}(v) \leq z_1,\ldots,Z_{Nd}(v) \leq z_N\right] = \prod_{o=1,\ldots,N} \Pr\left[Z_{od}(v) \leq z_o\right] = \exp\left(-\sum_{o=1}^{N} T_{od} z_o^{-\theta}\right).$$

This case corresponds to (1) with an additive correlation function,

$$(3) \qquad G^d(x_1,\ldots,x_N) = \sum_{o=1}^{N} x_o.$$

Because of the additive structure of G^d, in this special case, the shape parameter θ plays two distinct roles. First, since it controls dispersion in productivity across the continuum of goods, it determines the distribution of relative productivity between any two goods within a country, $Z_{od}(v)/Z_{od}(v')$. Second, because of independence, it also controls the joint distribution of relative productivity between any two countries, and therefore the strength of comparative advantage—determined by $\frac{Z_{od}(v)/Z_{od}(v')}{Z_{o'd}(v)/Z_{o'd}(v')}$. Consequently, this case leads to the result in ACR that θ alone governs the gains from trade in the EK model.

However, once we abandon the assumption of independent productivity, the strength of comparative advantage no longer solely depends on θ. Generally, it is the correlation function that controls comparative advantage since it determines the joint distribution of productivity between any two countries. Except for the knife-edge case of independence, *both* θ and the parameters defining G^d will matter for the gains from trade.

To illustrate this point, consider the case of a symmetric max-stable Fréchet distribution,

$$(4) \qquad \Pr\left[Z_{1d}(v) \leq z_1,\ldots,Z_{Nd}(v) \leq z_N\right] = \exp\left[-\left(\sum_{o=1}^{N}\left(T_{od} z_o^{-\theta}\right)^{\frac{1}{1-\rho}}\right)^{1-\rho}\right],$$

[6] Formally: $G^d(x_1,\ldots,x_N) \to \infty$ as $x_o \to \infty$ for any $o = 1,\ldots,N$; and the mixed partial derivatives of G^d exist and are continuous up to order N, with the oth partial derivative with respect to o distinct arguments nonnegative if o is odd and nonpositive if o is even.

where the correlation function is CES,

$$G^d(x_1,\ldots,x_N) = \left[\sum_{o=1,\ldots,N} x_o^{\frac{1}{1-\rho}}\right]^{1-\rho}. \tag{5}$$

The parameter $\rho \in [0,1)$ regulates correlation in productivity draws across origins o, and therefore similarity in relative productivity. When $\rho = 0$, we are back to the independence case. As $\rho \to 1$, relative productivity between any two goods becomes identical across countries. In this case, no country has a comparative advantage in any good, and, as we will see, there are no gains from trade. Despite the existence of heterogeneity in productivity across goods, regulated by θ, it is now ρ that determines the strength of comparative advantage across countries.[7]

Next, we focus on cross-nested CES correlation functions. This functional form constitutes the foundation of our procedure to estimate correlation patterns across countries.

B. *The Case of Cross-Nested CES*

We now present a flexible structure for correlation based on a cross-nested CES (CNCES) function. This case is relevant for several reasons. First, it approximates any correlation function. Second, it is the building block of many EK-type Ricardian models of trade, such as sectoral models. And third, it allows us to relax commonly made distributional assumptions through the introduction of latent nests.

Assume that productivity is distributed max-stable multivariate Fréchet, with scale T_{od}, shape θ, and the following correlation function:

$$G^d(x_1,\ldots,x_N) = \sum_{k=1}^{K}\left[\sum_{o=1}^{N}(\omega_{kod}x_o)^{\frac{1}{1-\rho_k}}\right]^{1-\rho_k}, \tag{6}$$

where $\rho_k \in [0,1)$, for each k, $\omega_{kod} > 0$, and $\sum_k \omega_{kod} = 1$. The weight ω_{kod} indicates the relative importance of each nest k for a given trading pair od. If ω_{kod} is high, nest k is particularly productive in country o for delivery to d. Within nest k, correlation in productivity across origins is measured by the correlation coefficient ρ_k. For $\rho_k = 0$, productivity is independent and the kth nest is additive. In contrast, as $\rho_k \to 1$, productivity becomes perfectly correlated within nest k, and the kth nest converges to a max function.

The specification in (6) is particularly useful as it can capture any max-stable structure.

PROPOSITION 1 (CNCES Approximation): *Any correlation function can be approximated uniformly on compact sets using a CNCES correlation function.*

[7] However, in this case, the parameter θ and ρ cannot be separately identified from the trade data because this correlation function leads to CES expenditure with an elasticity of $\theta/(1-\rho)$ (see also Eaton and Kortum 2002).

PROOF:
See Appendix A.

This result ensures that focusing on CNCES—as we do in our quantitative application in Section III—is without loss of generality.

The CNCES functional form also provides a bridge between our general framework and existing quantitative trade models based on EK—which arise as special cases after imposing additional restrictions on the nests.

As a first example, consider the case where each nest is specific to a single origin, meaning that $\omega_{kod} = \mathbf{1}\{k = o\}$ and $K = N$. The correlation function in (6) collapses to (3), corresponding to independent productivity across countries. Indeed, overlapping nests across countries are necessary for correlation in productivity.

Second, consider the case of only one nest, $K = 1$. This restriction means that (6) collapses to the expression in (5) corresponding to symmetric correlation across origins (i.e., same ρ). For $\rho = 0$, we get the productivity distribution in (2), as in EK. But even for $\rho > 0$, correlation is innocuous because it has no impact on trade patterns—i.e., it leads to a CES import demand system, as we make clear below. We need more than one nest so that correlation is heterogeneous across countries and empirically relevant.

Third, we can connect our aggregate model to disaggregate sectoral models in the literature by assuming that the nests in (6) correspond to sectors. In particular, the additive structure of the nests in the CNCES specification means that we will have closed-form solutions not only for aggregate variables but also, as we show in Section II, nest-level variables. Concretely, letting s index sectors, we can replace k by s in (6). Within each sector, productivity draws can be correlated $(0 \leq \rho_s < 1)$; across sectors, correlation can be heterogeneous $(\rho_s \neq \rho_{s'})$, with higher sectoral correlation due to more similar productivity draws across countries. However, this structure implies that, since each nest is specific to a single sector, productivity draws are independent across sectors and, within sector, correlation is homogeneous across origins. We would need overlapping nests across sectors in order to relax these two assumptions.

An important feature of the nests in the correlation function in (6) is that they do not have to correspond to a category observed in the data, such as a sector. They can be treated as unobserved dimension of the data. In this case, we refer to them as latent factors. In the context of a multisector Ricardian model of trade, we propose to treat the k-nests as unobservable categories, and move away from the assumption of independence across sectors and homogeneous correlation within sectors. When sectors share latent factors, within-factor correlation, captured by ρ_k, induces both across-sector and across-origin correlation. The CNCES structure with latent factors constitutes a tractable and intuitive way of departing from the independence assumptions that are common in the literature.

Furthermore, in the context of Ricardian theory, these latent factors have a natural interpretation as technology classes applied to the production—and delivery—of goods, and may be shared across countries and sectors. Formally, suppose that there exist K technology classes, $k = 1, \ldots, K$, each corresponding to a set of related ideas for producing goods. The efficiency of k in country o to produce good v for destination d is a random variable $Z^*_{kod}(v)$, drawn from a max-stable multivariate

Fréchet distribution with scale T^*_{kod}, shape θ, and correlation function as in (5) with coefficient ρ_k.

Productivity is the result of applying the best set of ideas for production of a good v in a location o for delivery to d: $Z_{od}(v) = \max_{k=1,\ldots,K} Z^*_{kod}(v)$. Due to max-stability, productivity, $Z_{od}(v)$, is distributed max-stable multivariate Fréchet with scale $T_{od} = \sum_k T^*_{kod}$, shape θ, and a CNCES correlation function as in (6) with weights given by $\omega_{kod} \equiv T^*_{kod} / \sum_{k'} T^*_{k'od}$.

This simple example illustrates how the parameters of the correlation function can be linked to primitives related to technology and the nests of a CNCES correlation function can be interpreted as underlying technology classes.[8]

II. Expenditure, Prices, and Welfare

Our theory generates import demand systems belonging to the GEV class. This is a large subclass in the class of invertible demand systems with the gross substitute property, allows for rich patterns of substitution in expenditure, and leads to closed-form expenditure shares.

We first derive expenditure shares for the Ricardian model in Section I. Next, we present properties of the GEV class and focus on the subclass of CNCES. Finally, we characterize macro counterfactuals under GEV.

A. *GEV Import Demand*

Under perfect competition, the price of good v equals its marginal cost, and it is provided to country d by the lowest-cost supplier,

$$P_d(v) = \min_{o=1,\ldots,N} \frac{W_o}{Z_{od}(v)}, \tag{7}$$

with W_o denoting the nominal wage in country o.

The following proposition derives expressions for expenditure shares and the price index. These closed-form results are a direct consequence of max-stability.

PROPOSITION 2 (Trade Shares and Price Levels): *If productivity is distributed max-stable multivariate Fréchet with shape $\theta > \eta - 1$ and a continuously differentiable correlation function, then country d's expenditure share on goods from country o is*

$$\pi_{od} \equiv \frac{X_{od}}{X_d} = \frac{P_{od}^{-\theta} G_o^d\left(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}\right)}{G^d\left(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}\right)}, \tag{8}$$

[8] Lind and Ramondo (2021) develop a model of innovation and diffusion that gives rise to correlation in productivity across countries as a result of the dynamics of knowledge worldwide. Their model provides a microfoundation for the entire class of max-stable multivariate Fréchet distributions.

where $P_{od} \equiv \gamma T_{od}^{-1/\theta} W_o$, $\gamma \equiv \Gamma\left(\frac{\theta + 1 - \eta}{\theta}\right)^{\frac{1}{1-\eta}}$, $\Gamma(\cdot)$ is the gamma function, $G_o^d(x_1, \ldots, x_N) \equiv \partial G^d(x_1, \ldots, x_N)/\partial x_o$, and the price index in country d is given by

$$(9) \qquad P_d = G^d\left(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}\right)^{-\frac{1}{\theta}}.$$

PROOF:

See Appendix B.

First, the share of goods imported from o into d has the same form as choice probabilities in GEV discrete choice models, with $P_{od}^{-\theta}$ replacing choice-specific utility.[9] Second, as in EK, the share of expenditure of country d on goods from o equals the probability that o is the lowest cost producer, thanks to max-stability.[10] Finally, 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.

An important class of import demand systems within the GEV class is CES. An additive correlation function generates CES expenditure,

$$(10) \qquad \pi_{od} = \frac{P_{od}^{-\theta}}{\sum_{o'} P_{o'd}^{-\theta}}.$$

This specification includes most of the workhorse models of trade, such as Armington, Melitz, and EK (ACR). However, the GEV class is much larger than the CES class, allowing for richer substitution patterns.

To clearly see this result, we compute the cross-price elasticity $(o' \neq o)$ of (8):

$$(11) \qquad \varepsilon_{oo'd} \equiv \frac{\partial \ln \pi_{od}}{\partial \ln P_{o'd}/P_d} = -\theta \frac{P_{o'd}^{-\theta} G_{oo'}^d\left(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}\right)}{G_o^d\left(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}\right)} \geq 0,$$

where $G_{oo'}^d(x_1, \ldots, x_N) \equiv \partial G_o^d(x_1, \ldots, x_N)/\partial x_{o'}$. Due to the sign-switching property of the correlation function, these elasticities are nonnegative, implying the gross substitutes property. Further, max-stability implies that the elasticities sum up to $-\theta$, so that the own-price elasticity $(o' = o)$ is simply $\varepsilon_{ood} = -\theta - \sum_{o' \neq o} \varepsilon_{oo'd} < 0$.

When the correlation function is additive, the cross-price elasticity in (11) is zero. That is, CES entails IIA. When the correlation function is not additive, the cross-price elasticity is not zero, generating departures from IIA. Since linearity is associated with independence, more curvature in G^d is associated with more correlation and stronger departures from IIA.

[9] We can map the scale parameters T_{od} into a *productivity index*, $A_o \equiv T_{oo}^{1/\theta}$, which measures a country's ability to produce goods in their domestic market, and an *iceberg trade cost index*, $\tau_{od} \equiv (T_{dd}/T_{od})^{1/\theta}$, which measures efficiency losses associated with delivering goods to market d. In this way, we get the familiar expression $P_{od} = \gamma \tau_{od} W_o / A_o$.

[10] Since the conditional and unconditional distributions of the maximum are identical, $\pi_{od} = E\left[\left(P_d(v)/P_d\right)^{1-\eta} \times \mathbf{1}\left\{W_o/Z_{od}(v) = P_d(v)\right\}\right] = E\left[\left(P_d(v)/P_d\right)^{1-\eta}\right] \Pr\left[W_o/Z_{od}(v) = P_d(v)\right] = \Pr\left[W_o/Z_{od}(v) = P_d(v)\right]$. This result does not rely on CES preferences (see online Appendix O.5).

For the case of the CNCES correlation function in (6), expenditure shares are the result of adding nest-level expenditure shares, which, from a Ricardian perspective, we interpret as expenditure on goods made with each latent technology:

$$(12)\quad \pi_{od} = \sum_{k=1}^{K} \pi^{*}_{kod} \text{ with } \pi^{*}_{kod} = \underbrace{\frac{\left(\omega_{kod} P_{od}^{-\theta}\right)^{\frac{1}{1-\rho_k}}}{\sum_{o'=1}^{N}\left(\omega_{ko'd} P_{o'd}^{-\theta}\right)^{\frac{1}{1-\rho_k}}}}_{\pi^{W}_{kod}} \underbrace{\frac{\left[\sum_{o'=1}^{N}\left(\omega_{ko'd} P_{o'd}^{-\theta}\right)^{\frac{1}{1-\rho_k}}\right]^{1-\rho_k}}{\sum_{k'=1}^{K}\left[\sum_{o'=1}^{N}\left(\omega_{k'o'd} P_{o'd}^{-\theta}\right)^{\frac{1}{1-\rho_{k'}}}\right]^{1-\rho_{k'}}}}_{\pi^{B}_{kd}}.$$

The variable π^{*}_{kod} is the share of overall expenditure on goods made with latent factor k that destination d sources from origin o. The variable $\pi^{W}_{kod} \equiv \pi^{*}_{kod}/\sum_{o'} \pi^{*}_{ko'd}$ is the *within-factor share*, and the variable $\pi^{B}_{kod} \equiv \sum_{o'} \pi^{*}_{ko'd}/\sum_{k'}\sum_{o'} \pi^{*}_{k'o'd}$ is the *between-factor share*. In this case, the cross-price expenditure elasticity $(o \neq o')$ is

$$(13)\qquad \varepsilon_{oo'd} = \sum_{k=1}^{K} \frac{\pi^{*}_{kod}}{\pi_{od}} \frac{\partial \ln \pi^{*}_{kod}}{\partial \ln P_{o'd}/P_d} = \theta \sum_{k=1}^{K} \frac{\rho_k}{1-\rho_k} \pi^{W}_{kod} \pi^{W}_{ko'd} \frac{\pi^{B}_{kd}}{\pi_{od}}.$$

When two origins have similar within-factor expenditure shares in a destination, they are strong head-to-head competitors and this elasticity is high. Similarly, when two countries have expenditure concentrated on factors with high correlation across countries (high ρ_k), they are more substitutable. In contrast, elasticities are low for competitors with dissimilar within-factor shares and/or in factors with low correlation across countries (low ρ_k). When $\rho_k = 0$ for all k (or there is no similarity in within-factor expenditure between o and o'), we are back to the CES case where cross-price elasticities are zero and IIA holds.

Virtually all models in the existing quantitative literature inspired by EK have a CNCES demand system, as in (12). That is, they fit into the GEV class. The connection arises from interpreting nests as corresponding to observable categories such as sectors, regions, multinational firms, or global value chains. For the case of sectors, this means pairing each latent factor with a unique sector, which amounts to assuming that sectors do not share technologies. This restriction implies, as we show in Section IIIA, that cross-price expenditure elasticities are zero across sectors.[11]

These cases are examples of the following general equivalence result.

COROLLARY 1 (GEV Equivalence): *For any trade model that generates a GEV import demand system, there exists a Ricardian model with max-stable multivariate Fréchet productivity that generates the same import demand system.*

Corollary 1 provides an "umbrella" for a large class of models in the trade literature by pairing any model with expenditure in the GEV class to a max-stable

[11] For instance, in the multinational production model in Ramondo and Rodríguez-Clare (2013), where the home country of a technology may differ from the location where it is used for production, each nest in (12) is paired with the home country of the technology. See online Appendix O.2.

multivariate Fréchet Ricardian model.[12] Despite their distinct microfoundations, all the models in the GEV class can be tied to a common Ricardian interpretation where aggregate productivity is max-stable multivariate Fréchet. Moreover, these models share identical macro counterfactuals, as we explain next.

B. *Macro Counterfactuals with GEV*

We next show that heterogeneity in correlation leads to heterogeneity in the gains from trade.[13] Specializing (8) to self-trade, and using the expression for the price index in (9), we can write the real wage in country d as

$$\frac{W_d}{P_d} = \gamma^{-1} T_{dd}^{1/\theta} (\tilde{\pi}_{dd})^{-\frac{1}{\theta}}, \tag{14}$$

where $\tilde{\pi}_{dd} \equiv \pi_{dd}/G_d^d\big(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}\big) = \big(P_{dd}/P_d\big)^{-\theta}$ reflects the real price of domestically produced goods which, in turn, summarizes correlation of d with the rest of the world. Using (14) for fixed T_{dd}, the change in real wages between two equilibria reflects the change in correlation-adjusted self-trade shares:

$$\frac{W_d'/P_d'}{W_d/P_d} = \left(\frac{\tilde{\pi}_{dd}'}{\tilde{\pi}_{dd}}\right)^{-\frac{1}{\theta}}. \tag{15}$$

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 (15) collapses to

$$\frac{W_d/P_d}{W_d^A/P_d^A} = (\tilde{\pi}_{dd})^{-\frac{1}{\theta}}. \tag{16}$$

This expression generalizes the sufficient-statistic approach of ACR to the class of models with GEV import demand systems. Crucially, the sufficient statistic is no longer the self-trade share—it is now necessary to adjust self trade to account for cross-country correlation. Under independence, the correlation function is additive, and the gains from trade in (16) simplify to the ones in ACR, $\tilde{\pi}_{dd} = \pi_{dd}$: two countries with the same self-trade share have the same gains from trade. However, the expression in (16) admits the possibility that two countries with the same self-trade share have different gains depending on how similar they are to other countries. In particular, when productivity is more similar across countries, the forces of comparative advantage weaken and trade produces lower gains.

We next focus on the CNCES case, which yields a closed-form expression for (16). Using the expenditure shares in (12), the gains from trade relative to autarky are (see Appendix C for derivations)

$$\frac{W_d/P_d}{W_d^A/P_d^A} = \pi_{dd}^{-1/\theta} \left[\sum_{k=1}^{K} \frac{\left(\pi_{kdd}^W\right)^{1-\rho_k} \pi_{kd}^B}{\pi_{dd}}\right]^{-\frac{1}{\theta}}. \tag{17}$$

[12] By adapting results from the discrete choice literature (Dagsvik 1995), we go a step further and show that GEV import demand systems are dense in the space of import demand system generated by Ricardian models with *any* productivity distributions (see online Appendix O.3).

[13] Online Appendix O.4 presents the model equilibrium formally and exact hat-algebra methods.

The second term on the right-hand side captures how correlation affects the gains from trade relative to the case of independent productivity. Conditional on factor-level expenditure, more correlation within any nest k reduces the gains from trade, and more so if self-trade expenditure for that factor is high. For $\rho_k = 0$ for all k, the gains from trade reduce to the ACR formula, and θ is the only parameter regulating the gains from trade. In contrast, for $\rho_k \rightarrow 1$ for all k, there are no gains from trade regardless of the value of θ. Intuitively, if all countries have identical production possibilities, there are no gains for trade—dispersion in productivity across goods within a country is no longer relevant—and there is no trade. For intermediate values of ρ_k, both dispersion in productivity across goods, controlled by θ, and correlation in productivity across countries, controlled by ρ_k, matter. Either higher θ or higher ρ_k reduces the strength of comparative advantage and decreases the gains from trade. Summing up, the gains from trade depend not only on the shape parameter θ, as in ACR, but also on factor-level correlation parameters ρ_k, and factor-level expenditure shares.

Next, for further intuition, we provide a three-country example.

A Three-Country Example.—Consider a world with three countries with identical size. Assume that productivity is max-stable Fréchet with common scale and correlation function $G^d(x_1, x_2, x_3) = \left(x_1^{1/(1-\rho)} + x_2^{1/(1-\rho)}\right)^{1-\rho} + x_3$. This nested CES correlation function—a special case of CNCES with $K = 2$—includes countries 1 and 2 in the first nest and only country 3 in the second nest. 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. The gains from trade are

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

The gains from trade for country 3 simply reflect their self-trade share. But the gains from trade for countries 1 and 2 depend on the degree of correlation in their technology. Conditional on expenditure, when $\rho = 0$, we get the ACR formula; for $\rho > 0$, correlation lowers the gains from trade; for $\rho \rightarrow 1$, the two countries are effectively a single country and the gains from trade depend on their combined self trade.[14]

If we do not condition on expenditures, but rather solve for their equilibrium values, the intuition carries over. In an otherwise identical world with frictionless trade, although wages equalize between countries 1 and 2, heterogeneity in correlation precludes wage equalization with country 3. Specifically, countries 1 and 2 have lower gains from trade because they have correlated productivity.[15]

[14] Normalizing the (common) price level to 1 and solving for trade shares yields $\pi_{od} = 2^{-\rho} W_o^{-\theta}$, for $o = 1, 2$, and $\pi_{3d} = W_3^{-\theta}$. While expenditure on goods from country 3 does not depend on ρ, trade shares from countries 1 and 2 decrease with ρ.

[15] The real wage in country 3 is $W_3 = \left(1 + 2^{\frac{1+\theta-\rho}{1+\theta}}\right)^{1/\theta}$, while it is $W_o = 2^{-\frac{\rho}{1+\theta}} W_3$ in countries 1 and 2. Wages are decreasing in the parameter ρ.

III. Quantitative Application

In this section, we estimate the Ricardian model of trade with a cross-nested CES (CNCES) correlation function. We treat the nests of the correlation function as an unobserved dimension of the data. In order to recover these latent factors, we estimate a multisector version of our model using disaggregate sectoral trade flow and tariff data. Our estimation procedure infers factor-level expenditure from the sectoral data. By not pairing each nest to a sector, this procedure regroups observed sectors into latent technology classes. In this way, we can avoid imposing a gravity structure at the sector level—a constant own-price elasticity within each sector and IIA across sectors. Our estimation procedure provides a tractable and intuitive way to relax these restrictions.

A. *Multisector Model with Latent Factors*

We use our results from the one-sector model in Section I and reinterpret an origin country o as a sector-origin pair so. Consumers have CES preferences over the continuum of goods, with elasticity of substitution $\eta > 1$.[16] Each good $v \in [0,1]$ is produced using one of many latent factors, $k = 1,\ldots,K$, which can be thought of as unobserved technology classes. Additionally, goods are assigned a sectoral label, so that sectors $s = 1,\ldots,S$ consist of groupings of goods *observed* in the data. With this interpretation of sectors, firms choose under which sectoral label to produce a good.

Productivity for a good v assigned to sector s in country o for delivery to d is $Z_{sod}(v)$. It captures both the production-and-delivery technology, which includes components of trade costs such as geography, as well as the efficiency losses associated with choosing a particular sectoral category. We assume that productivity is distributed multivariate max-stable Fréchet with scale $T_{sod} > 0$, shape $\theta > 0$, and a CNCES correlation function with weights $\omega_{ksod} > 0$, $\sum_k \omega_{ksod} = 1$, and correlation parameters $\rho_k \in [0,1)$. Using (6) yields

$$(18)\qquad \Pr\left[Z_{sod}(v) \leq z_{so}, \forall s,o\right] = \exp\left[-\sum_{k=1}^{K}\left(\sum_{s=1}^{S}\sum_{o=1}^{N}\left(T^{*}_{ksod} z_{so}^{-\theta}\right)^{\frac{1}{1-\rho_k}}\right)^{1-\rho_k}\right],$$

where $T^{*}_{ksod} \equiv \omega_{ksod} T_{sod}$. Within k, correlation is symmetric across origins and sectors and parameterized by ρ_k. Across k, productivity draws are independent. However, because both sectors and origins can share latent factors, productivity draws are not independent across sectors and countries.

Goods shipped from country o to d in sector s are subject to tariffs t_{sod}. Destinations source goods from the sector-origin pair with the lowest unit cost, $\min_{s,o} \frac{t_{sod} W_o}{Z_{sod}(v)}$. Thanks to max-stability, sectoral expenditure can be solved in closed

[16] In contrast to the literature, consumers have preferences over individual goods directly rather than over a composite sectoral good that aggregates individual goods. See online Appendix O.2.

form and equals the share of goods sourced from sector s and origin o, taking the same form as (12): $\pi_{sod} = \sum_{k=1}^{K} \pi^*_{ksod}$ with

$$\pi^*_{ksod} = \underbrace{\frac{\left[T^*_{ksod}\left(t_{sod}W_o\right)^{-\theta}\right]^{\frac{1}{1-\rho_k}}}{\sum_{s'=1}^{S}\sum_{o'=1}^{N}\left[T^*_{ks'o'd}\left(t_{s'o'd}W_{o'}\right)^{-\theta}\right]^{\frac{1}{1-\rho_k}}}}_{\pi^W_{ksod}} \underbrace{\frac{\left\{\sum_{s'=1}^{S}\sum_{o'=1}^{N}\left[T^*_{ks'o'd}\left(t_{s'o'd}W_{o'}\right)^{-\theta}\right]^{\frac{1}{1-\rho_k}}\right\}^{1-\rho_k}}{\sum_{k'=1}^{K}\left\{\sum_{s'=1}^{S}\sum_{o'=1}^{N}\left[T^*_{k's'o'd}\left(t_{s'o'd}W_{o'}\right)^{-\theta}\right]^{\frac{1}{1-\rho_{k'}}}\right\}^{1-\rho_{k'}}}}_{\pi^B_{kd}}. \tag{19}$$

Here, π^W_{ksod} is the within-factor share across sectors and origins, and π^B_{kd} is the between-factor share.

The cross-price elasticities $(so \neq s'o')$ of (19) are

$$\varepsilon_{sos'o'd} = \theta \sum_{k=1}^{K} \frac{\rho_k}{1-\rho_k} \pi^W_{ksod}\pi^W_{ks'o'd}\frac{\pi^B_{kd}}{\pi_{sod}} \geq 0, \tag{20}$$

with the own-price elasticity $(so = s'o')$ coming from the restriction $\sum_{s',o'} \varepsilon_{sos'o'd} = -\theta$. These elasticities are of the same form as the elasticities in (13), but now they include the possibility of cross-sector—in addition to cross-origin—substitution through the sharing of latent factors: increases in the real import price in sector s' can increase expenditure shares in a different sector s. A high expenditure elasticity between two sectors can be due to high factor-level correlation ρ_k and/or two sectors with similar within-factor expenditures—high $\pi^W_{ksod}\pi^W_{ks'o'd}$. When $\rho_k = 0$ for all k, $\varepsilon_{sos'o'd} = -\theta \times \mathbf{1}\{so = s'o'\}$. This is the CES case in (10).

However, for the purpose of estimation, this model is overparameterized because the productivity distribution depends on both the observable dimensions of the data—sectors, origins, and destinations—as well as on the unobservable latent-factor dimension. To ensure that the model is not underidentified, it is necessary to add some structure to the productivity distribution so that we have at most as many parameters as available observations.

To such end, our LFM assumes that factor-level scale parameters are separable between sector-factor and factor-origin-destination components,

$$T^*_{ksod} = \left(B_{sk}A_{kod}\right)^{\theta}. \tag{21}$$

The component B_{sk} captures how useful factor k is for sector s, while A_{kod} measures the productivity of origin o in factor k when delivering to destination d, capturing barriers to apply technologies in a country as well as geographical barriers to trade (e.g., distance). The key consequence of this separability assumption is that, because B_{sk} is identical across countries, sectoral comparative advantage arises from an origin's ability to use each latent factor, measured by A_{kod}.[17]

[17] Notice that the parameter T^*_{ksod}, which is unobservable, does not include tariffs, t_{sod}, which are observable.

The separability assumption reduces the number of model parameters and helps to identify the latent factors. Concretely, replacing T^*_{ksod} in (19) by the condition in (21) yields

$$\pi_{sod} = \sum_{k=1}^{K} \left(\frac{t_{sod}}{t^*_{kod}}\right)^{-\sigma_k} \lambda_{sk} \pi^*_{kod}, \tag{22}$$

where

$$\sigma_k \equiv \frac{\theta}{1-\rho_k}, \quad \lambda_{sk} \equiv \frac{B_{sk}^{\sigma_k}}{\sum_{s=1}^{S} B_{sk}^{\sigma_k}}, \quad \text{and} \quad t^*_{kod} \equiv \left(\sum_{s'=1}^{S} t_{s'od}^{-\sigma_k} \lambda_{s'k}\right)^{-\frac{1}{\sigma_k}}, \tag{23}$$

are, respectively, the within-factor elasticity, the weight that sector s carries on factor k, and a factor-level tariff index. The key feature of (22) is that sectors load on factor-level expenditure shares π^*_{kod} through (relative) tariffs and the factor weights, λ_{sk}. These weights are identical across countries—reflecting the assumption that sectoral comparative advantage arises from a country's ability to use each latent factor.

Under (21), the model is no longer underidentified provided that the following rank condition is also satisfied:

$$K \le \frac{S \times N^2}{S + N^2}. \tag{24}$$

Note that $\frac{S \times N^2}{S + N^2} < S$, so that estimating (22) requires compressing the sectoral data on tariffs and expenditure to a lower-dimensional latent-factor level, similar to principal-component analysis.

The factor-level expenditure share in (22) is given by

$$\pi^*_{kod} = \underbrace{\frac{\left(t^*_{kod} W_o / A_{kod}\right)^{-\sigma_k}}{\sum_{o'=1}^{N}\left(t^*_{ko'd} W_{o'}/A_{ko'd}\right)^{-\sigma_k}}}_{\pi^W_{kod}} \underbrace{\frac{\left[\sum_{o'=1}^{N}\left(t^*_{ko'd} W_{o'}/A_{ko'd}\right)^{-\sigma_k}\right]^{\frac{\theta}{\sigma_k}}}{\sum_{k'=1}^{K}\left[\sum_{o'=1}^{N}\left(t^*_{k'o'd} W_{o'}/A_{k'o'd}\right)^{-\sigma_{k'}}\right]^{\frac{\theta}{\sigma_{k'}}}}}_{\pi^B_{kd}}, \tag{25}$$

where, again, the first term on the right-hand side is the within-factor expenditure share, and the second term is the between-factor share. This expression has a gravity structure as defined by ACR: IIA holds within each factor and within-factor elasticities are constant and equal to σ_k.

The functional form in (25) is reminiscent of sectoral trade models in the gravity literature. The LFM reduces to those models if we force latent factors to be specific to sectors, which amounts to restricting $B_{sk} = 0$ for $s \neq k$, in (21). In this case, $\lambda_{sk} = \mathbf{1}\{k = s\}$, and we get a gravity specification at the sectoral level:

$$\pi_{sod} = \frac{\left(t_{sod} W_o / A_{sod}\right)^{-\sigma_s}}{\sum_{o'=1}^{N}\left(t_{so'd} W_{o'}/A_{so'd}\right)^{-\sigma_s}} \frac{\left[\sum_{o'=1}^{N}\left(t_{so'd} W_{o'}/A_{so'd}\right)^{-\sigma_s}\right]^{\frac{\theta}{\sigma_s}}}{\sum_{s'=1}^{S}\left[\sum_{o'=1}^{N}\left(t_{s's'd} W_{o'}/A_{s'o'd}\right)^{-\sigma_{s'}}\right]^{\frac{\theta}{\sigma_{s'}}}}. \tag{26}$$

This sectoral gravity model (SGM) is a special case of LFM, with factor-level tariffs corresponding to observed tariffs, and factor-level expenditure corresponding to sectoral expenditure.

There are two key consequences of assuming sector-specific technology. First, the within-sector own-price elasticity of substitution is constant and equal to σ_s. Second, there is no cross-sector substitution since the elasticity in (20) collapses to

$$\varepsilon_{sos'o'd} = (\sigma_s - \theta)\pi^W_{so'd} \times \mathbf{1}\{s = s'\}. \tag{27}$$

Now, $\varepsilon_{sos'o'd} = 0$ whenever $s \neq s'$, implying that IIA holds within each sector.

These assumptions make the SGM convenient for estimation because one can exploit within-sector variation in tariffs and expenditure to estimate σ_s. The model, however, is still underidentified because $K = S$, and we need to estimate the parameters A_{sod} and elasticities σ_s. A common and relatively flexible approach to reducing the dimensionality of the SGM is to introduce a fixed-effect specification where A_{sod} is multiplicatively separable into origin-destination, sector-origin, and sector-destination components. Under these additional restrictions, SGM estimates correspond to structural estimates of σ_s.

Although the SGM restrictions are convenient for estimation, they are a direct and testable consequence of assuming that factors are specific to sectors.[18]

B. *LFM Estimation*

We propose a tractable and flexible estimation procedure based on our LFM, which departs from sectoral gravity by relaxing the assumption that factors are specific to sectors. In this way, we allow for heterogeneous and nonzero elasticities of substitution across countries and sectors. Like principal-component analysis, the LFM procedure entails compressing disaggregate sectoral data into a few latent factors. Despite being a data-compression procedure, principal-component analysis typically produces estimates with negative entries, so that we cannot use it to structurally estimate the LFM. Our theory implies that latent-factor weights and expenditure shares are all nonnegative. To estimate the LFM, we not only need an alternative to sectoral gravity, but we also need an alternative to principal-component analysis.

We use disaggregate four-digit SITC sectoral tariff and trade flow data from Comtrade, combined with World Input-Output Database aggregate sectoral expenditure data, for 1999–2007. Our sample has 787 sectors and 31 countries (see online Appendix O.9 for details on the data construction).

Our LFM estimator infers latent-factor expenditure and latent-factor weights from observed sectoral expenditure and tariffs. We assume that factor weights and factor-level elasticities are time invariant, $\lambda_{skt} = \lambda_{sk}$ and $\sigma_{kt} = \sigma_k$, for all t, while factor-level expenditure and tariffs can vary over time. For convenience, we define $\phi^*_{kodt} \equiv (t^*_{kodt})^{\sigma_k}\pi^*_{kodt}$, and rewrite (22) as

$$\pi_{sodt} = \sum_{k=1}^{K} t_{sodt}^{-\sigma_k}\lambda_{sk}\phi^*_{kodt}. \tag{28}$$

Observed sectoral shares are linear in the unobserved components λ_{sk} and ϕ^*_{kodt}, allowing us to build an estimation algorithm based on nonnegative matrix factorization,

[18] Online Appendix O.6 shows reduced-form evidence on departures from IIA within and across sectors. These results are in line with the evidence in Adao, Costinot, and Donaldson (2017).

combined with the pseudo-Poisson maximum likelihood method used in the gravity literature (Santos Silva and Tenreyro 2006; Fally 2015).

For a given choice of K, we set $\Sigma = \{\sigma_k\}_k$, $\Lambda = \{\lambda_{sk}\}_{s,k}$, and $\Phi^* = \{\phi^*_{kodt}\}_{k,o,d,t}$ to solve

$$\hat{\Sigma}, \hat{\Lambda}, \hat{\Phi}^* = \underset{\Sigma \geq 0, \Lambda \geq 0, \Phi^* 0}{\arg\min} \sum_{s,o,d,t} \ell\left(\pi_{sodt}, \sum_{k=1}^{K} t_{sodt}^{-\sigma_k} \lambda_{sk} \phi^*_{kodt}\right), \tag{29}$$

where $\ell(x, \hat{x}) \equiv 2[x \ln(x/\hat{x}) - (x - \hat{x})]$. One convenient feature of pseudo-Poisson maximum likelihood is that, as established by Fally (2015), it is the unique likelihood-based criterion that preserves the restriction that predicted aggregate bilateral expenditure matches observed expenditure. By using a Poisson deviance, we ensure that our estimates of bilateral factor-level expenditure exactly aggregate to observed bilateral trade flows, consistent with the model—i.e., our prediction for $\sum_k \pi^*_{kodt}$ matches exactly the data on π_{odt}.

To perform the data compression embedded in the LFM and solve (29), we adapt techniques from the literature on nonnegative matrix factorization. Specifically, we extend the multiplicative updating algorithm in Lee and Seung (1999, 2000) to accommodate both missing data and simultaneous estimation of σ_k. To better see the connection to those procedures, suppose for a moment that there are no tariffs. Then, we can write (28) in matrix form as $\Pi = \Lambda\Phi^*$. The nonnegative matrix factorization procedure decomposes the observed expenditure share matrix, Π, into two matrices: Λ containing the sector-specific components and Φ^* containing the country-pair-time components of sectoral expenditure. As in all factor models, an important concern is a lack of identification coming from transformations of the latent factors. However, the presence of nonnegativity constraints in (29) restricts the transformations that are feasible, meaning that the factorization is unique (up to permutation and scale of factors) under relatively general conditions (Fu et al. 2019).[19] In a nutshell, uniqueness is ensured when a large amount of data are used for the factorization.

The case of no tariffs also clarifies that tariffs are not used for the estimation of latent-factor weights and expenditures. However, they are crucial cost shifters to estimate the within-factor elasticities, σ_k. In fact, LFM uses within- and cross-sector variation in the data to estimate those elasticities—in contrast to SGM, which only uses within-sector variation to estimate sectoral elasticities σ_s. Identification of σ_k comes from the conditional independence assumption embedded in our Poisson criterion: $E[\upsilon_{sodt} | t_{sodt}, \Lambda, \Phi^*] = 0$ for $\upsilon_{sodt} \equiv \frac{\pi_{sodt}}{\sum_k t_{sodt}^{-\sigma_k} \lambda_{sk} \phi^*_{kodt}} - 1$.

We choose the number of latent factors by estimating (29) 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 $K + 1$.[20] The rank condition in (24) indicates that we could fit as many as 432 latent factors.

[19] Online Appendix O.10 describes in detail the algorithm as well as the (sufficient) conditions for identification of nonnegative matrix factorization.

[20] The Poisson deviance function is homogeneous 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.

Table 1—LFM Selection: Likelihood Ratio Test

Number of factors, K	1	2	3	4	5	6	7	8
R^2 4-digit SITC expenditure	0.725	0.79	0.804	0.826	0.835	0.938	0.937	0.936
within *odt*	0.092	0.158	0.197	0.24	0.266	0.306	0.334	0.362
Deviance	377,451	333,999	310,594	292,161	278,379	266,955	256,823	248,288
Degrees of freedom[a]	9,436	18,872	28,308	37,744	47,180	56,616	66,052	75,488
Null hypothesis		1	2	3	4	5	6	7
χ^2		43,452	23,405	18,433	13,783	11,423	10,133	8,535
Degrees of freedom		9,436	9,436	9,436	9,436	9,436	9,436	9,436
p-value		0.0	0.0	0.0	0.0	0.0	0.0	1.0

Notes: Results from estimating (29) with $K = 1, \ldots, 8; 14$. Number of observations $= 5{,}528{,}764$.
[a] Model's degrees of freedom. Last panel shows likelihood ratio tests comparing specifications across columns.

Lastly, we need to estimate the shape parameter θ and the correlation function parameters, ρ_k for each k. Recall that $\sigma_k = \theta/(1 - \rho_k)$, so that given our estimates of factor-level elasticities, choosing a value for θ pins down each ρ_k. Additionally, we have the structural restriction that $\theta > 0$ and $\rho_k \geq 0$ so that $\theta \in (0, \sigma_k]$ for each k. The largest possible value of θ that is consistent with our estimates of factor-level elasticities is $\min_{k=1,\ldots,K} \hat{\sigma}_k$. We use this upper bound on the shape parameter as our baseline estimate. This value ensures that we conservatively estimate of the gains from trade because, *conditional on our estimates of expenditure shares and elasticities at the factor-level*, the gains from trade decrease as θ increases.[21] For robustness, we implement an alternative estimation of θ, based on the factor-level gravity structure of (25), that uses the between-factor variation produced by the LFM procedure—i.e., the estimated factor-level expenditures and tariff indices. This two-step procedure yields estimates of θ that are not statistically different from our baseline estimate (see online Appendix O.8 for details).

C. *Results*

We next analyze the results from the LFM estimation. When we show variables at a higher level of aggregation than the four-digit SITC level, we use the factor weights λ_{sk} to aggregate.

We estimate that the number of latent factors is $K = 7$. Table 1 shows that $K = 8$ is not significantly different from $K = 7$. Seven factors explain about 94 percent of the variation in the sectoral trade flow data, and more than 33 percent of the variation in expenditure shares within each origin-destination.[22]

Next, we examine our estimates of factor-level elasticities, factor weights, and factor-level expenditure.

The first panel of Table 2 shows estimates of factor elasticities, σ_k. We rank the latent factors according to their elasticities from largest (F1) to lowest (F7). Given

[21] From (17) and Hölder's inequality, for any $\theta' \in (0, \theta]$,

$$\frac{W_d/P_d}{W_d^A/P_d^A} = \left[\sum_{k=1}^{K} \left(\pi_{kdd}^W\right)^{\frac{\theta}{\sigma_k}} \pi_{kd}^B\right]^{-\frac{1}{\theta}} \leq \left[\sum_{k=1}^{K} \left(\pi_{kdd}^W\right)^{\frac{\theta'}{\sigma_k}} \pi_{kd}^B\right]^{-\frac{1}{\theta'}} \xrightarrow{\theta' \to 0} \prod_{k=1}^{K} \left(\pi_{kdd}^W\right)^{-\frac{\pi_{kd}^B}{\sigma_k}}.$$

[22] In online Appendix O.7, we show that our LFM estimates capture quite accurately departures from IIA within and across sectors. We also show that estimates of a version of LFM where Λ is constrained to match the restrictions of SGM are statistically different from estimates of LFM with $K = 7$ and has less explanatory power.

Table 2—Estimates of Latent-Factor Elasticities, Weights, and Expenditure; Summary

	Factor						
	F1	F2	F3	F4	F5	F6	F7
σ_k	5.175	4.869	4.625	1.482	0.671	0.390	0.375
	(0.142)	(0.091)	(0.142)	(0.130)	(0.076)	(0.182)	(0.091)
ρ_k	0.927	0.923	0.919	0.747	0.44	0.038	0.00
	Factor weights: four-digit SITC sectors						
Zero share	0.108	0.177	0.062	0.113	0.16	0.233	0.174
Ninetieth percentile	0.003	0.002	0.003	0.003	0.002	0.001	0.001
Ninety-ninth 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: WIOD sectoral aggregates						
1. Agriculture, hunting, forestry, and fishing	0.071	0.004	0.02	0.019	0.23	0.003	0.027
2. Mining and quarrying	0.007	0.0	0.002	0.003	0.122	0.003	0.587
3. Food, beverages, and tobacco	0.109	0.005	0.117	0.051	0.247	0.008	0.035
4. Textiles and leather	0.43	0.032	0.024	0.017	0.057	0.01	0.005
5. Wood and products of wood and cork	0.017	0.003	0.004	0.05	0.017	0.001	0.009
6. Pulp, paper, paper, printing, and publishing	0.004	0.004	0.031	0.126	0.04	0.039	0.014
7. Coke, refined petroleum, and nuclear fuel	0.005	0.001	0.005	0.118	0.002	0.001	0.035
8. Chemicals, rubber, and plastics	0.066	0.092	0.337	0.167	0.03	0.047	0.062
9. Other nonmetallic mineral	0.038	0.022	0.009	0.03	0.009	0.002	0.003
10. Basic metals and fabricated metal	0.064	0.066	0.037	0.221	0.074	0.012	0.169
11. Machinery, nec	0.051	0.132	0.166	0.094	0.015	0.042	0.011
12. Electrical and optical equipment	0.05	0.147	0.131	0.046	0.016	0.778	0.007
13. Transport equipment	0.02	0.475	0.083	0.021	0.128	0.034	0.01
14. Manufacturing, nec; recycling	0.067	0.017	0.033	0.038	0.011	0.019	0.028
	Factor-level expenditure shares						
Expenditure share	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: Standard errors for σ_k are in parentheses. $\theta = \min_k \sigma_{k=1,\ldots,K} = 0.375$ with $\rho_k = 1 - \theta/\sigma_k$.

that $\theta = \min_k \sigma_k = 0.375$, the factor with the highest correlation across countries is F1, with $\rho_1 = 0.927$, and the factor with the lowest correlation is F7, with $\rho_7 = 0$. Additionally, the standard errors indicate that elasticities are tightly estimated.[23]

The second panel of Table 2 presents statistics for the factor weights. First, the fraction of factor weights that are zero ranges from 6.2 to 23.3 percent (e.g., 23.3 percent of four-digit SITC sectors do not use technologies related to F6). Second, each factor is concentrated in a few four-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 2 shows that each factor has some weight on essentially every WIOD-aggregate sector.[24]

[23] The average across σ_ks is 2.51, very close to the estimate of the gravity tariff elasticity (see online Appendix O.6). This should not be surprising since the LFM captures very well the correlation between tariffs and expenditure observed in the data. Online Appendix O.11 shows that while the heterogeneity in estimates of σ_k increases with K, the average across σ_ks remains around 2.5 to 3 for all K.

[24] For additional results, see online Appendix O.11.

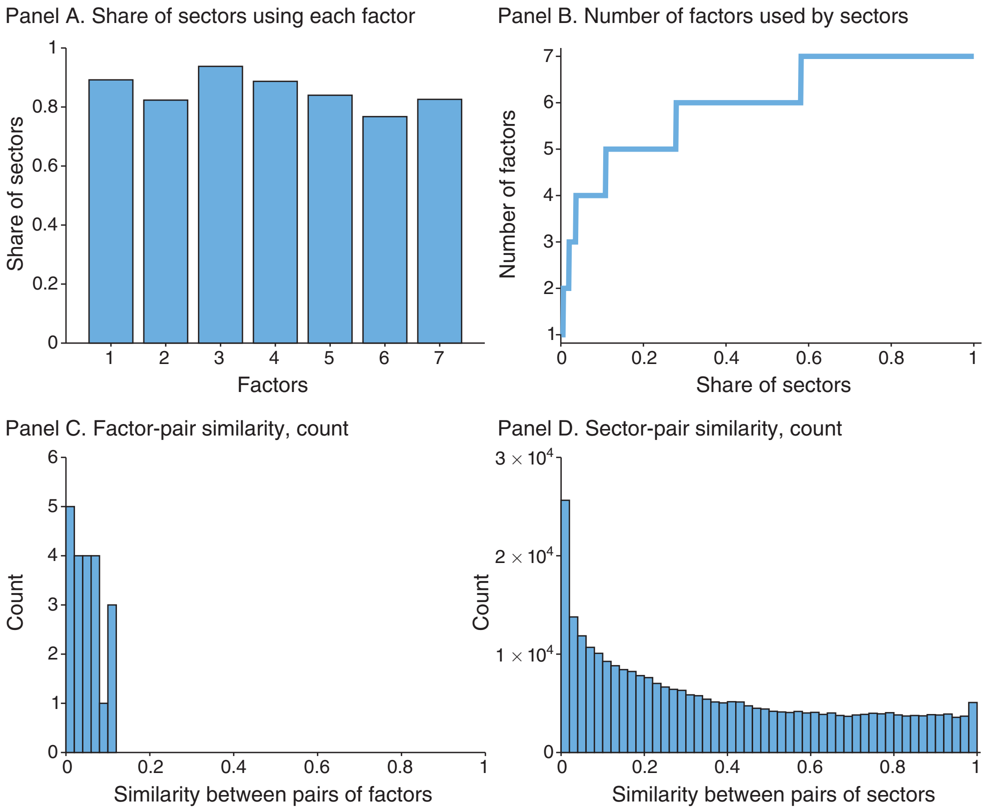

Figure 1. Factor Weights: Extensive and Intensive Margins

Notes: Sectors are four-digit SITC. Similarity refers to, in panel C, $\sum_s \lambda_{sk}\lambda_{sk'}/\sqrt{\sum_s \lambda_{sk}^2 \sum_s \lambda_{sk'}^2}$, and in panel D, $\sum_k \lambda_{sk}\lambda_{s'k}/\sqrt{\sum_k \lambda_{sk}^2 \sum_k \lambda_{s'k}^2}$.

Figure 1 panels A and B show that factors are not unique to sectors. Less than 15 percent of four-digit SITC sectors use less than four factors, while about 75 percent use at least six out of the seven factors. Additionally, Figure 1 panels C and D examine how intensively sectors (factors) use pairs of factors (sectors) by plotting histograms of a similarity measure constructed using factor weights (0 $=$ orthogonal weights; 1 $=$ identical weights). Similarity is concentrated close to zero for all factor-pairs, consistent with the interpretation that latent factors are groups of distinct technologies, so that they weigh on sectors in distinct ways. But many sector pairs never load on the same factors (low similarity) and many sector pairs weigh on factors similarly.

To get some interpretation for each factor, we use our estimates of λ_{sk} to examine how factors load on sectors, and our estimates of π^*_{kod} to get patterns of expenditure, export intensity, and domestic absorption by factor. First, the bottom panel of Table 2 shows that F4 and F5 make up the majority of global expenditure, and are barely traded. In contrast, the remaining factors are heavily traded, with self-trade shares ranging from around 40 to 50 percent.

To get a sense of the identity of each factor, we turn to the use of factors across sectors. Table 3 reports the top-three two-digit SITC sectors with the highest weights

Table 3—Factor Weights: Top-Three Two-Digit SITC Sectors

Factor	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	Nonferrous metals	0.092

Notes: Factor weights from estimating (29) and aggregating to two-digit SITC. NES = not elsewhere specified.

in each factor. For example, F2 is mainly used in the production of "machinery and transport equipment," and Germany, Japan, and the United States are the countries using this technology the most, as measured by each country's share of total exports that rely on this factor. F6 relates to highly specialized manufactured goods such as electronics and scientific instruments, and, according to Table 2, China overtook the United States as the main exporter of this factor between 1999 and 2007. 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. For the remaining factors, we see that F3 relates to medical products, chemicals, and industrial equipment, while F4 and F5 are used for making less complex products, related to basic materials, and are the two factors most related to self trade.

The estimates of factor-level expenditure and elasticities shape the structure of the correlation function.[25] Next, we show correlation patterns and aggregate elasticities implied by the LFM estimates. We compare them with estimates using the SGM and CES model. For SGM, we use a gravity specification to estimate sectoral elasticities for each WIOD aggregate sector, and use the same estimate of θ as for LFM so that we ensure that differences in results solely come from the correlation function. For the CES model, we use a gravity specification that assumes that elasticities are the same across sectors, and estimate an elasticity of 2.65 (this is the trade elasticity under the independence restrictions that lead to CES).[26]

[25] The weights of G^d can be recovered as $\omega_{kod} = \frac{\left(\pi^W_{kod}\right)^{1-\rho_k}\pi^B_{ko'd}}{\sum_{k'=1}^{K}\left(\pi^W_{k'od}\right)^{1-\rho_{k'}}\pi^B_{k'o'd}}$.

[26] See online Appendix O.6 and O.8 for details on the estimation of the SGM and CES models.

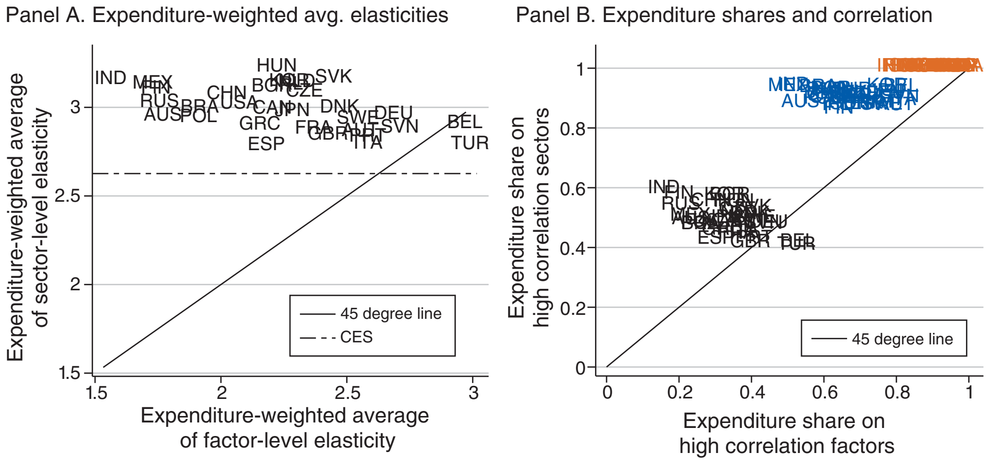

Figure 2. Elasticities, Expenditure Shares, and Correlation: Latent Factors versus Sectors

Notes: Panel A: average across factor-level (sector-level) elasticities weighted by between-factor (between-sector) expenditure share in country d. Panel B: share of country d's total expenditure on factors (sectors) with correlation coefficient higher than 0.4 (orange), 0.7 (blue), and 0.85 (black). Year 2007.

Correlation Patterns.—How much substitutability and correlation do our estimates imply? How do they compare with estimates from the SGM and CES model? These patterns are important for understanding the quantitative predictions of both models in counterfactual exercises.

First, we calculate averages of estimated factor-level elasticities σ_k, weighting by the share of total expenditure in country d on each factor k. For SGM, we calculate the averages of sector-level elasticities, weighting by the share of total expenditure in country d on each sector. While the unweighted average across these elasticities is around 2.5 to 2.7 in both models, Figure 2 panel A shows that the expenditure-weighted averages are very different between models. SGM (and CES) estimates predict very similar (the same) average elasticities across countries, ranging from around 2.7 for Spain, Italy, and Turkey to around 3.2 for Hungary, the heterogeneity coming from countries on the lower end concentrating expenditure on less substitutable sectors. In contrast, LFM estimates predict large variation across countries, ranging from 1.5 for India to almost 3 for Turkey, the difference coming from the distribution of expenditure across factors, in each country. Note that the larger average elasticities are similar between the two models—for instance Belgium and Turkey—which reveals that the heterogeneity in LFM is driven by some countries having more expenditure concentrated on low-elasticity factors. This suggests that LFM will predict that there is more heterogeneity across countries in terms of how their productivity correlates with the rest of the world. Given that $\rho_k = 1 - \theta/\sigma_k$ (sectors for SGM), Figure 2 panel B confirms that the share of expenditure in high correlation factors in LFM is almost always lower than the share in high correlation sectors in SGM for all countries, regardless of the how we define the cutoff for "high correlation." This means that, according to LFM, most expenditure is difficult to substitute, while according to SGM, countries can substitute more easily.

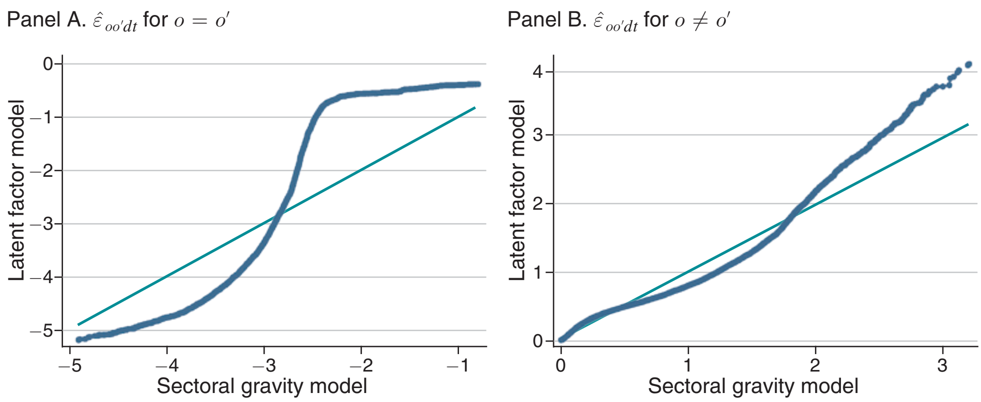

Figure 3. Expenditure Elasticities: LFM versus SGM

Notes: Expenditure elasticities $\varepsilon_{oo'dt}$ calculated using (30) and LFM (SGM) estimates. Quantiles of $\varepsilon_{oo'dt}$ for LFM versus SGM. Subscript t refers to years 1999 to 2007.

Expenditure Elasticities.—We next compute the aggregate expenditure elasticities, $\varepsilon_{oo'd}$, implied by our LFM estimates, and compare with those from the SGM.

Aggregating (22) across sectors and taking log derivatives yields the aggregate elasticities

$$\varepsilon_{oo'd} = \sum_{s,s'} \frac{\pi_{sod}}{\pi_{od}} \varepsilon_{sos'o'd}, \tag{30}$$

with $\varepsilon_{sos'o'd}$ given by (20) for LFM and (27) for SGM.

Both models deliver positive aggregate effects for $o \neq o'$. However, elasticities differ substantially between the two models, with the difference coming from the restriction of SGM to $\varepsilon_{sos'o'd} = 0$ for $s \neq s'$. Figure 3 shows that this restriction matters quantitatively. We use a quantile-quantile plot for visual purposes. Implied aggregate elasticities are different across the two models, particularly the own-price elasticity of substitution in Figure 3 panel A. This is related mainly to the different predictions of the models in terms of the distribution of expenditure across factors (sectors) with different degrees of correlation, as shown in Figure 2.

Figure 4 compares the values for elasticities for the LFM and the SGM focusing on expenditure by US consumers. Figure 4 panel A zooms into China and its competitors in serving the United States, $\varepsilon_{o,CHN,USA}$, for $o \neq CHN$. Figure 4 panel B considers the own-price elasticity for each origin country in our sample that serves the US market, $\varepsilon_{o,o',USA}$, for $o = o'$.

LFM estimates indicate that Chinese goods are close substitutes for goods from Turkey, Bulgaria, and Greece, for US consumers, while they are very poor substitutes for goods from Ireland, the Netherlands, Russia, and the United States itself. In contrast, estimates from the SGM imply more similar cross-price elasticities across alternative origins serving the US market and a larger own-price elasticity for China

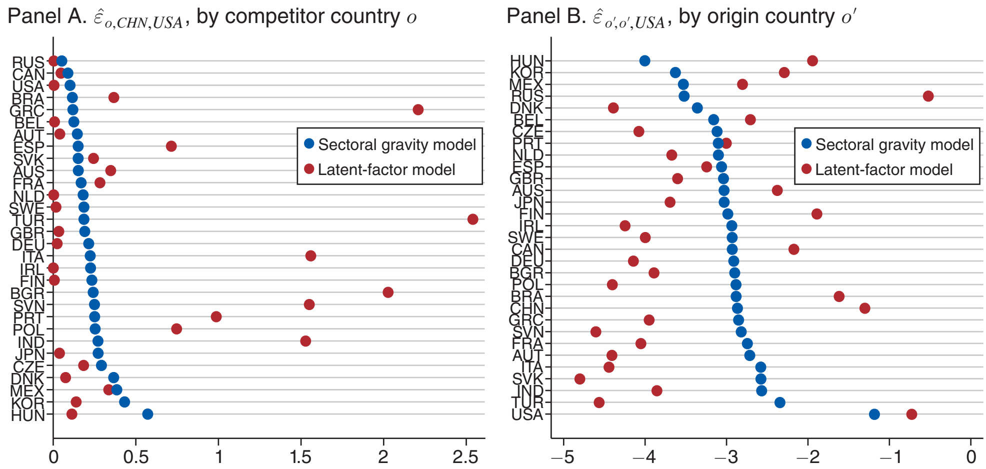

FIGURE 4. EXPENDITURE ELASTICITIES, US MARKET: LFM VERSUS SGM

Notes: Estimates of expenditure elasticities $\varepsilon_{o,o',USA}$ calculated using (30) and LFM (SGM) estimates. Panel A plots $o \neq o'$ when $o' = CHN$. Panel B plots $o = o'$ for each country o' in our sample. Year 2007.

$(-2.9$ vs $-1.3)$. In general, elasticities are much more similar across countries in the SGM than in LFM.[27]

The quantitative differences between LFM and SGM will create very different answers to counterfactual exercises, as we show next.

IV. Quantitative Exercises

Armed with our 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, aggregate expenditure, and factor-level expenditure.

A. *The Gains from Trade*

Figure 5 shows the gains from trade against self-trade shares, using the estimated versions of the LFM, the SGM, and the CES model.[28]

For LFM, we use our estimates in Section IIIC and calculate gains from trade according to (17). For the SGM, we also calculate gains using (17), but under the restriction that each latent factor k corresponds to a sector.

Finally, for the CES model, we calculate gains from trade using (17), but restricting $\rho_k = 0$ for all k (i.e., the ACR case).

In panel A, LFM estimates show that countries with the same self-trade share but different degrees of correlation with the rest of the world have different gains from trade. For instance, Canada has the same self-trade as Germany, but its implied LFM

[27] In online Appendix O.11, we plot all the implied aggregate elasticities for the US market, providing a comprehensive visualization of the differences between the two models.

[28] Online Appendix O.11 reports the exact numbers.

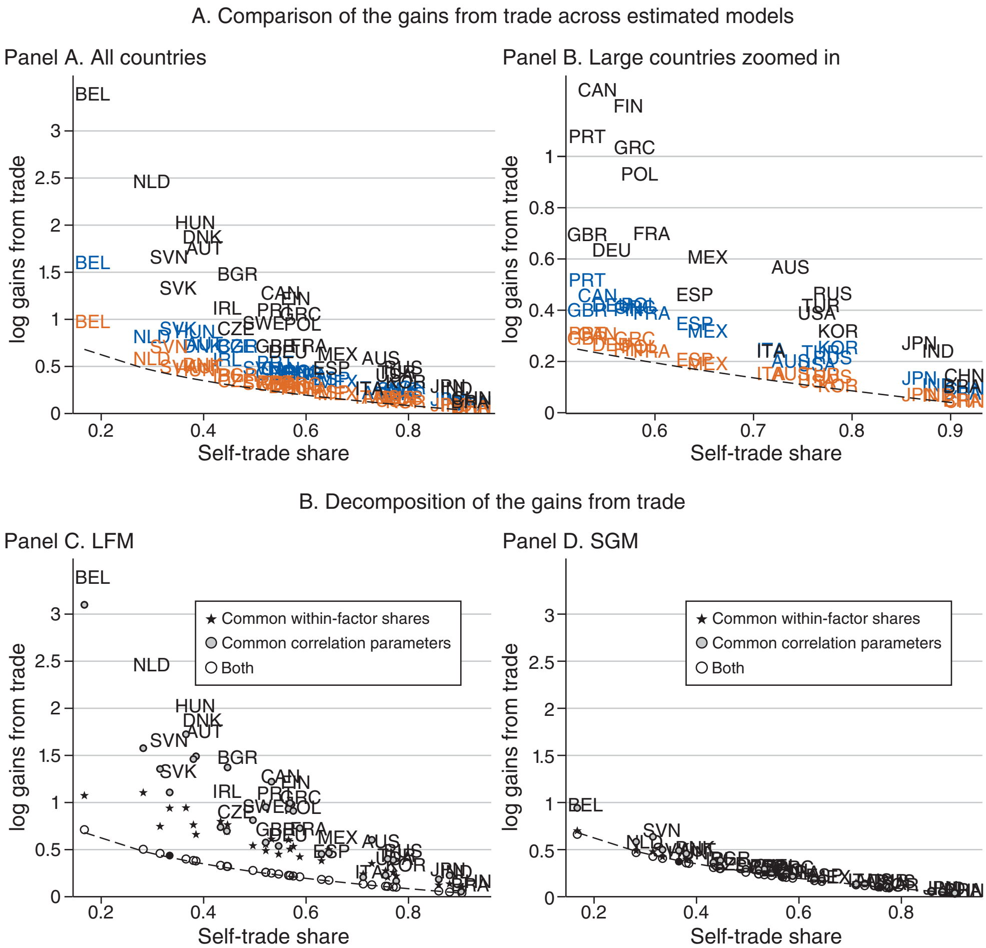

Figure 5. The Gains from Trade

Notes: Panels A and B: gains from trade = real wages in the observed equilibrium relative to autarky real wages. Calculations using estimates from LFM (black dots), SGM (orange dots), SGM augmented by input-output links (blue dots), and CES model (dashed line). Year 2007. Panels C and D: gains from trade calculated using (16). Stars removed heterogeneity in within-factor (within-sector) self-trade shares. Gray dots removed heterogeneity in correlation coefficients. White dots remove both sources of heterogeneity. The country names are baseline LFM estimates.

gains are almost 90 percent higher because it is less correlated with the rest of the world. Examining the patterns in Table 2 reveals that Canada is the top exporter of factor F7, the factor related to the production of energy and minerals and the one with the lowest correlation across countries ($\rho_7 = 0$). In contrast, Germany specializes in factors F2 and F3, which present high correlation in productivity across countries, and in factor F4, for which we estimate very high self-trade shares (i.e., it is barely traded).

The heterogeneity in correlation that we estimate under LFM leads to gains from trade that are much more heterogeneous than the gains calculated using the estimates from CES and SGM—controlling for self-trade, the standard deviation differs by an order of magnitude (2.6 vs 0.07). For instance, the CES model delivers virtually the same gains for Canada and Germany—because they have almost identical self-trade shares. The SGM, with its restrictive way of incorporating correlation, barely increases the differences in gains among these countries relative to CES (from 1 to 4 percent).

With respect to LFM, the quantitative version of SGM also delivers lower gains, including for the large countries in our sample. Given the estimates shown in Section IIIC, this result should not be surprising: the LFM estimates that less expenditure happens in factors with high correlation, while the SGM estimates more expenditure in sectors with high correlation, implying a correlation function for SGM with much higher similarity between trading partners than the one estimated under LFM.

Panel B of Figure 5 further explores the sources of the quantitative differences in gains between the two models by decomposing the gains in (16). The country names show the baseline LFM estimates. The stars show what gains would be if we removed the heterogeneity in within-factor (within-sector for SGM) self-trade expenditure shares and replace them by observed self-trade shares. The gray dots show gains if we, instead, removed heterogeneity in correlation coefficients and replace them by an average given by $\bar{\rho} = 1 - \theta/\bar{\sigma}$, with $\bar{\sigma}$ denoting the average over factors (sectors for SGM). Finally, the white dots show gains if we removed both heterogeneity in within-factor self-trade shares and heterogeneity in correlation coefficients. By construction, when both sources of heterogeneity are removed, the gains from trade reduce to the ACR formula, and, since the average elasticity of LFM (and SGM) is about the same as the trade elasticity estimated under CES, we get nearly identical gains.

We can see from this decomposition that both sources of heterogeneity matter, but a larger portion of the gains under LFM relative to ACR and SGM are driven by heterogeneity in within-factor self-trade expenditure shares. Importantly, these within-factor shares are what the LFM procedure uses to match the substitution patterns in the data that the SGM fails to capture. It is primarily the difference between within-factor versus within-sector expenditure shares that explains the quantitative difference in gains between the two models.

The large quantitative difference in the gains from trade implied by the different models indicates that it matters how correlation in productivity is introduced. In particular, it is important to let the data reveal correlation patterns rather than restricting those patterns across sectors and countries. Having a tractable and flexible procedure, like LFM, to estimate those patterns is key.

Input-Output Linkages.—Even though input-output sectoral linkages are a different economic mechanism from correlation in productivity, they may result in similar quantitative predictions. Here, we compare the gains from trade calculated using our LFM estimates with the gains from trade implied by the estimated SGM augmented by input-output linkages.[29]

Following the literature (e.g., Costinot and Rodríguez-Clare 2014), we introduce these linkages assuming that each sector s and country o has a Cobb-Douglas production function that combines labor, with share $1 - \alpha_{so} \in [0,1]$, and a composite

[29] Online Appendix O.2 presents the sectoral model with input-output linkages in detail.

input from each sector, with shares $\alpha_{ss'o} \in [0,1]$ and $\sum_{s'} \alpha_{ss'o} = \alpha_{so}$. Each sectoral input aggregates individual goods according to a CES function. Consequently, the cost of the input bundle in country o and sector s is $c_{so} = A_s W_o^{1-\alpha_{so}} \prod_{s'} P_{s'o}^{\alpha_{ss'o}}$, with $A_s > 0$ and $P_{s'o}$ the CES price index associated with the composite sectoral good.

This structure results in the same sectoral expenditure shares as in (26), with the sectoral input bundle c_{so} replacing the wage W_o. In particular, the gravity structure of SGM is preserved and the elasticities σ_s are the same as for the model without the input-output structure. The variable c_{so}, which captures the impact of input-output linkages on unit costs, is simply absorbed into a sector-origin fixed effect of the gravity specification.

Gains from trade, however, do change, and are given by

$$\frac{W_d/P_d}{W_d^A/P_d^A} = \left\{ \sum_{s=1}^{S} \left[\prod_{s'=1}^{S} \left(\pi_{s'dd}^W\right)^{-\frac{a_{ss'd}}{\sigma_{s'}^m}} \right]^{-\theta} \pi_{sd}^B \right\}^{-\frac{1}{\theta}}, \tag{31}$$

where $a_{ss'd}$ is elements of the Leontief inverse matrix, $(\mathbf{I} - \mathbf{A}_d)^{-1}$, with $\alpha_{ss'd}$ the typical element of $\mathbf{A}_d$. This formula collapses to the one in Costinot and Rodríguez-Clare (2014) when $\theta \to 0$.

The blue dots in Figure 5 are the gains from trade implied by the SGM with input-output sectoral linkages. Gains present a similar pattern to the gains coming from the SGM without those linkages, but, as is well known, since these linkages act as an amplification mechanism for trade, gains are higher for all the countries in the sample. However, gains are still less heterogeneous—and lower—than the gains from trade implied by LFM. For instance, the difference in gains between Canada and Germany does not increase. The overall variation in gains across countries (controlling for self-trade) only increases from 0.07 to 0.26, far from the standard deviation of 2 implied by the LFM estimates. These results suggest that the forces captured by LFM are different from the ones captured by sectoral input-output linkages.

B. *The Cost of Protectionism*

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

$$\frac{d\ln W_d/P_d}{d\ln t_{o'd}} = \underbrace{(1 - \pi_{dd})\frac{d\ln W_d/W_{o'}}{d\ln t_{o'd}}}_{\text{domestic wage effect}} \underbrace{\sum_{o \neq o' \text{ and } o \neq d} \pi_{od} \frac{d\ln W_o/W_{o'}}{d\ln t_{o'd}}}_{\text{third party effect}} \underbrace{\pi_{o'd}}_{\text{direct tariff effect}}, \tag{32}$$

where $t_{o'd} \equiv \left[\sum_{k=1}^{K} \left(t_{ko'd}^*\right)^{-\theta}\right]^{-\frac{1}{\theta}}$. 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 on d of increasing tariffs on o'.

Figure 6 focuses on the effects of increasing US tariffs on China from 0 to 100 percent. We compute each component of the change in US real wages by integrating each term of (32) from 0 to Δt where Δt is the total change in tariffs shown on the x-axis. Our computations reveal that, for instance, the US welfare cost of imposing a 50 percent tariff on China doubles in the LFM. The cumulative effect

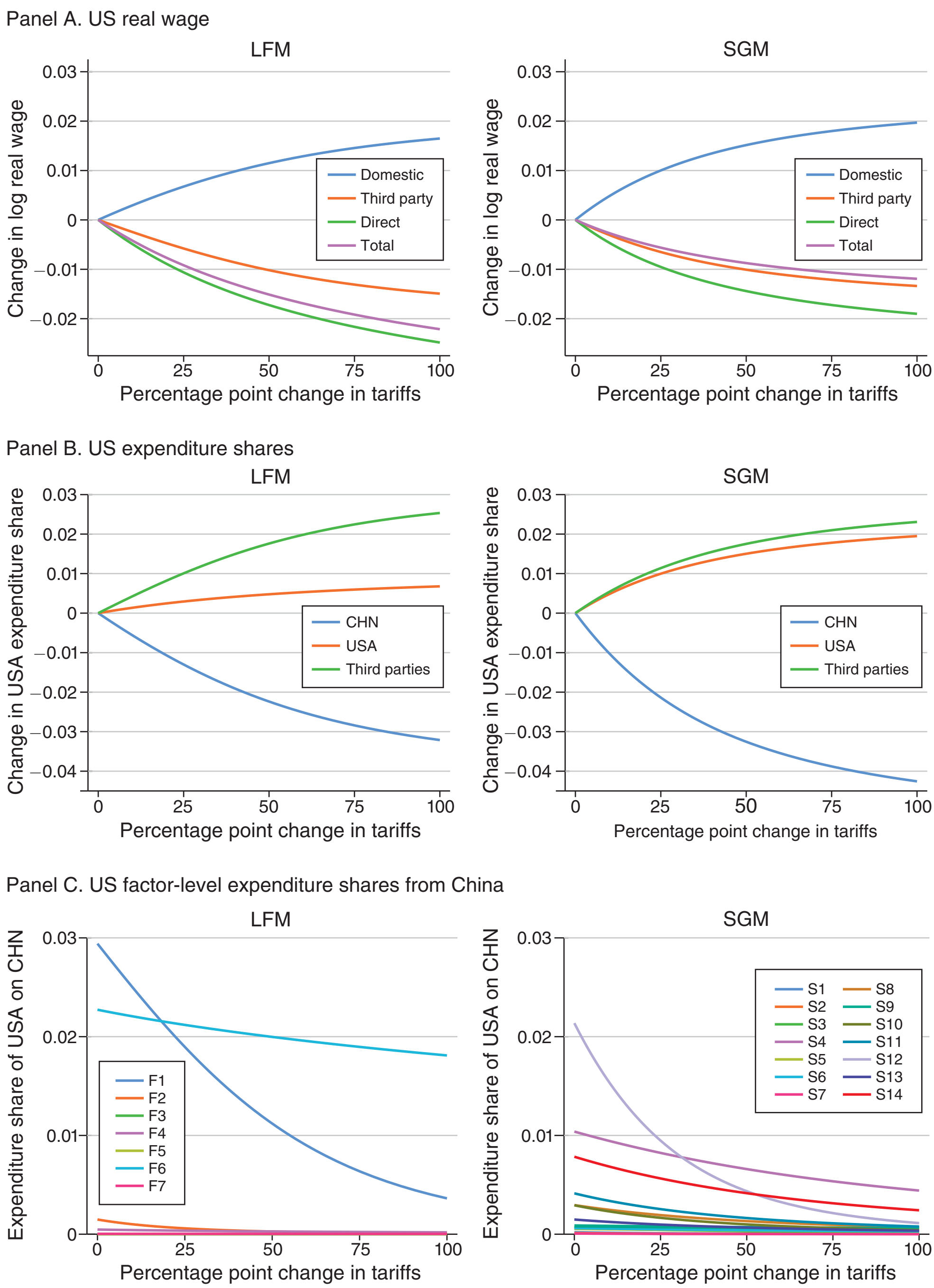

Figure 6. Effect of Increases in US Tariffs on China: LFM versus SGM

Notes: Changes in log US real wage are decomposed using (32). Year 2007.

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 toward their own goods and more toward third parties in the LFM (second panel of

Figure 6). Additionally, the cumulative direct effect of higher tariffs is larger in the LFM because this effect is proportional to expenditure shares, and, as tariffs rise, US consumers shift expenditure away from China, dampening the effect. However, US consumers substitute less away from China in the LFM, and the cumulative direct cost on US consumers from rising tariffs is larger in the LFM than in the SGM.[30]

The difference in substitution patterns between the two models comes from differences in expenditure shares across latent factors, which correspond to sectors in the SGM. The bottom panel of Figure 6 shows that, for all factors, US expenditure shifts away from China when tariffs rise. However, it does so much 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 goods such as electronics has a very low correlation across countries and US consumers do not rapidly shift their expenditure away from China. When latent factors correspond to sectors, own-price sectoral elasticities tend to be more elastic, creating more similarity across exporters. Consequently, shifts in US expenditure away from Chinese goods occur more rapidly.

V. 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 model 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. We propose a CNCES structure for correlation that departs from the existing models by treating the nests as unobserved dimensions of the data, and allows us to relax commonly made distributional assumptions. In the context of a multisector trade model, we develop a flexible estimation procedure based on compressing highly disaggregate (sectoral) data into few latent factors. Our estimates successfully capture the rich substitution patterns observed across countries and sectors, and find substantial heterogeneity in correlation patterns. The implied gains from trade are much more heterogeneous across countries than from estimates of models that restrict correlation patterns.

Appendix A. Proof of Proposition 1

PROOF:

We show that for any max-stable multivariate Fréchet random vector there exists a sequence of CNCES correlation functions that converges uniformly on compact sets to the true correlation function. The proof is constructive and, to simplify notation, we suppress the destination index, d, and the variety index, v.

[30] In online Appendix O.11, we show moments of the elasticity in (32) and its components including each country pair in our sample. Results confirm that, on average, the smaller direct wage effect in the LFM together with the larger third-party effect combine such that the cost of increasing tariffs is typically larger in the LFM than in the SGM.

Let $\{Z_o\}_{o=1}^N$ be distributed max-stable multivariate Fréchet. Then by Theorem 1 in Kabluchko(2009), $\{Z_o\}_{o=1}^N$ has a spectral representation—there exists a σ-finite measure space $(\mathcal{X}, \mathbb{X}, \mu)$; spectral functions $A_o : \mathcal{X} \to \mathbb{R}_+$ with $\int_{\mathcal{X}} A_o(\chi)^\theta d\chi < \infty$ for each $o = 1, \ldots, N$; and a Poisson process on $\mathbb{R}_+ \times \mathcal{X}$ with points $\{Q_i, \chi_i\}_{i=1,2,\ldots}$ and intensity $\theta q^{-\theta-1} dq d\mu(\chi)$ such that $Z_o = \max_{i=1,2,\ldots} Q_i A_o(\chi_i)$ for each $o = 1, \ldots, N$. Given this Poisson process, we can express the joint distribution of $\{Z_o\}_{o=1}^N$ as

$$\begin{aligned}
\Pr[Z_1 \le z_1, \ldots, Z_N \le z_N] &= \Pr\Big[\max_{i=1,2,\ldots} Q_i A_o(\chi_i) \le z_o, \forall o = 1, \ldots, N\Big] \\
&= \Pr\Big[Q_i \le \min_{o=1,\ldots,N} z_o / A_o(\chi_i), \forall i = 1, 2, \ldots\Big] \\
&= \Pr\Big[Q_i > \min_{o=1,\ldots,N} z_o / A_o(\chi_i), \text{for no } i = 1, 2, \ldots\Big] \\
&= \exp\Big[-\int_{\mathcal{X}} \int_{\min_{o=1,\ldots,N} z_o / A_o(\chi)}^{\infty} \theta q^{-\theta-1} dq d\mu(\chi)\Big] \\
&= \exp\Big[-\int_{\mathcal{X}} \max_{o=1,\ldots,N} A_o(\chi)^\theta z_o^{-\theta} d\mu(\chi)\Big].
\end{aligned}$$

This spectral representation provides us with an integral representation for the scale parameters and the correlation function. In particular, the marginal distribution of Z_o is Fréchet with scale $T_o \equiv \int_{\mathcal{X}} A_o(\chi)^\theta d\mu(\chi)$ and shape θ: $\Pr[Z_o \le z_o] = \lim_{z'_o \to \infty, \forall o' \neq o} e^{-\int_{\mathcal{X}} \max_{o=1,\ldots,N} A_o(\chi)^\theta z_o^{-\theta} d\mu(\chi)} = e^{-\int_{\mathcal{X}} A_o(\chi)^\theta d\mu(\chi) z_o^{-\theta}}$. Also, the joint distribution satisfies (1) for $G : \mathbb{R}_+^N \to R_+$ defined by

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

where $f_o(\chi) \equiv A_o(\chi)^\theta / T_o, \forall o = 1, \ldots, N$. This function is the correlation function of the max-stable multivariate Fréchet random vector.

We now use this representation to construct a sequence of CNCES correlation functions that converges uniformly to G. To do so, we first construct a sequence of auxiliary functions that are monotone increasing and converge pointwise to G.

For each $o = 1, \ldots, N$, since A_o is measurable, f_o is measurable and there exists a sequence of monotone increasing simple functions, $\{f_{no}\}_{n=1,2,\ldots}$, that converges pointwise to f_o. For each integer n, we construct an "auxiliary" function $F_n : \mathbb{R}_+^N \to \mathbb{R}_+$ as follows. Since f_{no} is simple for each o, there exists a common finite partition $\{\mathcal{X}_{kn}\}_{k=1}^{K_n}$ for each n such that $f_{no}(\chi) = \sum_{k=1}^{K_n} a_{kno} \mathbf{1}\{\chi \in \mathcal{X}_{kn}\}$ for some $\{a_{kno}\}_{k=1}^{K_n} \subset \mathbb{R}_+^{K_n}$. Set $F_n(x_1, \ldots, x_N) = \frac{n}{n+1} \sum_{k=1}^{K_n} \max_{o=1,\ldots,N} a_{kno} x_o \mu(\mathcal{X}_{kn})$. Note that we have $\int_{\mathcal{X}} \max_{o=1,\ldots,N} f_{no}(\chi) x_o d\mu(\chi) = \sum_{k=1}^{K_n} \max_{o=1,\ldots,N} a_{kno} x_o \mu(\mathcal{X}_{kn})$. By monotone convergence, $\{F_n\}_{n=1,2,\ldots}$ converges pointwise to G since $\{f_{no}\}_{n=1,2,\ldots}$ is monotone increasing and converges pointwise to f_o for each o:

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

$$= G(x_1,\ldots,x_N).$$

We now construct CNCES correlation functions that, up to a sequence of scaling constants, lie between each sequential pair of auxiliary functions and converge uniformly to G. For each n, choose a $\rho_n \in \left[\max\{0,\tilde{\rho}_n\},1\right)$ for $\tilde{\rho}_n \equiv 1 - \frac{\ln\frac{n^2+2n+1}{n^2+2n}}{\ln N} < 1$. Choose any $\rho_{kn} \in [\rho_n, 1)$ for $k = 1,\ldots,K_n$ and define $G_n:\mathbb{R}_+^N \to \mathbb{R}_+$ by $G_n(x_1,\ldots,x_N) \equiv \sum_{k=1}^{K_n}\left[\sum_{o=1}^{N}(\omega_{kno}x_o)^{\frac{1}{1-\rho_{kn}}}\right]^{1-\rho_{kn}}$ where $\omega_{kno} \equiv \delta_{no}^{-1}a_{kno}\mu(\mathcal{X}_{kn})$ for each $k = 1,\ldots,K_n$ with $\delta_{no} \equiv \sum_{k=1}^{K_n}a_{kno}\mu(\mathcal{X}_{kn}) \le 1$ for each $o = 1,\ldots,N$. Because $\sum_{k=1}^{K_n}\omega_{kno} = 1$ and $\rho_{kn} \in [0,1)$, G_n is a CNCES correlation function. Then

$$F_n(x_1,\ldots,x_N) = \frac{n}{n+1}\sum_{k=1}^{K_n}\max_{o=1,\ldots,N} a_{kno}x_o\mu(\mathcal{X}_{kn})$$

$$\le \frac{n}{n+1}G_n(\delta_{n1}x_1,\ldots,\delta_{nN}x_N)$$

$$\le \frac{n}{n+1}\sum_{k=1}^{K_n}N^{1-\rho_{kn}}\max_{o=1,\ldots,N} a_{kno}x_o\mu(\mathcal{X}_{kn})$$

$$\le \frac{n}{n+1}N^{1-\rho_n}\int_{\mathcal{X}}\max_{o=1,\ldots,N} a_{no}(\chi)x_o\,d\mu(\chi)$$

$$\le \frac{n^2+2n}{n^2+2n+1}N^{1-\rho_n}\frac{n+1}{n+2}\int_{\mathcal{X}}\max_{o=1,\ldots,N} a_{n+1,o}(\chi)x_o\,d\mu(\chi)$$

$$\le F_{n+1}(x_1,\ldots,x_N),$$

where the first and second inequalities use $\max_{o=1,\ldots,N}x_o \le \left[\sum_{o=1}^{N}x_o^{1/(1-\rho)}\right]^{1-\rho} \le N^{1-\rho}\max_{o=1,\ldots,N}x_o$ for any $\rho \in [0,1)$, and the last inequality uses $\frac{n^2+2n}{n^2+2n+1}N^{1-\rho_n} \le 1$ due to our choice of ρ_n. Define $\tilde{G}_n(x_1,\ldots,x_N) \equiv \frac{n}{n+1}G_n(\delta_{n1}x_1,\ldots,\delta_{nN}x_N)$. Then we have $F_n \le \tilde{G}_n \le F_{n+1} \le \tilde{G}_{n+1} \le G$. Since $F_n \to G$ pointwise, we also have $\tilde{G}_n \to G$ pointwise. Moreover, since (1) $\{\tilde{G}_n\}_{n=1,2,\ldots}$ is monotone increasing, (2) $\tilde{G}_n$ is continuous for each $n = 1, 2, \ldots$, and (3) G is continuous, we also have $\tilde{G}_n \to G$ uniformly on compact sets by Dini's theorem (Theorem 7.13 in Rudin 1964).

Finally, we show that the sequence of CNCES correlation functions converges uniformly on compacts sets to G. Fix any compact set $X \subset \mathbb{R}_+^N$. We have

$$\lim_{n\to\infty} \sup_{(x_1,\ldots,x_N)\in X} \left|G_n(x_1,\ldots,x_N) - G(x_1,\ldots,x_N)\right|$$

$$\leq \lim_{n\to\infty} \sup_{(x_1,\ldots,x_N)\in X} \left|G_n(x_1,\ldots,x_N) - \tilde{G}_n(x_1,\ldots,x_N)\right|$$

$$= \lim_{n\to\infty} \sup_{(x_1,\ldots,x_N)\in X} \left|\frac{n+1}{n}\tilde{G}_n\left(\delta_{1n}^{-1}x_1,\ldots,\delta_{Nn}^{-1}x_N\right) - \tilde{G}_n(x_1,\ldots,x_N)\right|$$

$$\leq \lim_{n\to\infty}\left|\frac{n+1}{n}\max_{o=1,\ldots,N}\delta_{no}^{-1} - 1\right| \lim_{n\to\infty} \sup_{(x_1,\ldots,x_N)\in X} \tilde{G}_n(x_1,\ldots,x_N)$$

$$= \lim_{n\to\infty}\frac{1}{n}\lim_{n\to\infty} \sup_{(x_1,\ldots,x_N)\in X} \tilde{G}_n(x_1,\ldots,x_N)$$

$$= \lim_{n\to\infty}\frac{1}{n} \sup_{(x_1,\ldots,x_N)\in X} G(x_1,\ldots,x_N) = 0,$$

where the first line uses the triangle inequality and $\tilde{G}_n \to G$ uniformly on compact sets, the second line uses the definition of $\tilde{G}_n$, the third line uses the fact that $\delta_{no} \leq 1$, $\forall o = 1,\ldots,N$, the fourth line uses $\lim_{n\to\infty}\delta_{no} = \lim_{n\to\infty}\sum_{k=1}^{K_n} a_{kno}\,\mu(\mathcal{X}_{kn}) = \lim_{n\to\infty}\int_{\mathcal{X}} a_{no}(\chi)d\mu(\chi) = \int_{\mathcal{X}} a_o(\chi)d\mu(\chi) = 1$, and the last line uses $\tilde{G}_n \to G$ uniformly on compact sets. Therefore, $G_n \to G$ uniformly on compact sets. ∎

Appendix B. Proof of Proposition 2

Since destination prices are given by (7), the price index in destination d is

$$P_d = \left[\int_0^1 \min_{o=1,\ldots,N}\left(W_o/Z_{od}(v)\right)^{1-\eta}dv\right]^{\frac{1}{1-\eta}}$$

$$= \left[E \max_{o=1,\ldots,N}\left(Z_{od}(v)/W_o\right)^{\eta-1}dv\right]^{\frac{1}{1-\eta}}$$

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

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

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 online Appendix Lemma O.2 and O.5.

The expenditure share of d on o is

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

using part 2 of online Appendix Lemma O.6 and the previous result for the price level. By part 1 of online Appendix Lemma O.6,

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

Appendix C. Derivation of the Gains from Trade in (17)

Using the within-factor component in (12), calculate

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

The denominator on the left-hand-side can be recovered from the between-factor component in (12),

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

Together, we have

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

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

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

The gains from trade relative to autarky are then

$$\frac{W_d/P_d}{W_d^A/P_d^A} = \left[\sum_{k=1}^{K} \left(\frac{\pi^*_{kdd}}{X_d}\right)^{1-\rho_k} \left(\sum_{o=1}^{N} \pi^*_{kod}\right)^{\rho_k}\right]^{-\frac{1}{\theta}}$$

$$= \pi_{dd}^{-1/\theta} \left[\sum_{k=1}^{K} \frac{\pi^*_{kdd}}{\pi_{dd}} \left(\sum_{o=1}^{N} \frac{\pi^*_{kod}}{\pi^*_{kdd}}\right)^{\rho_k}\right]^{-\frac{1}{\theta}}.$$

Further replacing $\pi^*_{kod} = \pi^W_{kod}\pi^B_{kd}$ and $\pi^*_{kdd} = \pi^W_{kdd}\pi^B_{kd}$, we get the expression in (17). ∎

REFERENCES

Adao, Rodrigo, Arnaud Costinot, and Dave Donaldson. 2017. "Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade." *American Economic Review* 107 (3): 633–89.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2015. "The Economics of Density: Evidence from the Berlin Wall." *Econometrica* 83 (6): 2127–89.

Alviarez, Vanessa. 2019. "Multinational Production and Comparative Advantage." *Journal of International Economics* 119 (C): 1–54.

Antràs, Pol, and Alonso de Gortari. 2020. "On the Geography of Global Value Chains." *Econometrica* 84 (4): 1553–98.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review* 102 (1): 94–130.

Bas, Maria, Thierry Mayer, and Mathias Thoenig. 2017. "From Micro to Macro: Demand, Supply, and Heterogeneity in the Trade Elasticity." *Journal of International Economics* 108: 1–19.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4): 841–90.

Brooks, Wyatt J., and Pau S. Pujolas. 2019. "Gains from Trade with Variable Trade Elasticities." *International Economic Review* 60 (4): 1619–46.

Bryan, Gharad, and Melanie Morten. 2019. "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia." *Journal of Political Economy* 127 (5): 2229–68.

Caliendo, Lorenzo, and Fernando Parro. 2015. "Estimates of the Trade and Welfare Effects of NAFTA." *Review of Economic Studies* 82 (1): 1–44.

Caron, Justin, Thibault Fally, and James R. Markusen. 2014. "International Trade Puzzles: A Solution Linking Production Factors and Demand." *Quarterly Journal of Economics* 129 (3): 1501–52.

Costa Tavares, Samia. 2006. "The Political Economy of the European Customs Classification." *Public Choice* 129 (1-2): 107–30.

Costinot, Arnaud, Dave Donaldson, and Ivana Komunjer. 2012. "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas." *Review of Economic Studies* 79 (2): 581–608.

Costinot, Arnaud, and Andrés Rodríguez-Clare. 2014. "Trade Theory with Numbers: Quantifying the Consequences of Globalization." In *Handbook of International Economics,* Vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 197–261. Amsterdam: Elsevier.

Dagsvik, John K. 1995. "How Large Is the Class of Generalized Extreme Value Random Utility Models?" *Journal of Mathematical Psychology* 39 (1): 90–98.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70 (5): 1741–79.

Eaton, Jonathan, and Samuel Kortum. 2012. "Putting Ricardo to Work." *Journal of Economic Perspectives* 26 (2): 65–90.

Fally, Thibault. 2015. "Structural Gravity and Fixed Effects." *Journal of International Economics* 97 (1): 76–85.

Feenstra, Robert C., Philip Luck, Maurice Obstfeld, and Katheryn N. Russ. 2017. "In Search of the Armington Elasticity." *Review of Economics and Statistics* 100 (1): 135–50.

Frankel, Todd. 2018. "Ford's Creative Efforts to Avoid $250 Million in 'Chicken Tax' Tariffs under Scrutiny." *Chicago Tribune*, July 9. https://www.chicagotribune.com/business/ct-biz-ford-tariff-chicken-tax-20180709-story.html.

French, Scott. 2016. "The Composition of Trade Flows and the Aggregate Effects of Trade Barriers." *Journal of International Economics* 98 (C): 114–37.

Fu, Xiao, Kejun Huang, Nicholas D. Sidiropoulos, and Wing-Kin Ma. 2019. "Nonnegative Matrix Factorization for Signal and Data Analytics." *IEEE Signal Processing Magazine* 36 (2): 59–80.

Gudendorf, Gordon, and Johan Segers. 2010. "Extreme-Value Copulas." In *Copula Theory and Its Applications*, edited by Piotr Jaworski, Fabrizio Durante, Wolfgang Karl Härdle, and Toasz Rychlik, 127–45. New York: Springer.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. 2019. "The Allocation of Talent and US Economic Growth." *Econometrica* 87 (5): 1439–74.

Irwin, Douglas. 2017. *Clashing over Commerce: A History of US Trade Policy*. Chicago: Chicago University Press.

Kabluchko, Zakhar. 2009. "Spectral Representations of Sum-and Max-Stable Processes." *Extremes* 12 (4): 401–24.

Lagakos, David, and Michael E. Waugh. 2013. "Selection, Agriculture, and Cross-Country Productivity Differences." *American Economic Review* 103 (2): 948–80.

Lashkari, Danial, and Marti Mestieri. 2016. "Gains from Trade with Heterogeneous Income and Price Elasticities." Unpublished.

Lashkaripour, Ahmad, and Volodymyr Lugovskyy. 2017. "Industry-Level Scale Economies: From Micro-Estimation to Macro-Implications." Unpublished.

Lee, Daniel D., and H. Sebastina Seung. 1999. "Learning the Parts of Objects by Non-negative Matrix Factorization." *Nature* 401 (6755): 788–91.

Lee, Daniel D., and H. Sebastian Seung. 2000. "Algorithms for Non-negative Matrix Factorization." In *Advances in Neural Information Processing Systems*, edited by Todd K. Leen, Thomas Dietterich, and Volker Tresp, 556–62. Cambridge, MA: MIT Press.

Levchenko, Andrei A., and Jing Zhang. 2014. "Ricardian Productivity Differences and the Gains from Trade." *European Economic Review* 65: 45–65.

Lind, Nelson, and Natalia Ramondo. 2021. "Global Innovation and Knowledge Diffusion." NBER Working Paper 29629.

Lind, Nelson, and Natalia Ramondo. 2023. "Replication Data for: Trade with Correlation." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E173601V1.

McFadden, Daniel. 1978. "Modeling the Choice of Residential Location." *Transportation Research Record:* 673.

McFadden, Daniel. 1981. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. Manski and Daniel McFadden, 198–272. Cambridge, MA: MIT Press.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2015. "Commuting, Migration and Local Employment Elasticities." *American Economic Review* 108 (12): 3855–90.

Ramondo, Natalia, and Andrés Rodríguez-Clare. 2013. "Trade, Multinational Production, and the Gains from Openness." *Journal of Political Economy* 121 (2): 273–322.

Ramondo, Natalia, Andrés Rodríguez-Clare, and Milagro Saborío-Rodríguez. 2016. "Trade, Domestic Frictions, and Scale Effects." *American Economic Review* 106 (10): 3159–84.

Redding, Stephen J. 2016. "Goods Trade, Factor Mobility and Welfare." *Journal of International Economics* 101: 148–67.

Ricardo, David. 1817. *On the Principles of Political Economy and Taxation*. 1st ed. London: John Murray.

Rudin, Walter. 1964. *Principles of Mathematical Analysis*, Vol. 3. New York: McGraw-Hill.

Samuelson, Paul A. 1954. "The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments." *Economic Journal* 64 (254): 264–89.

Santos Silva, João M. C., and Silvana Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics* 88 (4): 641–58.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press.