# Trade with Correlation*

Nelson Lind
Emory University

Natalia Ramondo
Boston University and NBER

September 18, 2021

**Abstract**

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 multi-sector 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 Codes: F1. Key Words: Ricardian model; generalized extreme value; Fréchet distribution; gains from trade; gravity.

*We thank our discussants Costas Arkolakis and Arnaud Costinot for their very helpful comments. We benefit from comments from Rodrigo Adao, Roy Allen, Dave Donaldson, Jonathan Eaton, Pablo Fajgelbaum, Teresa Fort, Stefania Garetto, and Fernando Parro, as well as from participants at various seminars and conferences. This research is supported by NSF grant #1919372. All errors are our own.

# 1 Introduction

Two hundred years ago, Ricardo (1817) proposed the idea that cross-country differences in production technologies can lead to gains from trade. Ricardo's work led to the following insight: Two countries gain more from trade when they have dissimilar production possibilities.

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.

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 multi-sector Ricardian trade model where these latent nests have an intuitive interpretation as technology classes used to produce goods classified in different sectors, allowing sectors to share technologies. Multi-sector models in the literature typically preclude sectors from sharing technologies, instead pairing each latent nest with an observed sectoral category. This independence assumption 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 2 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. However, while the independence assumption in EK entails that bilateral trade flows follow a CES structure, in Section 3, we show that our model 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 et al., 2012; Costinot and Rodrìguez-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, 2017), and domestic geography (Ramondo et al., 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 et al. (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. 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 4 presents the quantitative application of our framework to a multi-sector trade model. We use a cross-nested CES structure for productivity and treat each nest as a latent factor.[1] This *latent factor model* (LFM) allows for factors to be shared across sectors, and includes the case of independent productivity across sectors as the special case where each factor is unique to a single sector.

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.

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 (e.g. 2 digits SITC), corresponds to some fundamental aspect of the production process, such as a technology class. Instead, it re-groups observed sectors into fewer latent factors. Basically, while we treat latent factors as fundamental aspects of the production process, sectors are

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

categories designed to apply policies, such as import tariffs.[2]

Because LFM encompasses sectoral gravity as a special case — the case where factors are restricted to be specific to sectors — we can directly test the IIA restrictions implied by this assumption. We provide evidence suggesting that the sectoral gravity model is misspecified and that the LFM specification is consistent with correlation patterns observed in the expenditure data.

Our procedure to estimate latent factors is based on compressing the data from a high to a lower dimension, similar to principal-component analysis. To perform this compression, we use disaggregate sectoral expenditure and tariff data, and adapt techniques from the literature on non-negative matrix factorization (Lee and Seung, 1999, 2001; Fu et al., 2019) coupled with a pseudo Poisson maximum likelihood criterion for estimation (Silva and Tenreyro, 2006; Fally, 2015).[3]

Our LFM estimates show that seven latent factors are enough to explain almost 95 percent of the variation in 4-digit 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 heterogenous. These estimates, in turn, shape the amount of correlation across countries and sectors predicted by each model.

By estimating significant heterogeneity in correlation across sectors and countries, we show in Section 5 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 productiv-

[2] 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 vs. Citroen, The United States vs. Heartland By-products), starting with the case of U.S. import duties to sugar in 1881 (see Irwin, 2017), as well as a flurry of articles in the business press (e.g. ChicagoTribune, 2018).

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

ity across countries. The difference in gains between these two countries is only four 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., 2013; Bryan and Morten, 2018), or trade models used in the urban literature (Ahlfeldt et al., 2015; Monte et al., 2015).

Our paper is closely related to Adao et al. (2017), who provide sufficient conditions for non-parametric identification of invertible import demand systems using aggregate trade data. Their approach departs from CES demand, but does not necessarily lead to closed-form results. By focusing on the subclass of GEV import demand systems, we operationalize a 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 et al. (1995) procedure in Adao et al. (2017).[4]

Relatedly, papers such as Caron et al. (2014), Lashkari and Mestieri (2016), Brooks and Pujolas (2017), Feenstra et al. (2017), and Bas et al. (2017), among others, 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 multi-sector 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 heterogenous, suggesting that the correlation structure of this model captures economic forces that are distinct from input-output linkages.

[4] The mixed CES specification used in Adao et al. (2017) belongs to the GEV class, and it is obtained as the limiting case of the cross-nested CES specification we use when the number of nests goes to infinity (see Appendix C.2).

# 2 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 $X_d(v) = (\frac{P_d(v)}{P_d})^{1-\eta} X_d$, where $P_d(v)$ is the price of good $v$, $P_d = (\int_0^1 P_d(v)^{1-\eta} \mathrm{d}v)^{\frac{1}{1-\eta}}$ 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. The EK model, which is built on independent Fréchet random variables, gets its tractability from the property of max-stability. By applying the tools developed originally for random utility models (McFadden, 1978, 1981) to Ricardian models of trade, we are able to relax the independence assumption in EK, and get a flexible, yet tractable, model of trade in the generalized extreme value (GEV) class. Models in this class capture Ricardo's insight that the degree of technological similarity determines the gains from trade.

## 2.1 Max-Stable Multivariate Fréchet Productivity

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

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

where $T_{od}$ is the scale parameter and $\theta$ the shape parameter characterizing the marginal (Fréchet) distributions, $\mathbb{P}[Z_{od}(v) \le z] = e^{-T_{od} z^{-\theta}}$. 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] This function is homogeneous of degree one, ensuring the property of max-stability. Max-stability implies that the distribution of the maximum is Fréchet with shape $\theta$, 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.[7]

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

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

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

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

Because of the additive structure of $G^d$, in this special case, the shape parameter $\theta$ 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 $\theta$ alone governs the gains from trade in the EK model.

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

[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 $o$'th partial derivative with respect to $o$ distinct arguments non-negative if $o$ is odd and non-positive if $o$ is even.

[7] See Online Appendix O.1 for details on properties of Fréchet random variables and the representation of their joint distribution with correlation functions.

However, once we abandon the assumption of independent productivity, the strength of comparative advantage no longer solely depends on $\theta$. 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* $\theta$ 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,

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

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 $\theta$, it is now $\rho$ that determines the strength of comparative advantage across countries.

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

## 2.2 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 $\theta$, 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 $\omega_{kod}$ indicates the relative importance of each nest $k$ for a given trading pair $od$. If $\omega_{kod}$ 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 $\rho_k$. For $\rho_k = 0$, productivity is independent and the $k$'th nest is additive. In contrast, as $\rho_k \to 1$, productivity becomes perfectly correlated within nest $k$, and the $k$'th nest converges to a max function.

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

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

*Proof.* See Appendix A.1. □

This result ensures that focusing on CNCES — as we do in our quantitative application in Section 4 — 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 $\rho$). 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 heterogenous 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 3, 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 homogenous 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 multi-sector 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 homogenous correlation within sectors. When sectors share latent factors, within-factor correlation, captured by $\rho_k$, 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 $\theta$, and correlation function as in (5) with coefficient $\rho_k$.

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 $\theta$, 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]

[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 micro-foundation for the entire class of max-stable multivariate Fréchet distributions.

# 3 Expenditure, Prices, and Welfare

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

We first derive expenditure shares for the Ricardian model in Section 2. Next, we present properties of the GEV class and focus on the sub-class of CNCES. Finally, we characterize macro counterfactuals under GEV.

## 3.1 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*

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

*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, $G_o^d\left(x_1, \ldots, x_N\right) \equiv \partial G^d\left(x_1, \ldots, x_N\right)/\partial x_o$, and the price index in country $d$ is given by*

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

*Proof.* See Appendix A.2. □

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]

[9] Notice that 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*

These GEV import demand systems are uniquely characterized by the shape parameter $\theta$ and the correlation function $G^d$. 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,[11]

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

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

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

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 non-negative, 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 $\varepsilon_{ood} = -\theta - \sum_{o' \neq o} \varepsilon_{oo'd} < 0$.[12]

When the correlation function is additive, the cross-price elasticity in (11) is zero. That is, CES entails independence of irrelevant alternatives (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.

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,

---

*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, $\pi_{od} = \mathbb{E}\left[(P_d(v)/P_d)^{1-\eta} \mathbf{1}\{W_o/Z_{od}(v) = P_d(v)\}\right] = \mathbb{E}\left[(P_d(v)/P_d)^{1-\eta}\right] \mathbb{P}\left[W_o/Z_{od}(v) = P_d(v)\right] = \mathbb{P}\left[W_o/Z_{od}(v) = P_d(v)\right]$. This result does not rely on CES preferences (see Online Appendix O.5).

[11] The correlation function in (5) also leads to CES but with an elasticity equal to $\theta/(1-\rho)$.

[12] Max-stability requires that $G^d$ is homogenous of degree 1. Consequently, $G_o^d(x_1, \ldots, x_N)$ is homogenous of degree zero and $\sum_{o'=1}^N x_{o'} G_{oo'}^d(x_1, \ldots, x_N) = 0$. Since $\varepsilon_{ood} = -\theta - \theta \frac{P_{od}^{-\theta} G_{oo}^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta})}{G_o^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta})}$, then $\sum_{o'} \varepsilon_{oo'd} = -\theta$.

we interpret as expenditure on goods made with each latent technology:

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

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

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

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 $\rho_k$), 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 $\rho_k$). When $\rho_k = 0$ for all $k$, we are back to the CES case.

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.[13]

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

[13] 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 Appendix C.1 for the multi-sector model, and Online Appendix O.2 for the multinational production model, regional model, and global-value chain model.

multivariate Fréchet Ricardian model.[14] Despite their distinct micro-foundations, 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.

### 3.2 Macro Counterfactuals with GEV

We next show that heterogeneity in correlation leads to heterogeneity in the gains from trade.[15] 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

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

where $\widetilde{\pi}_{dd} \equiv \pi_{dd}/G_d^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}) = (P_{dd}/P_d)^{-\theta}$ reflects the real price of domestically produced goods 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:

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

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

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

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, $\widetilde{\pi}_{dd} = \pi_{dd}$: 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.

[14] 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).

[15] Online Appendix O.4 presents the model equilibrium formally and how to compute counterfactuals using exact hat-algebra methods.

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 B for derivations),

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

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 $\theta$ is the only parameter regulating the gains from trade. In contrast, for $\rho_k \to 1$ for all $k$, there are no gains from trade regardless of the value of $\theta$. 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. For intermediate values of $\rho_k$, both dispersion in productivity across goods, controlled by $\theta$, and correlation in productivity across countries, controlled by $\rho_k$, matter. Either higher $\theta$ or higher $\rho_k$ reduce the strength of comparative advantage and decrease the gains from trade. Summing up, the shape $\theta$, factor-level correlation parameters $\rho_k$, and factor-level expenditure shares combine to determine the gains from trade.

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 of $G^d(x_1, x_2, x_3) = \left( x_1^{1/(1-\rho)} + x_2^{1/(1-\rho)} \right)^{1-\rho} + x_3$. Countries 1 and 2 are technological peers, with the parameter $\rho$ 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:

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

The gains from trade for 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 technology between them. Conditional on expenditure, when $\rho = 0$, we get the ACR formula; for $\rho > 0$, correlation lowers the gains from trade; for $\rho \to 1$, the two countries are effectively a single country and the gains from trade depend on their combined self trade.

If we do not condition on expenditures, but rather solve for their equilibrium values, the intuition carries over. In an otherwise identical world, although wages equalize between countries one and two, heterogeneity in correlation precludes wage equalization with country 3. Specifically, countries 1 and 2 have lower gains from trade because they have correlated productivity.[16]

# 4 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 multi-sector 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 re-groups observed sectors into latent technology classes. In this way, we can avoid 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. Such a choice, common in the literature, is not innocuous because it implies 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.

## 4.1 Multi-Sector Model with Latent Factors

We use our results from the one-sector model in Section 2 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 $\eta > 1$.[17] 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. While the technology classes captured by the latent factors are fundamental aspects of the production process, sectors are not — they are categories designed to apply policies, such as import tariffs. With this interpretation of sectors, firms choose under which sectoral label to produce a good.

[16]The wage in country 3 is $W_3 = \left(1 + 2^{\frac{1+\theta-\rho}{1+\theta}}\right)^{1/\theta}$, while the wage in countries 1 and 2 is $W = 2^{-\frac{\rho}{1+\theta}} W_3$. Trade shares are: $\pi_{od} = 2^{-\rho} W^{-\theta} = 2^{-\frac{\rho}{1+\theta}} W_3^{-\theta}$, for $o = 1, 2$, and $\pi_{3d} = W_3^{-\theta}$. Wages are decreasing in the parameter $\rho$, while trade shares from country 3 increase with $\rho$, and trade shares from countries 1 and 2 decrease with $\rho$.

[17] In contrast to the literature, consumers have preferences over individual goods directly rather than over a composite sectoral good that aggregates individual goods. See Appendix C.1.

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 $T_{sod}$ 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

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

where $T^*_{ksod} \equiv \omega_{ksod} T_{sod}$. Within $k$, correlation is symmetric across origins and sectors and parameterized by $\rho_k$. 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 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

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

Here, $\pi^W_{ksod}$ is the within-factor share across sectors and origins, and $\pi^B_{kd}$ 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 $\rho_k$ and/or two sectors with similar within-factor expenditures — high $\pi^W_{ksod}\pi^W_{ks'o'd}$. When $\rho_k = 0$ for all $k$, $\varepsilon_{sos'o'd} = -\theta\mathbf{1}\{so = s'o'\}$. This is the CES case in (10).

The form of these cross-price elasticities motivates the reduced-form evidence we provide in the next section, which uses correlation in different dimensions of expenditure shares to construct indices of exposure to third-party tariffs.

However, for the purpose of estimation, this model is over-parameterized 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 under-identified, 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 *latent-factor model* (LFM) assumes that factor-level scale parameters are separable between sector-factor and factor-origin-destination components,

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

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 $A_{kod}$.

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 $\pi^*_{kod}$ through (relative) tariffs and the factor weights, $\lambda_{sk}$. 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 under-identified provided that the following rank condition is also satisfied:

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

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

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

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 $\sigma_k$.

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:

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

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 $\sigma_s$. Second, there is no cross-sector substitution since the elasticity in (20) collapses to

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

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 $\sigma_s$. The model, however, is still under-identified because $K = S$, and we need to estimate the parameters $A_{sod}$ and elasticities $\sigma_s$. 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 $\sigma_s$.

Although the SGM restrictions are convenient for estimation, they are a direct and testable consequence of assuming that factors are specific to sectors. We next present reduced-form evidence against a constant sectoral own-price elasticity and zero cross-sectoral elasticities. This evidence suggests that the SGM restrictions should be relaxed.

## 4.2 Reduced-Form Evidence

We estimate various specifications of a sectoral gravity-type equation and find evidence that the sectoral gravity model (SGM) is misspecified and that our latent-factor model (LFM) is consistent with correlation patterns observed in the expenditure data.

We use 14 aggregate sectoral categories from the World Input-Output Database (WIOD), denoted by $j$, and add a time subscript $t$ to estimate the following specification:

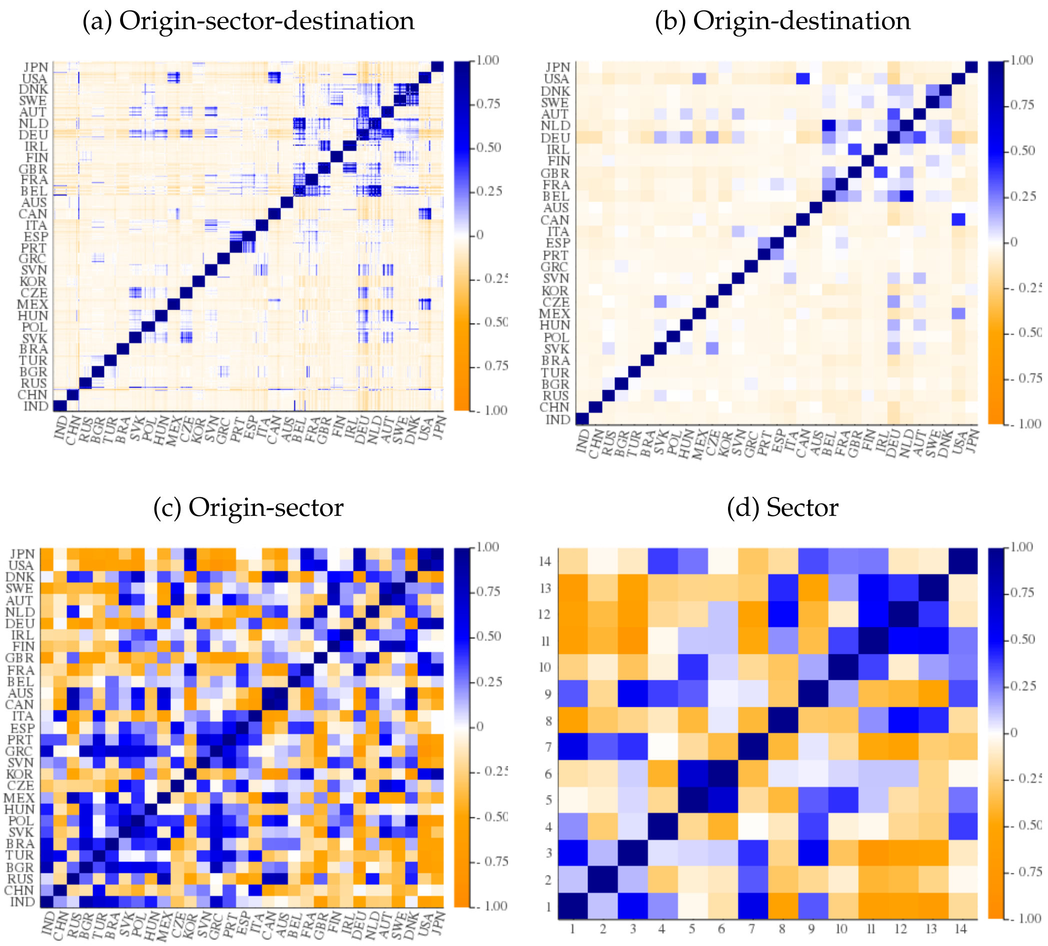
$$\pi_{jodt} \equiv \frac{X_{jodt}}{X_{dt}} = \exp\left[D^1_{jot} + D^2_{jdt} + D^3_{jod} + (\beta_j + \alpha' Geo_{od}) \ln t_{jodt} + \delta' I_{jodt}\right] \nu_{jodt}. \quad (28)$$

The variable $t_{jodt}$ is a tariff index for sector $j$.[18] $D^1_{jot}$, $D^2_{jdt}$ and $D^3_{jod}$ are sector-origin-time, sector-destination-time, and sector-origin-destination fixed effects, respectively. $Geo_{od}$ includes bilateral variables, such as geographical and income distance between the origin and destination. We include interactions of these variables with tariffs to allow the own-price elasticity to vary across origins within a sector. The variable $I_{jodt}$ includes three indices that capture potential departures from IIA.

To construct the indices, we use correlation patterns of expenditure observed in the data, shown in Figures 1b, 1c, and 1d. The figures use sector-origin-destination expenditure shares averaged over time, and sort sectors by WIOD classification code (see Appendix Table D.1) and countries by GDP per capita.

[18]This index is the result of aggregating 4-digit SITC tariffs into the WIOD sectoral categories, following a model-consistent procedure (see Appendix D.4).

Figure 1: Correlation Matrices for Expenditure Shares. WIOD sectoral data.




Notes: Each entry shows expenditure correlations: (1a) across destinations between a sector-origin pair; (1b) across destinations between two origins, $C_{oo'}^{\text{Origin - Geo}}$; (1c) across sectors between two origins, $C_{oo'}^{\text{Origin-Sectors}}$; and (1d) across origins between two sectors, $C_{jj'}^{\text{Sectors}}$. Axes are sorted by WIOD classification code for sectors and/or GDP per capita for countries. $j$ refers to a WIOD sectoral category. See Appendix Table D.1 for the WIOD sectoral classification.

Figure 1a is shown for comparison purposes and depicts correlation in expenditure shares across destinations, for each sector-origin pair. This figure shows that most correlation arises within a country across sectors, but we also observe correlation between origins across sectors.

Our first index of exposure to third-party tariffs exploits correlation in expenditure across destinations between two origins, after aggregating sectors at the origin-destination level (to remove correlation induced by sectoral export patterns).

$$I_{jodt}^{\text{Origin-Geo}} = \sum_{o' \neq o} C_{oo'}^{\text{Origin - Geo}} \ln t_{jo'dt}. \tag{29}$$

$C_{oo'}^{\text{Origin - Geo}}$ are the entries in Figure 1b and reveal correlation patterns that are highly geographic and related to income levels — in fact, "geography" explains 95 percent of the variation observed in Figure 1a. This index increases for country $o$ when tariffs rise in other countries with similar shares of destination expenditure.

Our second index uses correlation between two origins induced by their exporting sectors. In this case, we first average over destinations and compute sector-origin level expenditure relative to worldwide expenditure in the sector (so that correlation between two origins reflects similarity in comparative advantage).

$$I_{jodt}^{\text{Origin-Sector}} = \sum_{o' \neq o} C_{oo'}^{\text{Origin-Sector}} \ln t_{jo'dt}. \tag{30}$$

$C_{oo'}^{Origin-Sector}$ are the entries in Figure 1c, and show that correlation primarily arises between exporters of similar income. This index increases for country $o$ with tariffs in other countries with similar sectoral export shares.

Our final index is constructed based on correlation in expenditure between sectors within exporters. In this case, we use correlation in sector-origin level expenditure relative to the origin's total expenditure.

$$I_{jodt}^{\text{Sector}} = \sum_{j' \neq j} C_{jj'}^{\text{Sector}} \ln t_{j'odt}. \tag{31}$$

$C_{jj'}^{Sector}$ denotes the entries in Figure 1d, which are identical across countries. They show that, for instance, sectors related to more sophisticated manufacturing goods, such as "Electrical and Optical Equipment" (12) and "Transport Equipment" (13) are correlated with each other, as are sectors related to commodities, such as "Agriculture, Hunting, Forestry and Fishing" (1) and "Mining and Quarrying" (2). This index increases with tariffs on an origin in a different but correlated sector, and will allow us to detect patterns of cross-sector substitution within an origin country.

Table 1 presents PPML estimates of (28). If the SGM is correctly specified, we should find that $\alpha = \delta = 0$ in (28). If the CES model is correctly specified, we should further find that elasticities are the same across sectors.

We start in column 1 by restricting the sectoral elasticities to be common across sectors and exclude additional covariates. The coefficient on tariffs corresponds to a structural estimate of the ACR model where $\rho_k = 0$ for all $k$ in (19). In this case, $\alpha = \delta = 0$ and $\beta_j = -\theta = -2.63$ for all $j$.[19]

In columns 2 to 5, we allow for $\beta_j$ to be heterogenous across sectors. The estimates in column 2 correspond to structural estimates of the SGM where $\beta_j = -\sigma_j$, with

[19]This estimate is in the range estimated in the literature using sectoral data and the restriction to a uniform coefficient across sectors and countries (e.g. Boehm et al., 2021).

Table 1: Sectoral Gravity Model and Specification Tests. PPML.

| Dep. variable | $\pi_{jodt} \equiv X_{jodt}/X_{dt}$ | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $\beta$ | -2.63*** | | | | |
| | (0.221) | | | | |
| $\bar{\beta} = \sum_j^J \beta_j / J$ | | -2.70*** | -9.07*** | -2.49*** | -8.07*** |
| | | (0.233) | (1.676) | (0.261) | (1.679) |
| $\ln Dist_{od} \times \ln \bar{t}_{jodt}$ | | | 0.99*** | | 0.87** |
| | | | (0.293) | | (0.293) |
| $\|\ln Y_{ot} - \ln Y_{dt}\| \times \bar{t}_{jodt}$ | | | 1.40** | | 0.92* |
| | | | (0.442) | | (0.439) |
| $I_{jodt}^{\text{Origin-Geo}}$ | | | | 0.79** | 0.28 |
| | | | | (0.265) | (0.271) |
| $I_{jodt}^{\text{Origin-Sector}}$ | | | | -0.005 | -0.08 |
| | | | | (0.057) | (0.058) |
| $I_{jodt}^{\text{Sector}}$ | | | | 1.13*** | 0.79*** |
| | | | | (0.175) | (0.173) |
| $\|\ln Y_{ot} - \ln Y_{dt}\|$ | No | No | Yes | No | Yes |
| Deviance | 7.025 | 7.003 | 6.908 | 6.940 | 6.886 |
| Degrees of Freedom† | 7,814 | 7,827 | 7,830 | 7,830 | 7,833 |
| Null Hypothesis | | $\beta_j = \beta$ | $\alpha = 0$ | $\delta = 0$ | $\alpha = \delta = 0$ |
| $\chi^2$ | | 49.025 | 55.612 | 58.777 | 73.666 |
| Degrees of Freedom | | 13 | 2 | 3 | 5 |
| P-Value | | 0.0 | 0.0 | 0.0 | 0.0 |

Notes: Estimates of (28). Number of observations = 121,086. $j$ refers to a WIOD sectoral category. $Dist_{od}$ = distance between origin $o$ and destination $d$. $Y_{ot}$ = GDP per capita in $o$ at time $t$. $\bar{t}_{jodt}$ denotes $t_{jodt}$ relative to the sectoral mean. $I_{jodt}^{\text{Origin-Geo}}$, $I_{jodt}^{\text{Origin-Sector}}$, and $I_{jodt}^{\text{Sector}}$ are defined in (29), (30), and (31). All specifications include $j \times o \times t$, $j \times d \times t$, and $j \times o \times d$ fixed effects. For columns 2-5 the average tariff coefficient across sectors is reported, with estimates by sector from column 2 reported in Appendix Table D.1. †: Model's degrees of freedom. Last panel shows results of Wald tests for the null hypothesis that: sectoral elasticities are equal (column 2); and the tariff interactions as well as all indices are jointly insignificant (columns 3 to 5). Standard errors clustered at the sector-origin-destination level are in parenthesis, with levels of significance denoted by *** p < 0.001, and ** p < 0.01 and * p<0.05.

$\sigma_j \neq \sigma_{j'}$, for $j \neq j'$, and $\alpha = \delta = 0$. Not surprisingly, these estimates imply an average of 2.7, almost identical to the estimate in column 1.[20] The Wald test strongly rejects that elasticities are equal across sectors.

Columns 3-5 add tariff interactions and our indices.[21] Column 3 shows that both interactions are significant, suggesting that the own-price elasticity becomes more inelastic when geographical and income distance between an origin and destination increases. This column's Wald test strongly rejects the SGM prediction of a constant own-price elasticity within each sector.

Column 4 includes our three indices of exposure to third-party tariffs and directly tests the SGM prediction that IIA holds within each sector. While our indices of "geographic" ($I_{jodt}^{\text{Origin-Geo}}$) and cross-sector ($I_{jodt}^{\text{Sector}}$) exposure to third-party tariffs are positive and significant, the index of cross-origin sectoral exposure ($I_{jodt}^{\text{Origin-Sector}}$) is not. The insignificance of this index paired with the significance of the sectoral index suggests that departures from IIA operate through sectoral similarity within exporters rather than through similarity in sectoral comparative advantage between exporters. Note that, in column 5, $I_{jodt}^{\text{Origin-Geo}}$ is no longer significant after tariff interactions are also included, suggesting that this index indeed captures departures from IIA related to bilateral geographic factors. The Wald tests for both columns 4 and 5 strongly reject that these indices and interactions are jointly insignificant, providing evidence that the SGM is misspecified.

Summing up, Table 1 provides strong evidence against constant within-sector elasticities and zero cross-sector elasticities. Since the assumption that sectors correspond to latent factors implies both these restrictions, these results point toward relaxing the restrictions of the SGM.

The results in Table 1 also suggest that the LFM is on the right track. The insignificance of the index of cross-country sectoral exposure to third-party tariffs, together with the significance of the index of cross-sector exposure, suggests that departures from IIA are associated with patterns of cross-sector substitution within an origin rather than with within-sector cross-origin expenditure patterns — that is, it is reasonable to assume that latent factors are re-grouping sectors, not exporters. Moreover, because the index of cross-sector exposure to third-party tariffs is based on global correlation patterns across sectors, our findings suggest that it is reasonable to focus on a latent-factor structure where factor weights are common across countries, for each sector. This is precisely what our identifying assumption does

[20] The estimates by sector are reported in Appendix Table D.1 and are in the range of the sectoral elasticities estimated in Caliendo and Parro (2015).

[21] In this case, $t_{jodt}$ is deflated by the sector mean because the inclusion of sector-destination-time fixed effects absorbs that variation. The coefficients on the tariff interactions are interpreted relative to the sectoral average.

— countries load on sectors through common weights $\lambda_{sk}$.

We provide further reduced-form support for the LFM assumption in (21) by performing a principal-component analysis that predicts $\pi_{jodt}$ based on decompositions of the average expenditure share across destinations $d$. This analysis reveals that a structure where sectors load on (a few) origin-destination-time specific latent factors through common weights captures the data better than a structure where origins load on sectors-destination-time specific latent factors.[22]

Note that although principal-component analysis is a data-compression procedure, it 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 non-negative. To estimate the LFM, we not only need an alternative to sectoral gravity, but we also need an alternative to principal-component analysis. We present our estimation procedure next.

## 4.3 Latent-Factor Model Estimation

Our reduced-form evidence suggests that departures from IIA remain even when considering sectoral WIOD aggregate categories. The implication is that sectoral gravity regressions do not recover the true import demand system. We propose a tractable and flexible procedure based on our latent-factor model (LFM), which departs from sectoral gravity by relaxing the assumption that factors are specific to sectors. In this way, we allow for heterogenous and non-zero elasticities of substitution across countries and sectors. Like principal-component analysis, the LFM procedure entails compressing disaggregate sectoral data into a few latent factors.

We use disaggregate 4-digit SITC sectoral tariff and trade flow data from COMTRADE, combined with WIOD aggregate sectoral expenditure data, for 1999-2007 (see Appendix D for details). Our sample has 787 sectors and 31 countries. We denote the 4-digit SITC sectors by $s$, in contrast to the index $j$ used for the more aggregate WIOD sectoral data.

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, while factor-level expenditure and tariffs

[22] The principal-component structures are: $\hat{\pi}^{\text{OPC}}_{jodt} = \sum_{k=1}^{K} \lambda^{\text{OPC}}_{ok} \phi^{\text{OPC}}_{kjdt}$ (origins load on latent sector-destination specific factors); and $\hat{\pi}^{\text{SPC}}_{jodt} = \sum_{k=1}^{K} \lambda^{\text{SPC}}_{jk} \phi^{\text{SPC}}_{kodt}$ (sectors load on latent origin-destination specific factors). The factor weights $\lambda^{\text{OPC}}_{ok}$ are the right eigenvalues of the matrix of average expenditure across destinations, while $\lambda^{\text{SPC}}_{jk}$ are the left eigenvalues of that matrix. The first four eigenvectors explain 96.2 percent of the variation in average cross-destination expenditure. Given the factors weights, we solve for $\phi^{\text{OPC}}_{kjdt}$ and $\phi^{\text{SPC}}_{kodt}$. Their predicted values explain, respectively, 19.9 and 90.7 percent of the variation in $\pi_{jodt}$.

can vary over time. For convenience, we define $\phi^*_{kodt} \equiv (t^*_{kodt})^{\sigma_k}\pi^*_{kodt}$, and re-write (22) as

$$\pi_{sodt} = \sum_{k=1}^{K} t_{sodt}^{-\sigma_k}\lambda_{sk}\phi^*_{kodt}. \tag{32}$$

Observed sectoral shares are linear in the unobserved components $\lambda_{sk}$ and $\phi^*_{kodt}$, allowing us to build an estimation algorithm based on non-negative matrix factorization, combined with the pseudo-Poisson maximum likelihood (PPML) method used in the gravity literature (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

$$\hat{\Sigma}, \hat{\Lambda}, \hat{\Phi}^* = \arg\min_{\sigma\geq 0,\Lambda\geq 0,\Phi\geq 0} \sum_{s,o,d,t} \ell\left(\pi_{sodt}, \sum_{k=1}^{K} t_{sodt}^{-\sigma_k}\lambda_{sk}\phi^*_{kodt}\right), \tag{33}$$

where $\ell(x,\hat{x}) \equiv 2(x\ln(x/\hat{x}) - (x-\hat{x}))$. One convenient feature of PPML 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 $\pi_{odt}$.

To perform the data compression embedded in the LFM and solve (33), we adapt techniques from the literature on non-negative matrix factorization. Specifically, we extend the multiplicative updating algorithm in Lee and Seung (1999, 2001) to accommodate both missing data and simultaneous estimation of $\sigma_k$ (see Appendix E for details). To better see the connection to those procedures, suppose for a moment that there are no tariffs. Then, we can write (32) in matrix form as $\Pi = \Lambda\Phi^*$. The non-negative matrix factorization procedure decomposes the observed expenditure share matrix, $\Pi$, into two matrices: $\Lambda$ containing the sector-specific components and $\Phi^*$ 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.[23] However, the presence of non-negativity constraints in (33) 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). We present — and provide intuition for — sufficient conditions for identification of non-negative matrix factorization in Appendix E.1. In a nutshell, uniqueness is ensured when a large amount of data is used for the factorization.

The case of no tariffs also clarifies that tariffs are not used for the estimation of

[23]For any invertible matrix $R$ we have $\Pi = \tilde{\Lambda}\tilde{\Phi}^*$ for $\tilde{\Lambda} = \Lambda R^{-1}$ and $\tilde{\Phi}^* = R\Phi^*$.

latent-factor weights and expenditures. However, they are crucial cost shifters to estimate the within-factor elasticities $\sigma_k$. 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 $\sigma_s$. Identification of $\sigma_k$ comes from the conditional independence assumption embedded in our Poisson criterion: $\mathbb{E}[\upsilon_{sodt} \mid t_{sodt}, \Lambda, \Phi^*] = 0$ for $\upsilon_{sodt} \equiv \frac{\pi_{sodt}}{\sum_k t_{sodt}^{-\sigma_k} \lambda_{sk} \phi^*_{kodt}} - 1$. This assumption is analogous to the conditional independence assumption of PPML gravity estimation of tariff elasticities (e.g. equation 28).

We choose the number of latent factors by estimating (33) for $K = 1, 2, \dots$, and perform likelihood ratio tests until we fail to reject that the number of latent factors is $K$ versus the alternative of $K + 1$.[24] The rank condition in (24) indicates that we could fit as many as 432 latent factors.

Lastly, we need to estimate the shape parameter $\theta$ and the correlation function parameters $\rho_k$, for each $k$. Recall that $\sigma_k = \theta/(1 - \rho_k)$, so that given our estimates of factor-level elasticities, choosing a value for $\theta$ pins down each $\rho_k$. 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 $\theta$ that is consistent with our estimates of factor-level elasticities is $\min_{k=1,\dots,K} \sigma_k$. 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 $\theta$ increases.[25] For robustness, we implement an alternative estimation of $\theta$, 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 $\theta$ that are not statistically different from our baseline estimate (see Appendix G).

## 4.4 Results

We next analyze the results from the LFM estimation. When we show variables at a higher level of aggregation than 4-digit SITC level, we use the factor weights $\lambda_{sk}$

[24] The Poisson deviance function is homogenous of degree one and therefore its value depends on scaling of the data. The scaling does not impact the parameters' estimation, but it does matter for likelihood ratio tests. To address this scaling issue, we scale the Poisson deviance by the mean-variance ratio in the data.

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

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

Table 2: LFM Selection: Likelihood Ratio Test.

| Number of factors, $K$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | $14^r$ | 14 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $R^2$ 4-d SITC expenditure | 0.725 | 0.79 | 0.804 | 0.826 | 0.835 | 0.938 | 0.937 | 0.936 | 0.998 | 0.973 | |
| within $odt$ | 0.092 | 0.158 | 0.197 | 0.24 | 0.266 | 0.306 | 0.334 | 0.362 | 0.379 | 0.456 | |
| $R^2$ WIOD expenditure | 0.722 | 0.788 | 0.803 | 0.825 | 0.836 | 0.938 | 0.938 | 0.936 | 1.000 | 0.973 | |
| within $dt$ | 0.479 | 0.665 | 0.658 | 0.666 | 0.681 | 0.873 | 0.891 | 0.875 | 1.000 | 0.932 | |
| within $jdt$ | 0.849 | 0.885 | 0.901 | 0.912 | 0.920 | 0.957 | 0.955 | 0.955 | 1.000 | 0.971 | |
| within $odt$ | 0.221 | 0.382 | 0.458 | 0.521 | 0.614 | 0.657 | 0.693 | 0.673 | 1.000 | 0.787 | |
| Deviance | 377,451 | 333,999 | 310,594 | 292,161 | 278,379 | 266,955 | 256,823 | 248,288 | 260,822 | 210,554 | |
| Degrees of Freedom† | 9,436 | 18,872 | 28,308 | 37,744 | 47,180 | 56,616 | 66,052 | 75,488 | 121,873 | 132,104 | |
| Null Hypothesis | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | - | 7 | $14^r$ |
| $\chi^2$ | | 43,452 | 23,405 | 18,433 | 13,783 | 11,423 | 10,133 | 8,535 | - | 46,269 | 50,268 |
| Degrees of Freedom | | 9,436 | 9,436 | 9,436 | 9,436 | 9,436 | 9,436 | 9,436 | - | 66,052 | 10,231 |
| P-value | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 | - | 1.0 | 0.0 |

Notes: Results from estimating (33) with $K = 1, \ldots, 8; 14$. Number of observations = 5,528,764. $j$ refers to a WIOD sectoral category, while $s$ refers to a 4-digit SITC sector. $K = 14^r$ refers to a specification with 14 factors but factor weights at the WIOD-level restricted as in the sectoral gravity model (SGM). †: Model's degrees of freedom. Last panel shows likelihood ratio tests comparing specifications across columns.

to aggregate across 4-digit SITC sectors.

We estimate that the number of latent factors is $K = 7$. Table 2 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. The fit is also very high if we aggregate sectors at the WIOD level.[26] Additionally, a LFM model with $K = 14$ (i.e., the same number of WIOD sectors used in Table 1 for SGM) is not significantly different from $K = 7$. That said, although $K = 14$ is statistically indistinguishable from the LFM with $K = 7$, it is significantly different from LFM with $K = 14$ *plus* $\Lambda$ constrained to match the restrictions of SGM at the WIOD sector level, which we denote by $K = 14^r$.[27] This result provides further evidence that the SGM is misspecified. Indeed, it is notable that despite exactly fitting the data at the WIOD sectoral level (by construction) and using almost twice as many parameters, the deviance of $K = 14^r$ is higher than LFM with $K = 7$.

In Table 3, we go back to the same gravity-type regressions as in Table 1 adding the prediction for sectoral (WIOD-aggregate) expenditure from LFM. Are the tariff interactions and the indices capturing departures from IIA still significant? That is, does LFM capture the patterns in the data that SGM could not capture? Overall, LFM succeeds in capturing those patterns: Both tariff interactions as well as the

[26] Further linking the LFM estimates to the correlation matrices in Figure 1, and using $j$ for a WIOD sectoral aggregate, our estimated $\sum_k \hat{\lambda}_{jk}\hat{\lambda}_{j'k}$ and observed $\sum_{od} \pi_{jod}\pi_{j'od}$ have a correlation coefficient of 0.44, while for $\sum_k \hat{\phi}^*_{kod}\hat{\phi}^*_{ko'd'}$ and $\sum_j \pi_{jod}\pi_{jo'd'}$ the correlation is 0.81.

[27] Formally, let $j(s)$ be the WIOD sector that the 4-digit SITC sector $s$ belongs to. The restriction is that $\lambda_{sk} = 0$ if $j(s) \neq k$.

Table 3: Latent-Factor Model (LFM) and Specification Tests. PPML.

| Dep. variable | $\pi_{jodt} \equiv X_{jodt}/X_{dt}$ | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\ln \hat{\pi}^{\text{LFM}}_{jodt}$ | 1.01*** | | 0.982*** | 0.982*** | 0.981*** | 0.981*** |
| | (0.006) | | (0.01) | (0.01) | (0.01) | (0.01) |
| $\ln \hat{\pi}^{\text{LFM}}_{jodt} - \ln \hat{\pi}^{\text{U,LFM}}_{jodt}$ | | | | | | -0.526*** |
| | | | | | | (0.082) |
| $\ln Dist_{od} \times \ln \bar{t}_{jodt}$ | | 0.873** | 0.025 | | 0.021 | 0.004 |
| | | (0.293) | (0.131) | | (0.133) | (0.131) |
| $\lvert \ln Y_{ot} - \ln Y_{dt} \rvert \times \ln \bar{t}_{jodt}$ | | 0.915* | 0.3 | | 0.232 | 0.081 |
| | | (0.439) | (0.212) | | (0.216) | (0.216) |
| $I^{\text{Origin-Geo}}_{jodt}$ | | 0.277 | | -0.058 | -0.121 | -0.126 |
| | | (0.271) | | (0.143) | (0.157) | (0.153) |
| $I^{\text{Origin-Sector}}_{jodt}$ | | -0.084 | | -0.024 | -0.028 | -0.033 |
| | | (0.058) | | (0.032) | (0.033) | (0.032) |
| $I^{\text{Sector}}_{jodt}$ | | 0.793*** | | 0.288*** | 0.257** | 0.145 |
| | | (0.173) | | (0.082) | (0.084) | (0.086) |
| SGM Variables | No | Yes | Yes | Yes | Yes | Yes |
| $\lvert \ln Y_{ot} - \ln Y_{dt} \rvert$ | No | Yes | Yes | No | Yes | Yes |
| Deviance | 53.66 | 6.886 | 2.961 | 2.959 | 2.959 | 2.951 |
| Degrees of Freedom† | 2 | 7,833 | 7,831 | 7,831 | 7,834 | 7,835 |
| Null Hypothesis | LFM | $\alpha = \delta = 0$ | $\alpha = 0$ | $\delta = 0$ | $\alpha = \delta = 0$ | $\alpha = \delta = 0$ |
| $\chi^2$ | 2.757 | 73.666 | 4.34 | 12.342 | 13.111 | 3.97 |
| Degrees of Freedom | 1 | 5 | 2 | 3 | 5 | 5 |
| P-Value | 0.097 | 0.0 | 0.114 | 0.006 | 0.022 | 0.554 |

Notes: Estimates of (28) augmented by LFM predictions. Number of observations = 121,086. $j$ refers to a WIOD sectoral category. Column 2 corresponds to column 5 in Table 1. $\ln \hat{\pi}^{\text{LFM}}_{jodt}$ = LFM prediction for $\ln \pi_{jodt}$. $\ln \hat{\pi}^{\text{U,LFM}}_{jodt}$ is the prediction under uniform 4-digit SITC tariffs within each factor. $Dist_{od}$ = distance between $o$ and $d$. $Y_{ot}$ = GDP per capita in $o$ at time $t$. $\bar{t}_{jodt} = t_{jodt}$ relative to the sectoral mean. $I^{\text{Origin-Geo}}_{jodt}$, $I^{\text{Origin-Sector}}_{jodt}$, and $I^{\text{Sector}}_{jodt}$ are defined in (29), (30), and (31). SGM variables refers to sector-specific coefficients for log tariffs, and $j \times o \times t$, $j \times d \times t$, and $j \times o \times d$ fixed effects. †: Model's degrees of freedom. Last panel shows results of Wald tests for the null hypothesis that: the coefficient on $\ln \hat{\pi}^{\text{LFM}}_{jodt}$ is one (column 1), the tariff interactions and all indices are jointly insignificant (column 2 to 6). Standard errors clustered at the sector-origin-destination level are in parenthesis, with levels of significance denoted by *** p < 0.001, and ** p < 0.01 and * p<0.05.

origin-based indices are not significant, while the magnitude of the effect of the index capturing cross-sector correlation, $I^{\text{Sector}}_{jodt}$, is reduced more than three-fold (column 5). This index loses significance if we further control by the component of $\ln \hat{\pi}^{\text{LFM}}_{jodt}$ attributable to dispersion in 4-digit SITC tariffs within each factor. This means that, if anything, the LFM is predicting tariff effects that are too strong.

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

The first panel of Table 4 shows estimates of factor elasticities, $\sigma_k$. We rank the latent factors according to their elasticities $\sigma_k$ from largest ($F1$) to lowest ($F7$). Given 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. Notice that the average across $\sigma_k$ is 2.51, very close to the estimate of the reduced-form tariff elasticity in column 1 of Table 1. This should not be surprising since the LFM captures very well the correlation between tariffs and expenditure observed in the data.[28]

The second panel of Table 4 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 4-digit SITC sectors do not use technologies related to $F6$). Second, each factor is concentrated in a few 4-digit SITC sectors. The largest weight for each factor ranges from 0.045 to 0.281, with 90 percent of the weights below 0.003 for all factors. Since $\sum_s \lambda_{sk} = 1$, this indicates a very high level of sectoral concentration within each factor. Despite this concentration, the third panel of Table 4 shows that each factor has some weight on essentially every WIOD-aggregate sector.[29]

Figures 2a and 2b show that factors are not unique to sectors. Less than 15 percent of 4-digit SITC sectors use less than 4 factors, while about 75 percent use at least six out of the seven factors. Additionally, Figures 2c and 2d 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 $\lambda_{sk}$ to examine how factors load on sectors, and our estimates of $\pi^*_{kod}$ to get patterns of expenditure, export intensity, and domestic absorption by factor. First, the bottom panel of Table 4 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 5 reports the top-three 2-digit SITC sectors with the highest weights in each factor. For example, $F2$ is mainly used in the production of "Machinery

[28] Appendix Table F.1 shows the elasticity estimates for each LFM with $K = 1, \ldots, 8$. While the heterogeneity in estimates increases with $K$, the average across $\sigma_k$ remains around 2.5-3.

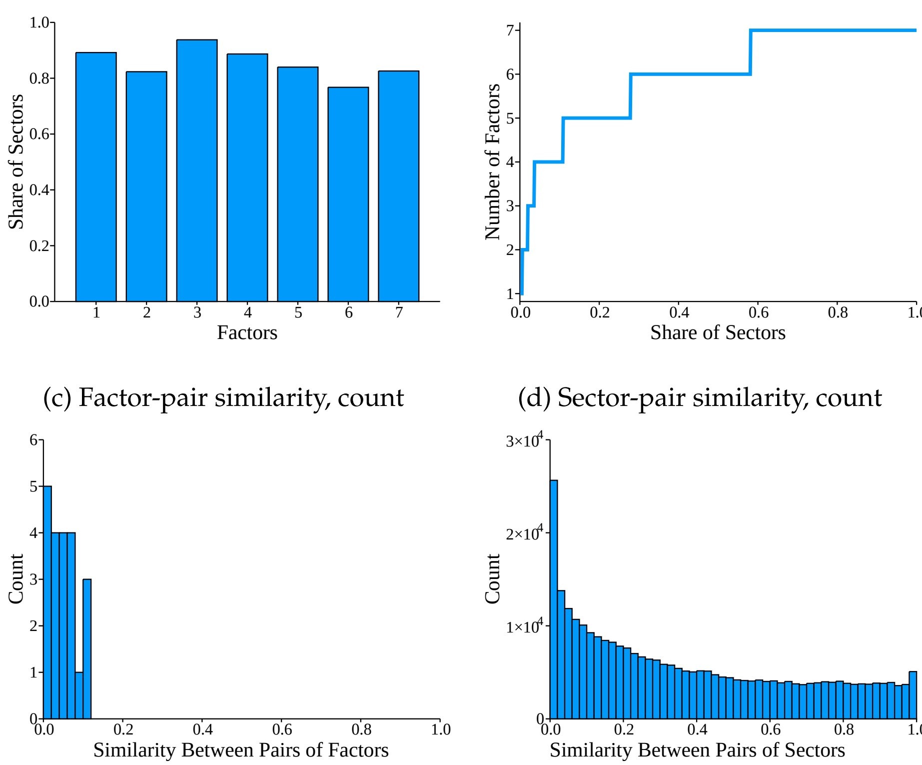
[29] For more details, see Online Appendix Figures O.1 and O.2.

Table 4: Estimates of latent-factor elasticities, weights, and expenditure. Summary.

| | Factor | | | | | | |
|---|---|---|---|---|---|---|---|
| | F1 | F2 | F3 | F4 | F5 | F6 | F7 |
| $\sigma_k$ | 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) |
| $\rho_k$ | 0.927 | 0.923 | 0.919 | 0.747 | 0.44 | 0.038 | 0.00 |
| | Factor Weights: 4-digit SITC Sectors | | | | | | |
| Zero Share | 0.108 | 0.177 | 0.062 | 0.113 | 0.16 | 0.233 | 0.174 |
| 90th Percentile | 0.003 | 0.002 | 0.003 | 0.003 | 0.002 | 0.001 | 0.001 |
| 99th Percentile | 0.016 | 0.011 | 0.013 | 0.019 | 0.029 | 0.026 | 0.018 |
| Maximum | 0.045 | 0.281 | 0.113 | 0.036 | 0.059 | 0.105 | 0.277 |
| | Factor Weights: 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 Non-Metallic 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 $\sigma_k$ are in parenthesis. $\theta = \min_k \sigma_{k=1,...,K} = 0.375$ with $\rho_k = 1 - \theta/\sigma_k$.

Figure 2: Factor Weights: Extensive and Intensive Margins.




(a) Share of sectors using each factor (b) Number of factors used by sectors

(c) Factor-pair similarity, count (d) Sector-pair similarity, count

Notes: Sectors are 4-digit SITC. Similarity refers to: in (2c), $\sum_s \lambda_{sk}\lambda_{sk'}/\sqrt{\sum_s \lambda_{sk}^2 \sum_s \lambda_{sk'}^2}$; and in (2d), $\sum_k \lambda_{sk}\lambda_{s'k}/\sqrt{\sum_k \lambda_{sk}^2 \sum_k \lambda_{s'k}^2}$.

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 4, 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.

The estimates of factor-level expenditure and elasticities shape the structure of the correlation function.[30] Next, we compare 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 the estimates of $\sigma_s$ coming from column 2 of Table 1 (reported in Appendix Table D.1), the WIOD sector-level data on expenditure, and the same estimate of $\theta$ as for LFM so that we ensure that differences

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

Table 5: 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 | Non-ferrous metals | 0.092 |

Notes: Factor weights from estimating (33) and aggregating to 2-digit SITC.

in results solely come from the correlation function.[31] For the CES model, we only need the elasticity estimated in column 1 of Table 1.

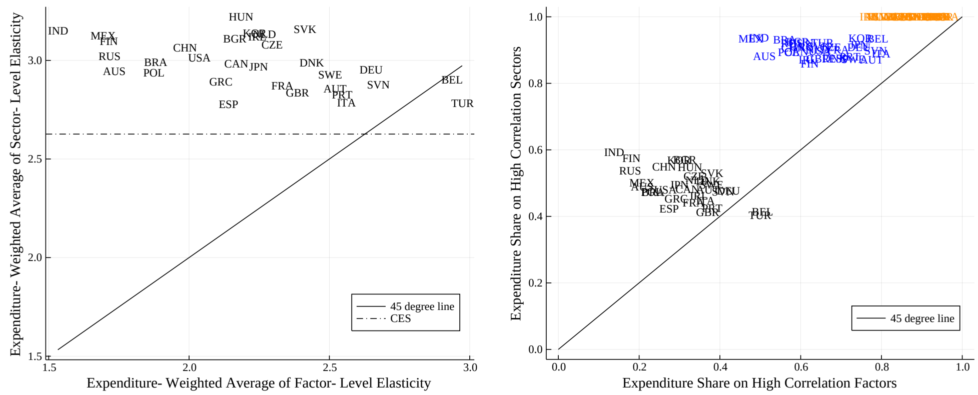
**Correlation patterns.** How much substitutability and correlation do our estimates imply? How do they compare with estimates from the SGM and CES model? These patterns are important for understanding the quantitative predictions of both models in counterfactual exercises.

First, we calculate averages of estimated factor-level elasticities $\sigma_k$, weighting by the the share of total expenditure in country $d$ on each factor $k$. For SGM, we calculate the averages of sector-level elasticities $\sigma_j$, weighting by the share of total expenditure in country $d$ on each WIOD sector $j$ (see Appendix Table D.1 for statistics on sectoral expenditure). While the unweighted average across these elasticities is around 2.5-2.7 in both models — matching the reduced-form average tariff elasticity of 2.6 in column 1 of Table 1 — Figure 3a shows that the expenditure-

[31] For the SGM, $\theta$ can be estimated with a two-step procedure, similar to the one in Appendix G for LFM, that uses cross-sectoral variation in expenditure and tariffs. Columns 4-6 in Appendix Table G.1 show that estimates are around 0.3, very close to our baseline estimate. This second step is not present in most of the SGM in the literature (e.g. Costinot and Rodrìguez-Clare, 2014) because the between-sector expenditure is constant — an assumption equivalent to assume that $\theta \to 0$. See Appendix C.1 and also Adao et al. (2017).

Figure 3: Elasticities, expenditure shares, and correlation: latent factors vs sectors.

(a) Expenditure-weighted avg elasticities (b) Expenditure shares and correlation




Notes: (3a): Average across factor-level (sector-level) elasticities $\sigma_k$ ($\sigma_j$) weighted by between-factor (between-sector) expenditure share in country $d$ $\pi^B_{kd}$ ($\pi^B_{jd}$). CES elasticity correspond to estimates in column 1 of Table 1. (3b): Share of country $d$'s total expenditure on factors (sectors) with correlation coefficient $\rho_k$ ($\rho_j$) higher than 0.4 (orange), 0.7 (blue), and 0.85 (black). $j$ denotes a WIOD sectoral category. Year 2007.

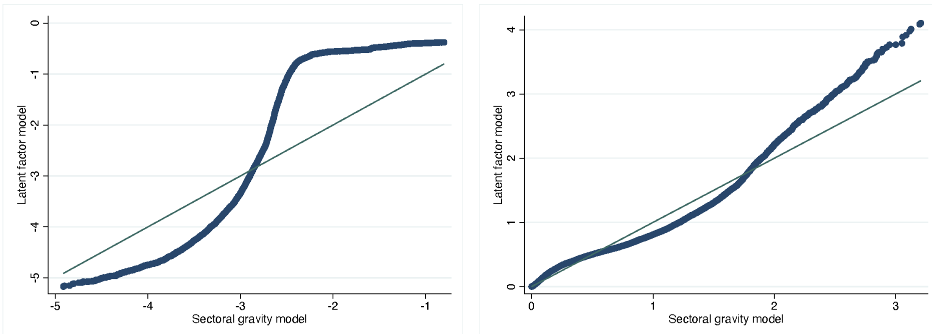
weighted averages are very different between models. SGM (and CES) estimates predict very similar (the same) average elasticities across countries, ranging from around 2.7 for Spain, Italy, and Turkey to around 3.2 for Hungary, the heterogeneity coming from countries on the lower end concentrating expenditure 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 $\rho_k = 1 - \theta/\sigma_k$ (subscript $j$ for SGM), Figure 3b 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.

**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

Figure 4: Expenditure Elasticities: LFM vs SGM.

(a) $\hat{\varepsilon}_{oo'dt}$ for $o = o'$ (b) $\hat{\varepsilon}_{oo'dt}$ for $o \neq o'$




Notes: Expenditure elasticities $\varepsilon_{oo'dt}$ calculated using (34) from latent-factor model (LFM) estimates, and sectoral gravity model (SGM) estimates. Quantiles of $\varepsilon_{oo'dt}$ for LFM vs SGM. Subscript $t$ refers to years 1999 to 2007.

elasticities

$$\varepsilon_{oo'd} = \sum_{s,s'} \frac{\pi_{sod}}{\pi_{od}} \varepsilon_{sos'o'd}, \tag{34}$$

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

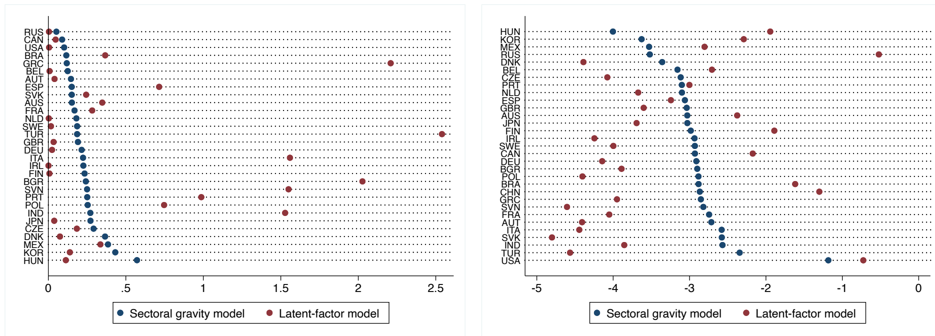
Both models deliver positive aggregate effects for $o \neq o'$. However, elasticities differ substantially between the two models, with the difference coming from the restriction of SGM to $\varepsilon_{sos'o'd} = 0$ for $s \neq s'$. Figure 4 shows that this restriction matters quantitatively. We use 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 4a. 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 3.

Figure 5 compares the values for elasticities for the LFM and the SGM focusing on expenditure by US consumers. Figure 5a zooms into China and its competitors in serving the United States, $\varepsilon_{o,CHN,USA}$, for $o \neq CHN$. Figure 5b 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, Netherland, 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 (-2.9 vs -1.3). In general, elasticities are much more similar in the SGM. Appendix Figure F.1 plots all the implied aggregate elasticities for the US

Figure 5: Expenditure Elasticities, US market: LFM vs SGM.

(a) $\hat{\varepsilon}_{o,CHN,USA}$, by competitor country $o$ (b) $\hat{\varepsilon}_{o',o',USA}$, by origin country $o'$




Notes: Estimates of expenditure elasticities $\varepsilon_{o,o',USA}$ calculated using (34) and estimates from the latent-factor model (LFM) and sectoral gravity model (SGM). Figure 5a plots $o \neq o'$ when $o' = CHN$. Figure 5b plots $o = o'$ for each country $o'$ in our sample. Year 2007.

market, providing a comprehensive visualization of the differences between the two models. For instance, the LFM implies lower substitution than SGM for goods produced in the United States for the US market, and more competition (higher substitutability) for goods from competing sources into the United States, in particular China.

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

# 5 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.

## 5.1 The Gains from Trade

Figure 6 shows the gains from trade against self-trade shares, using the estimated versions of the latent-factor model (LFM), the sectoral gravity model (SGM), and the CES model. Appendix Table F.2 reports the exact numbers.

For LFM, we use our estimates in Section 4.4 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 WIOD sector $j$. In this case, quantifying the gains from trade requires estimates of sectoral elasticities, which we take from the SGM estimates coming from column 2 of Table 1, data on sectoral expenditure shares, and an estimate of $\theta$. To ensure that differences in gains between the models are solely due to differences in correlation, we use the LFM estimate of 0.375 for both the SGM and the LFM — direct estimates of $\theta$ within the SGM are close to this value (see Footnote 31). 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 this case, we use the estimate in column 1 of Table 1, and set $\theta = 2.65$, which corresponds to the trade elasticity under the independence restrictions that lead to CES.

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 gains are almost 90 percent higher because it is less correlated with the rest of the world. Examining the patterns in Table 4 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 one to four percent).

With respect to LFM, the quantitative version of SGM also deliver lower gains, including for the large countries in our sample. Given the estimates shown in Section 4.4, 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 6 further explores the sources of the quantitative differences in gains between the two models by decomposing the gains in (16). 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 $\sigma_k$ ($\sigma_j$ for SGM). Finally, the white dots show gains if we remove 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 vs within-sector expenditure shares that explains the quantitative difference in gains between the two models.

The large quantitative differences in the gains from trade implied by the different models indicate that the way correlation in productivity is introduced matters. 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. Appendix C.1 presents the model in detail.

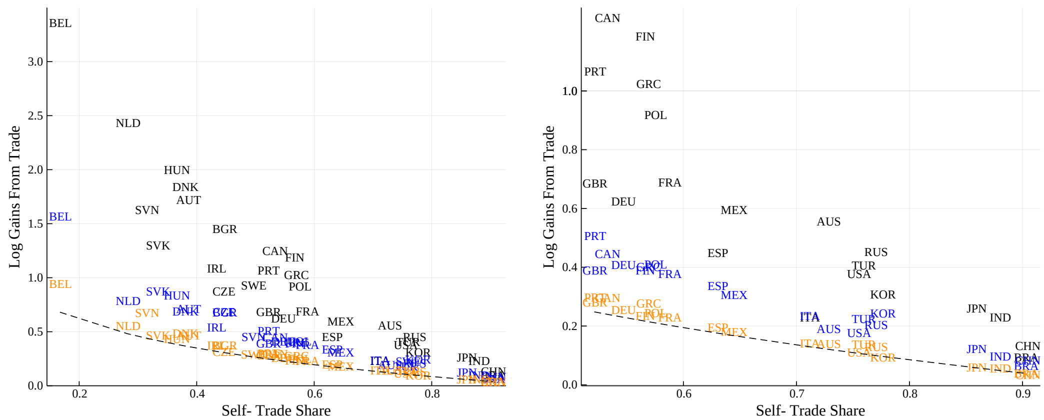
Following the literature (e.g. Costinot and Rodrìguez-Clare, 2014), we introduce these linkages assuming that each sector $s$ and country $o$ has a Cobb-Douglas production function that combines labor, with share $1 - \alpha_{so} \in [0, 1]$, and a composite input from each sector, with shares $\alpha_{ss'o} \in [0, 1]$ 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 estimates of $\sigma_s$ in column 2 of Table 1 do not change. The variable $c_{so}$, which captures the impact of input-output linkages on unit costs, is simply absorbed into the sector-origin fixed effect.

Figure 6: The Gains From Trade

A. Comparison of the gains from trade across estimated models.

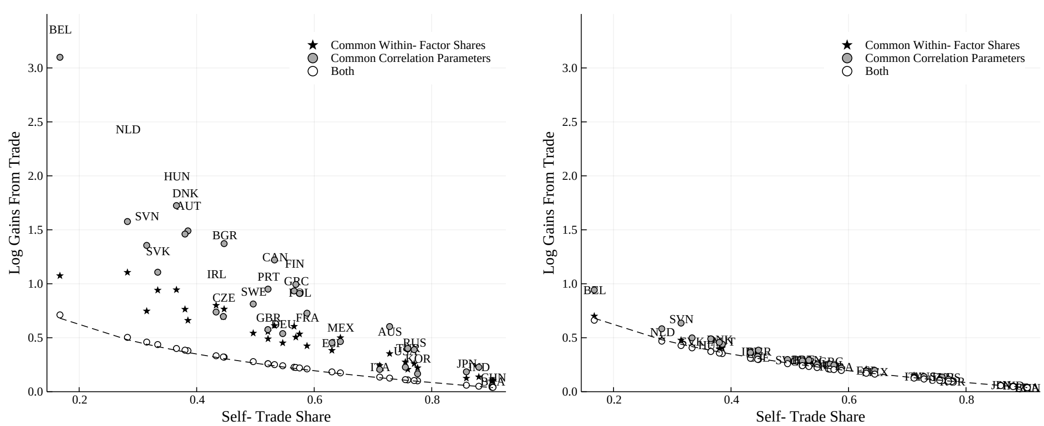
(a) All countries | (b) Large countries zoomed-in




Notes: Gains from trade = Real wages in the observed equilibrium relative to autarky real wages. Calculations using estimates from latent-factor model (LFM, black dots), sectoral gravity model (SGM, orange dots), SGM augmented by input-output links (blue dots), and CES model (dash line). Year 2007.

B. Decomposition of the gains from trade.

(c) LFM | (d) SGM




Notes: Gains from trade calculated using equation (16). Stars removed heterogeneity in within-$h$ self-trade shares. Gray dots removed heterogeneity in correlation coefficients. White dots remove both sources of heterogeneity. Here, $h$ refers to factor $k$ for the latent-factor model (LFM), and WIOD sector $j$ for the sectoral gravity model (SGM).

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

$$\frac{W_d/P_d}{W_d^A/P_d^A} = \left[\sum_{s=1}^{S} \left(\prod_{s'=1}^{S} (\pi^W_{s'dd})^{-\frac{a_{ss'd}}{\sigma^m_{s'}}}\right)^{-\theta} \pi^B_{sd}\right]^{-\frac{1}{\theta}}, \tag{35}$$

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

The blue dots in Figure 6 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 heterogenous — 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.[32]

## 5.2 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

$$\frac{\mathrm{d}\ln W_d/P_d}{\mathrm{d}\ln t_{o'd}} = \underbrace{(1-\pi_{dd})\frac{\mathrm{d}\ln W_d/W_{o'}}{\mathrm{d}\ln t_{o'd}}}_{\text{Domestic Wage Effect}} - \underbrace{\sum_{o\neq o' \text{ and } o\neq d} \pi_{od}\frac{\mathrm{d}\ln W_o/W_{o'}}{\mathrm{d}\ln t_{o'd}}}_{\text{Third Party Effect}} - \underbrace{\pi_{o'd}}_{\text{Direct Tariff Effect}}, \tag{36}$$

where $t_{o'd} \equiv [\sum_{k=1}^{K}(t^*_{ko'd})^{-\theta}]^{-\frac{1}{\theta}}$. 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 7 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 (36) from $0$ to $\Delta t$ where $\Delta t$ 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 of rising domestic wages is smaller, while the cumulative (negative) effect of rising third-party wages is larger for LFM. This is because US consumers substitute less towards their own goods and more towards third parties in the LFM (second panel of Figure 7). Additionally, the cumulative direct effect of higher tariffs is

[32] An additional piece of suggestive evidence is that the correlation between the entries of the cross-sectoral correlation matrix in Figure 1, $C^{Sector}_{jj'}$, and the input-output direct requirement coefficients from $j$ to $j'$ is only 0.2. This is an indication that those linkages do not account for most of the cross-sectoral correlation observed in the data.

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.

Appendix Figure F.2 and Appendix Table F.3 show moments of the elasticity in (36) 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.

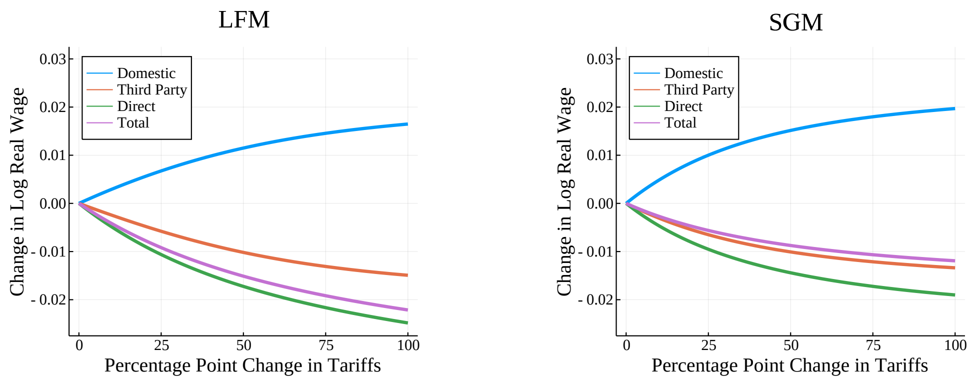
The difference in substitution patterns between the two models comes from differences in expenditure shares across latent factors, which correspond to sectors in the SGM. The bottom panel of Figure 7 shows that, for all factors, US expenditure shifts away from China when tariffs rise. However, it does 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.

# 6 Conclusions

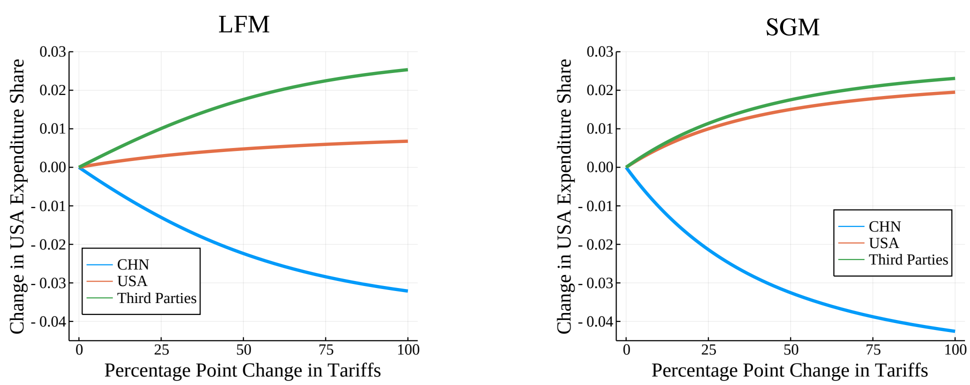
This paper is motivated by the old Ricardian idea that a country gains from trading with those countries who are technologically dissimilar. We develop a Ricardian model of trade that allows for rich patterns of correlation in technology between countries, retains all the tractability of EK-type tools, and spans the entire class of GEV import demand systems. We propose a cross-nested CES structure for correlation that departs from the existing models by treating the nests as unobserved "latent" dimensions of the data, and allows us to relax commonly made distributional assumptions. In the context of a multi-sector 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

Figure 7: Effect of increases in US tariffs on China: LFM vs SGM.

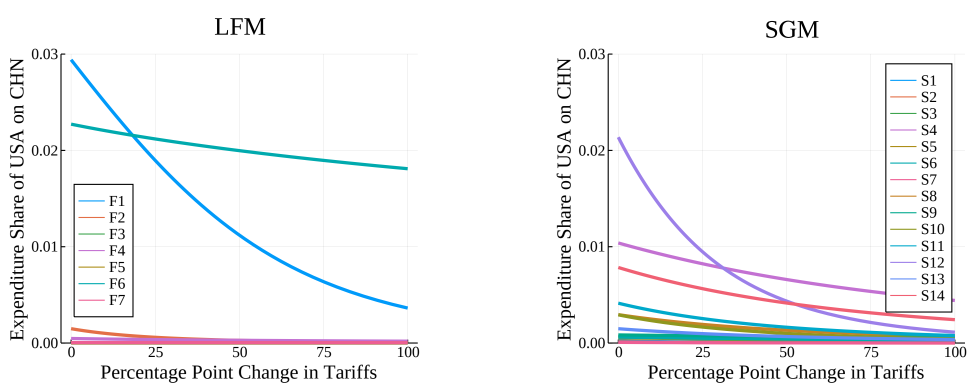
A. US Real Wage.




B. US Expenditure Shares.




C. US Factor-level Expenditure Shares from China.




Notes: LFM = Latent Factor Model; SGM = Sectoral Gravity Model. Changes in log US real wage are decomposed using (36). Year 2007.

correlation patterns. The implied gains from trade are much more heterogeneous across countries than from estimates of models that restrict correlation patterns.

# References


Adao, R., A. Costinot, and D. Donaldson (2017). Nonparametric counterfactual predictions in neoclassical models of international trade. *The American Economic Review 107*(3), 633–689.

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

Alviarez, V. (2019). Multinational production and comparative advantage. *Journal of International Economics 119*(C), 1–54.

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

Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2018). Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications. *Econometrica 86*(2), 685–720.

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

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

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

Boehm, C. E., A. Levchenko, and N. Pandalai-Nayar (2021). The long and short (run) of trade elasticities. *NBER working paper 27064*.

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

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

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

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

ChicagoTribune (09-07-2018). Ford's creative efforts to avoid $ 250 million in 'chicken tax' tariffs under scrutiny. *Frankel Todd*.

Costa Tavares, S. (2006). The political economy of the european customs classification. *Public Choice 129*, 107–130.

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

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

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

Ding, C., T. Li, W. Peng, and H. Park (2006). Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 126–135.

Donoho, D. and V. Stodden (2004). When does non-negative matrix factorization give a correct decomposition into parts? In *17th Annual Conference on Neural Information Processing Systems, NIPS 2003*. Neural information processing systems foundation.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica*, 1741–1779.

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

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

Faust, J. (1998). The robustness of identified var conclusions about money. In *Carnegie-Rochester conference series on public policy*, Volume 49, pp. 207–244. Elsevier.

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

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

Fry, R. and A. Pagan (2011). Sign restrictions in structural vector autoregressions: A critical review. *Journal of Economic Literature 49*(4), 938–60.

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

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

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

Huang, K., N. D. Sidiropoulos, and A. Swami (2014). Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition. *IEEE TRANSACTIONS ON SIGNAL PROCESSING 62*(1), 211.

Irwin, D. (2017). *Clashing over Commerce: A History of U.S. Trade Policy*. Chicago University Press.

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

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

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

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

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

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

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

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

Lind, N. and N. Ramondo (2021). Global innovation and knowledge diffusion. Mimeo.

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

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

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

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

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

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

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

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

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

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

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

Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics 52*(2), 381–419.

# A Proofs

## A.1 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, $v$.

Let $\{Z_o\}_{o=1}^N$ be distributed max-stable multivariate Frechet. Then by Theorem 1 in Kabluchko (2009), $\{Z_o\}_{o=1}^N$ has a spectral representation—there exists a $\sigma$-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

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

where the fourth equality follows from Campbell's theorem (see Kingman, 1992). 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 $\theta$: $\mathbb{P}[Z_o \le z_o] = \lim_{z'_o \to \infty \forall o' \neq o} e^{-\int_{\mathcal{X}} \max_{o=1,\ldots,N} A_o(\chi)^\theta z_o^{-\theta} d\mu(\chi)} = e^{-\int_{\mathcal{X}} A_o(\chi)^\theta d\mu(\chi) z_o^{-\theta}}$. 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 point-wise 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 point-wise to $f_o$. For each integer $n$, we construct an "auxiliary" function $F_n : \mathbb{R}_+^N \to \mathbb{R}_+$ as follows. First, define $R_n \equiv \cup_{o=1}^N \{f_{no}(\chi) \mid \chi \in \mathcal{X}\}$. Since $f_{no}$ is simple for each $o$, $R_n$ is finite. Next, let $\{\tilde{a}_{kn}\}_{k=1}^{K_n}$ be an enumeration of $R_n$ and for each $k = 1, \ldots, K_n$ define $\mathcal{X}_{kn} \equiv \{\chi \in \mathcal{X} \mid$

$\tilde{a}_{kn} = \max_{o=1,\ldots,N} f_{no}(\chi)\}$. Then for each $o$ we have $f_{no}(\chi) = \sum_{k=1}^{K_n} a_{kno}\mathbf{1}\{\mathcal{X}_{kn}\}$ for some $\{a_{kno}\}_{k=1}^{K_n} \subset \mathbb{R}_+^{K_n}$. Finally, 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 $\int_{\mathcal{X}} \max_{o=1,\ldots,N} f_{no}(\chi)x_o \mathrm{d}\mu(\chi) = \sum_{k=1}^{K_n} \max_{o=1,\ldots,N} a_{kno}x_o\mu(\mathcal{X}_{kn})$. 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 $o$:

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

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 $n$, chose a $\rho_n \in [\max\{0,\tilde{\rho}_n\},1)$ 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

$$\begin{aligned}
F_n(x_1,\ldots,x_N) &= \frac{n}{n+1}\sum_{k=1}^{K_n}\max_{o=1,\ldots,N} a_{kno}x_o\mu(\mathcal{X}_{kn}) \le \frac{n}{n+1}G_n(\delta_{n1}x_1,\ldots,\delta_{nN}x_N)\\
&\le \frac{n}{n+1}\sum_{k=1}^{K_n}N^{1-\rho_{kn}}\max_{o=1,\ldots,N} a_{kno}x_o\mu(\mathcal{X}_{kn}) \le \frac{n}{n+1}N^{1-\rho_n}\int_{\mathcal{X}}\max_{o=1,\ldots,N} a_{no}(\chi)x_o\mathrm{d}\mu(\chi)\\
&\le \frac{n^2+2n}{n^2+2n+1}N^{1-\rho_n}\frac{n+1}{n+2}\int_{\mathcal{X}}\max_{o=1,\ldots,N} a_{n+1,o}(\chi)x_o\mathrm{d}\mu(\chi) \le F_{n+1}(x_1,\ldots,x_N)
\end{aligned}$$

where the first and second inequalities use $\max_{o=1,\ldots,N} x_o \le (\sum_{o=1}^N x_o^{1/(1-\rho)})^{1-\rho} \le N^{1-\rho}\max_{o=1,\ldots,N} x_o$ 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 $\rho_n$. 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 $F_n \to G$ point-wise, we also have $\tilde{G}_n \to G$ point-wise. 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 et al., 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

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

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 $\delta_{no} \le 1\ \forall o = 1,\ldots,N$, the fourth line uses $\lim_{n\to\infty}\delta_{no} = \lim_{n\to\infty}\sum_{k=1}^{K_n} a_{kno}\mu(\mathcal{X}_{kn}) = \lim_{n\to\infty}\int_{\mathcal{X}} a_{no}(\chi)\mathrm{d}\mu(\chi) = \int_{\mathcal{X}} a_o(\chi)\mathrm{d}\mu(\chi) = 1$, and the last line uses $\tilde{G}_n \to G$ uniformly on compact sets. Therefore, $G_n \to G$ uniformly on compact sets.

$\square$

## A.2 Proof of Proposition 2

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

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

where $P_{od} \equiv \gamma T_{od}^{-1/\theta} W_o$, $\gamma = \Gamma\left(\frac{\theta+1-\eta}{\theta}\right)^{\frac{1}{1-\eta}}$, due to Online Appendix Lemma O.5 and Online Appendix Lemma O.2.

The expenditure share of $d$ on $o$ is

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

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

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

# B Derivation of the Gains from Trade in (17)

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

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

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

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

The gains from trade relative to autarky are then

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

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).

# C Models in the GEV Class

## C.1 Many Sectors

Assume that each country is composed of multiple sectors, $s = 1, \ldots, S$, each composed of a continuum of goods. As in Caliendo and Parro (2015), assume that productivity for good $v$ in sector $s$ is a random draw distributed independent Fréchet within each sector across origins, with sector-specific shape $\epsilon_s$ and scale $\widetilde{A}_{so}$. As in French (2016), further assume that consumers in each destination $d$ have CES preferences over sectoral aggregates with elasticity $\bar{\theta} > 0$. The sectoral composite good aggregates goods CES with elasticity $\eta_s$, where $\eta_s - 1 > \epsilon_s$. Given trade costs $\tau_{sod}$, the share of destination $d$'s expenditure on goods from origin $o$ and sector $s$ is

$$\pi_{sod} = \left( \frac{\tau_{sod}\frac{W_o}{A_{so}}}{P_{sd}} \right)^{-\epsilon_s} \left( \frac{P_{sd}}{P_d} \right)^{-\bar{\theta}}, \tag{C.1}$$

where $A_{so} \equiv \widetilde{A}_{so}^{1/\epsilon_s}$, $P_{sd}^{-\epsilon_s} \equiv \sum_{o'=1}^{N} \left( A_{so'}\tau_{so'd}\frac{W'_o}{A_{so'}} \right)^{-\epsilon_s}$, and $P_d^{-\bar{\theta}} \equiv \sum_s P_{sd}^{-\bar{\theta}}$.

This multi-sector model is isomorphic to a model where consumers have CES preferences over a continuum of goods and productivity is correlated within each sector across origins, as the sectoral gravity model presented in Section 4.1. Suppose that productivity for good $v$ in sector $s$ is a random vector drawn from a multivariate max-stable Fréchet distribution with scale parameter $T_{sod}$, shape $\theta$, and sector-level correlation function

$$G^{sd}(x_1, \ldots, x_N) = \left( \sum_{o=1}^{N} x_o^{1/(1-\rho_s)} \right)^{1-\rho_s}, \tag{C.2}$$

where $\rho_s$ measures the degree of correlation across origin countries in each sector. Sectoral expenditure shares are

$$\pi_{sod} = \left( \frac{T_{sod}^{-1/\theta} W_o}{P_{sd}} \right)^{-\frac{\theta}{1-\rho_s}} \left( \frac{P_{sd}}{P_d} \right)^{-\theta}, \tag{C.3}$$

where $P_{sd}^{-\frac{\theta}{1-\rho_s}} \equiv \sum_{o=1}^{N} (T_{sod}^{-1/\theta} W_o)^{-\frac{\theta}{1-\rho_s}}$, and $P_d^{-\theta} = \sum_s P_{sd}^{-\theta}$. This import demand system matches (C.1) for $T_{sod}^{1/\theta} = \tau_{sod}/A_{so}$, $\theta/(1-\rho_s) = \epsilon_s$, and $\theta = \bar{\theta}$. The first term on the right-hand side of (C.3)—and (C.1)—is expenditure within sector $s$ and is CES with elasticity $\theta/(1-\rho_s)$ — $\epsilon_s$ in (C.1). The second term refers to between-sector expenditure and is also CES with elasticity $\theta — \bar{\theta}$ in (C.1). If we further restrict $\theta \to 0$, the between-sector expenditure share become a constant – this is the case of a between-sector Cobb-Douglas aggregator.

**Multi-sector model with input-output linkages.** Assume that each sector $s$ combines domestic labor and a domestic aggregate input to produce sectoral tradable good $v$. The production function is Cobb-Douglas with $1-\alpha_{so} \in [0,1]$ the labor share in sector $s$ and country $o$. The aggregate input used by sector $s$ combines the composite sectoral good of each sector according to $\prod_{s'} M_{s'o}^{\alpha_{ss'o}}$, with $\sum_{s'} \alpha_{ss'o} = \alpha_{so}$. In turn, $M_{so}$ is a CES aggregator of the sectoral good $v$,

$$M_{so} = (\int_0^1 m_{so}^{\frac{\eta_s-1}{\eta_s}}(v)dv)^{\frac{\eta_s}{\eta_s-1}},$$

with $\eta_s > 1$ and $m_{so}(v)$ denoting the amount of $v$ used in the production of intermediate goods in country $o$ and sector $s$. Consumers in country $d$ have CES preferences over the composite sectoral good $C_{sd}$, with elasticity of substitution $\theta_m > 0$. $C_{sd}$ aggregates sectoral goods according to a CES function with elasticity of substitution $\sigma_s^m > 1$.

The cost of the domestic input bundle in country $o$ for sector $s$ is given by

$$c_{so} = A_s W_o^{1-\alpha_{so}} \prod_{s'} P_{s'o}^{\alpha_{ss'o}},$$

with $A_s > 0$ and $P_{s'o}$ the CES price index associated with the composite sectoral good.

Finally, productivity for good $v$ produced in $o$ by $s$ to deliver to $d$ is $Z_{sod}$, and distributed within each sector as an independent Fréchet with shape $\sigma_s^m$ and scale $T_{sod}^m$.

The sectoral expenditure shares are given by

$$\pi_{sod} = \frac{(P_{sod}^m)^{-\sigma_s^m}}{\sum_{o'=1}^N (P_{so'd}^m)^{\sigma_s^m}} \frac{\left[\sum_{o'=1}^N (P_{so'd}^m)^{-\sigma_s^m}\right]^{\frac{\theta}{\sigma_s^m}}}{\sum_{s'=1}^S \left[\sum_{o'=1}^N (P_{so'd}^m)^{-\sigma_{s'}^m}\right]^{\frac{\theta}{\sigma_{s'}^m}}} \quad \text{with} \quad P_{sod}^m \equiv (T_{sod}^m)^{-1/\theta^m} c_{so}. \tag{C.4}$$

Specializing (C.4) to the domestic pair, $\pi_{sdd}$, and after some algebra, we get the expression for the gains from trade in (35).

## C.2 Mixed CES

Consider a mixed-CES demand system (such as in Adao et al., 2017):

$$\pi_{od} = \int_{\mathbb{R}^M} \int_0^\infty \frac{e^{\beta' \mathrm{Geo}_{od}} W_o^{-\sigma}}{\sum_{o'=1}^M e^{\beta' \mathrm{Geo}_{o'd}} W_{o'}^{-\sigma}} F(\mathrm{d}\sigma, \mathrm{d}\beta)$$

where $F$ is a cumulative distribution function on $\mathbb{R}_+ \times \mathbb{R}^M$ and $\mathrm{Geo}_{od} \in \mathbb{R}^M$ denotes a vector of some bilateral variables (e.g. distance between the origin and destination, or dummy variables that allow for random effects).

To derive this demand system from a Ricardian model with max-stable multivariate Fréchet productivity, we use a CNCES correlation function, as in (6), but let $K \to \infty$:

$$G^d(x_1, \ldots, x_N) = \sum_{k=1}^{\infty} \left( \sum_{o=1}^{N} (\omega_{kod} x_o)^{\frac{1}{1-\rho_k}} \right)^{1-\rho_k} \tag{C.5}$$

where for each $o = 1, \ldots, N$ we have $\omega_{kod} \geq 0$ for each $k = 1, 2, \ldots$ and $\sum_{k=1}^{\infty} \omega_{kod} = 1$.

Assume that productivity when delivering to $d$ is distributed multivariate max-stable Fréchet across origins with shape $\theta$, scales of $\{T_{od}\}_{o=1}^{N}$, and correlation function as in (C.5). The implied demand system is

$$\pi_{od} = \sum_{k=1}^{\infty} \frac{(T^*_{kod} W_o)^{-\sigma_k}}{\sum_{o'=1}^{N} (T^*_{ko'd} W_{o'})^{-\sigma_k}} \frac{\left[\sum_{o'=1}^{N} (T^*_{ko'd} W_{o'})^{-\sigma_k}\right]^{\frac{\theta}{\sigma_k}}}{\sum_{k'=1}^{\infty} \left[\sum_{o'=1}^{N} (T^*_{k'o'd} W_{o'})^{-\sigma_{k'}}\right]^{\frac{\theta}{\sigma_{k'}}}}$$

where $\sigma_k \equiv \theta/(1 - \rho_k)$ and $T^*_{kod} \equiv \omega_{kod} T_{od}$.

Next, we add some additional structure to $T^*_{kod}$ and consider the limit as $\theta \to 0$. Assume that there exists sequences of $\beta_k \in \mathbb{R}^M$ and $\mu_k \geq 0$ for $k = 1, 2, \ldots$ such that $\sum_{k=1}^{\infty} \mu_k = 1$ and $T^*_{kod} = e^{-\beta'_k \text{Geo}_{od}/\sigma_k} \mu_k^{-1/\theta}$. Then

$$\pi_{od} = \sum_{k=1}^{\infty} \frac{e^{\beta'_k \text{Geo}_{od}} W_o^{-\sigma_k}}{\sum_{o'=1}^{N} e^{\beta'_k \text{Geo}_{o'd}} W_{o'}^{-\sigma_k}} \frac{\left[\sum_{o'=1}^{N} e^{\beta'_k \text{Geo}_{o'd}} W_{o'}^{-\sigma_k}\right]^{\frac{\theta}{\sigma_k}} \mu_k}{\sum_{k=1}^{\infty} \left[\sum_{o'=1}^{N} e^{\beta'_{k'} \text{Geo}_{o'd}} W_{o'}^{-\sigma_{k'}}\right]^{\frac{\theta}{\sigma_k}} \mu_{k'}}$$

Letting $\theta \to 0$ we get

$$\pi_{od} \to \sum_{k=1}^{\infty} \frac{e^{\beta'_k \text{Geo}_{od}} W_o^{-\sigma_k}}{\sum_{o'=1}^{N} e^{\beta'_k \text{Geo}_{o'd}} W_{o'}^{-\sigma_k}} \mu_k = \int_{\mathbb{R}^M} \int_0^{\infty} \frac{e^{\beta' \text{Geo}_{od}} W_o^{-\sigma}}{\sum_{o'=1}^{N} e^{\beta' \text{Geo}_{o'd}} W_{o'}^{-\sigma}} P(\mathrm{d}\sigma, \mathrm{d}\beta)$$

for

$$P(\sigma, \beta) \equiv \sum_{k=1}^{\infty} \mathbf{1}\{\sigma \leq \sigma_k, \beta \leq \beta_k\} \mu_k$$

Note that since $P$ is an empirical distribution function on $\mathbb{R}_+ \times \mathbb{R}^M$, and it can arbitrarily approximate $F$. As a consequence, this limiting case corresponds to a mixed-CES import demand system.

# D Data Construction

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

## D.1 Map from SITC Codes to WIOD Sectors

The WIOD data allow us to compute the total value of trade between a sample of 40 countries across 35 sectors from 1995 through 2011. While the sector classification in this dataset comes from aggregating underlying data classified according to the third revision of the International Standard Industrial Classification (ISIC), the COMTRADE tariff data are classified according to the second revision of the Standard International Trade Classification (SITC). In order to merge these data sources, we construct a mapping that assigns SITC codes to aggregates of WIOD sectors.

First, we match ISIC and SITC definitions using existing correspondences to Harmonized System (HS) product definitions. These correspondences come from the World Bank's World Integrated Trade Solution (WITS).[33] This merge matches 5,701 products out of 5,705 total HS products, creating a HS product dataset with 764 SITC codes and 35 ISIC codes. Note that there are 925 SITC codes in the tariff data to be classified into WIOD sectors.

Next, we map the ISIC definitions in this merge to 25 aggregates of WIOD sectors. This leaves products in the ISIC code 99 ("Goods n.e.c.") without a WIOD sector definition. This results in a HS-product-level dataset with labels for the 25 WIOD aggregates and 764 SITC codes.

At this point, there are two issues left to address: (1) classifying SITC codes that have products in multiple WIOD sectors; and (2) classifying the SITC codes in the tariff data that were either matched to ISIC code 99 or were not matched to any ISIC code. First, we determine the most common WIOD sector classification (including "unclassified") at the HS product level of each 4-digit SITC code within the merge. We re-classify all products within an 4-digit SITC sector as belonging to the most common WIOD sector, and break ties manually. This step resolves issue (1) and leaves us with 764 4-digit SITC codes mapped to a unique WIOD sector, and 161 4-digit SITC codes left unclassified. Second, we resolve issue (2) by refining the map by using the most common classification of HS products within each 3-digit SITC code, again breaking ties manually. In this step, we only use the most-common classification at the 3 digit level to classify previously unclassified 4-digit SITC codes, filling in the map. This step mostly resolves issue (2), leaving only 12 4-digit SITC codes unclassified. We complete the map by manually classifying the 12 remaining codes.This results in a map from 925 4-digit SITC codes to 25 WIOD aggregates.

## D.2 Reconciling WIOD and COMTRADE Data

We drop those countries in WIOD with completely missing data in COMTRADE, and aggregate the 35 WIOD sectors to the 25 aggregates in our concordance with 4-digit SITC codes, and restrict the sample to 1999 through 2007. These restrictions leave a balanced sample of 25 WIOD aggregates for 31 countries over 9 years.[34] Finally, we keep the 14 WIOD aggregates that correspond to traded goods. Table D.1 lists the sector and their code.

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

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

Table D.1: WIOD sectoral categories, SGM elasticities, and sectoral expenditure.

| Code | Name | Elasticity $\sigma_j$ | Self-Trade Share | Expenditure Share |
|---|---|---|---|---|
| 1 | Agriculture, Hunting, Forestry and Fishing | 4.15 (0.433) | 0.91 | 0.075 |
| 2 | Mining and Quarrying | 4.37 (1.598) | 0.69 | 0.057 |
| 3 | Food, Beverages and Tobacco | 2.21 (0.199) | 0.87 | 0.108 |
| 4 | Textiles and Leather | 1.81 (0.510) | 0.64 | 0.043 |
| 5 | Wood and Products of Wood and Cork | 1.13 (0.668) | 0.85 | 0.017 |
| 6 | Pulp, Paper, Paper , Printing and Publishing | 1.25 (0.492) | 0.86 | 0.048 |
| 7 | Coke, Refined Petroleum and Nuclear Fuel | 4.01 (1.569) | 0.87 | 0.058 |
| 8 | Chemicals, Rubber, and Plastics | 2.40 (0.511) | 0.69 | 0.125 |
| 9 | Other Non-Metallic Mineral | 0.66 (0.499) | 0.88 | 0.030 |
| 10 | Basic Metals and Fabricated Metal | 3.26 (0.463) | 0.80 | 0.124 |
| 11 | Machinery, Nec | 2.83 (0.683) | 0.61 | 0.071 |
| 12 | Electrical and Optical Equipment | 5.17 (1.568) | 0.51 | 0.108 |
| 13 | Transport Equipment | 2.36 (0.759) | 0.61 | 0.106 |
| 14 | Manufacturing, Nec; Recycling | 2.20 (0.493) | 0.54 | 0.026 |

Notes: SGM = sectoral gravity model. $\sigma_j$ from estimating by PPML the specification in (28) corresponding to column 2 of Table 1. Standard errors clustered at the sector-origin-destination level are in parenthesis. All coefficients are significant at the 0.01 level. Self-trade share calculated as sectoral self-trade relative to total expenditure in the sector. Expenditure share calculated as sectoral expenditure relative to total expenditure.

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

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

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

### D.3 Tariff Interpolation

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

### D.4 WIOD Aggregate-Level Tariffs

To estimate the gravity equation in (28), we require WIOD sector-level tariff data. We aggregate the COMTRADE tariff data to the WIOD aggregate sector level as follows. We use our model-based aggregation procedure to compute the aggregate applied tariff and total trade value in the COMTRADE data by SITC code, exporter, importer, and year. The model implies that when latent factors correspond to WIOD sectors, the within-WIOD-sector factor weights correspond to global expenditure shares. Then, up to a first order approximation around zero tariffs, WIOD sector-level tariff indices are equal to a weighted average of underlying 4-digit SITC tariffs using these global expenditure shares as weights. We use these global expenditure weighted tariff averages for WIOD sector-level tariffs.

## E Latent-Factor Model Estimation: Algorithm

We do not observe all sectors in (32). Additionally, we need to account for observed tariffs, and simultaneously estimate of $\sigma_k$ for $k = 1, \dots, K$. The presence of missing data requires to use an adjusted version of (33), which we describe in Section E.2. We solve this adjusted problem using an extension of the multiplicative-update non-negative matrix factorization (NMF) algorithm of Lee and Seung (1999, 2001) to accommodate covariates and missing data, which we present in Section E.3.

## E.1 Identification Conditions For NMF

Here, we present sufficient conditions from the literature on identification of non-negative matrix factorizations—see Fu et al. (2019) for a survey. Given a non-negative matrix $\Pi \in \mathbb{R}_+^{S\times M}$, any pair of matrices $(\Lambda, \Phi^*)$ with $\Pi = \Lambda\Phi^*$, $\Lambda \in \mathbb{R}_+^{S\times K}$, and $\Phi^* \in \mathbb{R}_+^{K\times M}$ is a *non-negative matrix factorization* (NMF). A NMF is *identified* if it is unique up to permutation and scaling of the columns of $\Lambda$ and the rows of $\Phi^*$. That is, the matrices of any other factorization can be written as $\Lambda R^{-1}$ and $R\Phi^*$ where $R$ is the product of a permutation matrix with a strictly positive diagonal matrix.

The intuition for identification of NMF is geometric. The rows of $\Lambda$ (or columns of $\Phi^*$), viewed as points in the factor space, $\mathbb{R}_+^K$, must be "spread out" in some sense that makes enough of the non-negativity constraints bind such that permutations are the only possible rotations of the factorization (with scale typically pinned down through some normalization). Intuitively if the non-negativity constraints are slack, then there might be a rotation that keeps all the constraints slack. In which case, the factorization would not be identified. This idea is analogous to the role of sign restrictions limiting rotations of latent structural shocks in structural VARs (Faust, 1998; Uhlig, 2005; Fry and Pagan, 2011; Arias et al., 2018).

It is useful to conceptualize the geometry using the cone generated by $\Lambda'$, $\text{cone}(\Lambda') = \{\Lambda' x \mid x \in \mathbb{R}_+^S\}$, which is the subset of $\mathbb{R}_+^K$ consisting of positive linear combinations of the rows of $\Lambda$. When this cone is large enough within $\mathbb{R}_+^K$, any rotation other than a permutation will violate non-negativity.

The following result provides a stark example of this logic and has a clear economic interpretation when the entries of $\Lambda$ correspond to how each sector, $s$, loads on each factor, $k$. In particular, it assumes that factors do not share sectors, forcing $\text{cone}(\Lambda')$ to entirely fill the positive orthant.

**Theorem E.1** (Ding et al. (2006))**.** *If $\Lambda$ is orthogonal so that $\Lambda'\Lambda = I$, then $(\Lambda, \Phi^*)$ is identified.*

First, the diagonal of the orthogonality constraint normalizes the scale of each column of $\Lambda$, removing the scale indeterminacy of the factorization. Second, the off-diagonal entries force the columns of $\Lambda$ to be mutually orthogonal. Since these columns have only non-negative entries, there can never be an $s$ such that $\lambda_{sk}$ and $\lambda_{sk'}$ are both positive unless $k = k'$, implying that each sector can only load on a single factor (although factors can put weight on many sectors). In this case, sectors are partitioned into $K$ groups which correspond to the factors. That is, factors do not share sectors and the non-zero entries of the columns of $\Lambda$ contain the weights across sectors within each group. Indeed, this type of restriction means that factors correspond to some aggregation of sectors—which is precisely the assumption of a SGM model at that aggregated level. Under this economic restriction, each row of $\Lambda$ lies along an axis of $\mathbb{R}_+^K$—it is a scaled standard basis vector. Geometrically, this means that the rows of $\Lambda$ are maximally spread out in $\mathbb{R}_+^K$, implying that $\text{cone}(\Lambda') = \mathbb{R}_+^K$ and only permutations preserve non-negativity.

Although this example clarifies the geometric intuition for why non-negativity constraints can ensure identification, orthogonality of $\Lambda$ is far from necessary. For instance, Donoho and Stodden (2004) provide a much weaker sufficient condition, which in our context can be interpreted as requiring that each factor is unique to at least one sector. In this case, most sectors can be shared across factors (breaking the restriction of the SGM). However, we still get the geometric result that $\text{cone}(\Lambda') = \mathbb{R}_+^K$ without requiring all rows of $\Lambda$ to correspond to scaled standard basis vectors.

One possible issue with this weaker assumption is that we may want to allow every sector to use multiple factors. Huang et al. (2014) provide a much weaker condition that allows for this possibility. It is based on the following notion of the rows of $\Lambda$ being "spread out" in $\mathbb{R}^K_+$.

**Definition 1** (Sufficiently Scattered)**.** $\Lambda \in \mathbb{R}^{S\times K}_+$ *is* sufficiently scattered *if:*

1. $\mathcal{C} \equiv \{x \in \mathbb{R}^K \mid x'\mathbf{1} \geq \sqrt{(K-1)x'x}\} \subseteq$ *cone*$(\Lambda')$.
2. *cone*$(\Lambda') \subseteq$ *cone*$(R)$ *does not hold for any orthonormal $R$ except the permutation matrices.*

To interpret the second-order cone, $\mathcal{C}$, we can project it onto the unit simplex in $\mathbb{R}^K_+$. This projection is the largest ($K-1$ dimensional) sphere contained inside the simplex and it is tangent to each facet of the simplex. (For the $K = 3$ case, this projection is a circle on the simplex that is tangent to each side of the simplex.) If the rows of $\Lambda$ (after projection onto the simplex) all were inside of this sphere, then they could be arbitrarily rotated without ever hitting the non-negativity constraints. However, if there are rows of $\Lambda$ that lie outside of $\mathcal{C}$, not all rotations become possible as they will eventually hit the facets of $\mathbb{R}^K_+$. When $\Lambda$ is sufficiently scattered, the rows of $\Lambda$ are spread out enough relative to $\mathcal{C}$ to rule out all rotations except permutations. The first condition implies that there are faces of cone$(\Lambda')$ that intersect the faces of $\mathbb{R}^K_+$ (ruling our small rotations), while the second is a regularity condition that means that cone$(\Lambda')$ is large enough to not simply tangentially contain $\mathcal{C}$ (ruling out large rotations, other than permutations).

This concept leads to the following sufficient condition for identification.

**Theorem E.2** (Huang et al. (2014))**.** *If $\Lambda$ and $\Phi^{*'}$ are sufficiently scattered, then $(\Lambda, \Phi^*)$ is identified.*

If we view the rows of $\Lambda$ (columns of $\Phi^*$) as being drawn from some distribution with full support on $\mathbb{R}^K_+$ and a positive probability of zero entries (necessary for the facets of cone$(\Lambda')$ to intersect the facets of $\mathbb{R}^K_+$), then it becomes very likely that this sufficient condition will hold as the number of rows (columns) get large. Indeed, Fu et al. (2019) use numerical examples to show that we get identification with high probability as the dimensions of the data get large for fixed $K$. In our context, this essentially means that we assume that $\Lambda$ and $\Phi^*$ contain zeros, and we use highly disaggregate sectoral data across many countries. Intuitively, each additional sector and country-pair adds additional non-negativity constraints, further restricting possible rotations in the low dimensional factor space, $\mathbb{R}^K_+$.

## E.2 Accounting for Missing Data

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

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

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

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

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

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

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

## E.3 NMF with Covariates and Missing Data

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The Poisson deviance is

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

It is useful to re-write this expression as

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

But then

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

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

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

We can therefore write the gradient in $\Lambda_k$ as

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

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

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

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

Similarly, we get an updating rule for $\Phi_k^*$ given by

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

Finally, the derivative in $\sigma_k$ is

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

The implied updating rule is

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

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

# F Additional Results

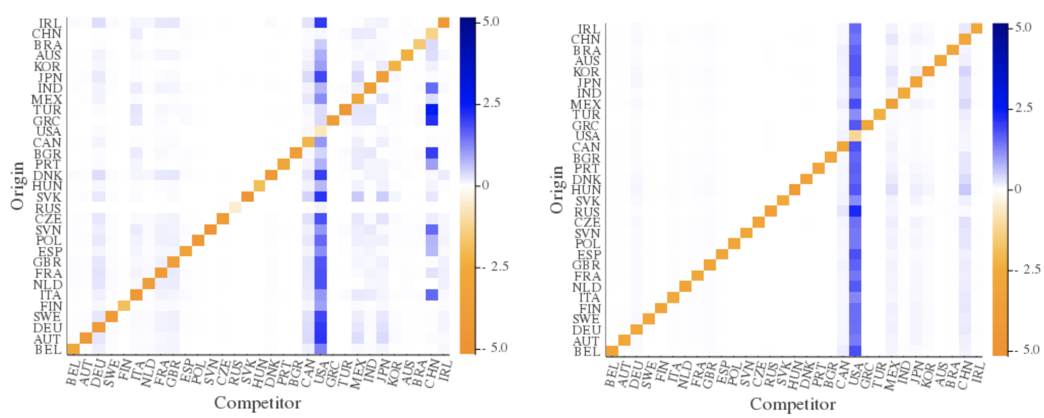
Table F.1: Elasticity estimates: LFM with different number of factors.

| | Number of factors $K$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $\sigma_1$ | 3.003 | 3.933 | 3.300 | 4.814 | 3.929 | 7.866 | 5.175 | 9.944 |
| $\sigma_2$ | | 2.767 | 2.638 | 3.342 | 3.780 | 3.536 | 4.868 | 5.471 |
| $\sigma_3$ | | | 1.592 | 2.614 | 3.573 | 2.559 | 4.624 | 4.417 |
| $\sigma_4$ | | | | 1.223 | 0.806 | 0.804 | 1.481 | 3.435 |
| $\sigma_5$ | | | | | 0.418 | 0.574 | 0.670 | 1.884 |
| $\sigma_6$ | | | | | | 0.163 | 0.390 | 1.594 |
| $\sigma_7$ | | | | | | | 0.375 | 0.111 |
| $\sigma_8$ | | | | | | | | 0.108 |
| $\theta = \min_k \sigma_k$ | 3.003 | 2.767 | 1.592 | 1.223 | 0.418 | 0.163 | 0.375 | 0.108 |
| Average $\sigma_k$ | 3.003 | 3.350 | 2.510 | 2.998 | 2.501 | 2.584 | 2.512 | 3.371 |

Notes: Estimates of factor-level elasticities, $\sigma_k$, for latent-factor models (LFM) with K = 1,. . . , 8. In each case, $F1$ is the factor with the highest elasticity, while $FK$ is the one with the lowest, with $\theta = \sigma_K$.

Figure F.1: Expenditure Elasticities, US market: LFM vs SGM.

(a) $\hat{\varepsilon}_{o,o',USA}$, LFM
(b) $\hat{\varepsilon}_{o,o',USA}$, SGM




Notes: Estimates of expenditure elasticities $\varepsilon_{o,o',USA}$ calculated using (34) and estimates from the latent-factor model (LFM) and sectoral gravity model (SGM). Year 2007.

Table F.2: Gains From Trade: Models' Comparison.

| Country Name | Country Code | Domestic share | Gains from Trade | | | |
|---|---|---|---|---|---|---|
| | | | CES | SGM | SGM + IO | LFM |
| Australia | AUS | 0.73 | 1.28 | 1.15 | 1.21 | 1.74 |
| Austria | AUT | 0.39 | 1.43 | 1.59 | 2.03 | 5.59 |
| Belgium | BEL | 0.17 | 1.97 | 2.56 | 4.79 | 28.69 |
| Bulgaria | BGR | 0.45 | 1.36 | 1.45 | 1.97 | 4.26 |
| Brazil | BRA | 0.90 | 1.04 | 1.04 | 1.07 | 1.10 |
| Canada | CAN | 0.53 | 1.28 | 1.34 | 1.56 | 3.48 |
| China | CHN | 0.90 | 1.04 | 1.03 | 1.09 | 1.14 |
| Czech Republic | CZE | 0.45 | 1.36 | 1.37 | 1.97 | 2.39 |
| Germany | DEU | 0.55 | 1.26 | 1.29 | 1.50 | 1.86 |
| Denmark | DNK | 0.38 | 1.45 | 1.62 | 1.98 | 6.29 |
| Spain | ESP | 0.63 | 1.19 | 1.22 | 1.40 | 1.57 |
| Finland | FIN | 0.57 | 1.24 | 1.26 | 1.48 | 3.27 |
| France | FRA | 0.59 | 1.22 | 1.26 | 1.46 | 1.99 |
| Great Britain | GBR | 0.52 | 1.29 | 1.32 | 1.48 | 1.98 |
| Greece | GRC | 0.57 | 1.24 | 1.32 | 1.49 | 2.78 |
| Hungary | HUN | 0.37 | 1.47 | 1.54 | 2.30 | 7.36 |
| India | IND | 0.88 | 1.05 | 1.06 | 1.10 | 1.26 |
| Ireland | IRL | 0.43 | 1.38 | 1.45 | 1.72 | 2.96 |
| Italy | ITA | 0.71 | 1.14 | 1.15 | 1.26 | 1.26 |
| Japan | JPN | 0.86 | 1.06 | 1.06 | 1.13 | 1.30 |
| Korea | KOR | 0.78 | 1.10 | 1.10 | 1.27 | 1.36 |
| Mexico | MEX | 0.64 | 1.19 | 1.20 | 1.36 | 1.81 |
| Netherlands | NLD | 0.28 | 1.63 | 1.73 | 2.19 | 11.38 |
| Poland | POL | 0.58 | 1.23 | 1.28 | 1.51 | 2.51 |
| Portugal | PRT | 0.52 | 1.29 | 1.35 | 1.66 | 2.90 |
| Russia | RUS | 0.77 | 1.11 | 1.14 | 1.23 | 1.57 |
| Slovakia | SVK | 0.33 | 1.53 | 1.59 | 2.39 | 3.67 |
| Slovenia | SVN | 0.31 | 1.57 | 1.96 | — | 5.08 |
| Sweden | SWE | 0.50 | 1.31 | 1.34 | 1.57 | 2.53 |
| Turkey | TUR | 0.76 | 1.11 | 1.15 | 1.25 | 1.50 |
| United States | USA | 0.76 | 1.11 | 1.12 | 1.19 | 1.46 |

Notes: Gains from trade = Real wages in the observed equilibrium relative to autarky real wages. Calculations using estimates from latent-factor model (LFM), sectoral gravity model (SGM), SGM augmented by input-output links (SGM + IO), and CES model as in ACR (CES). Year 2007.

Figure F.2: Tariff effects, densities.

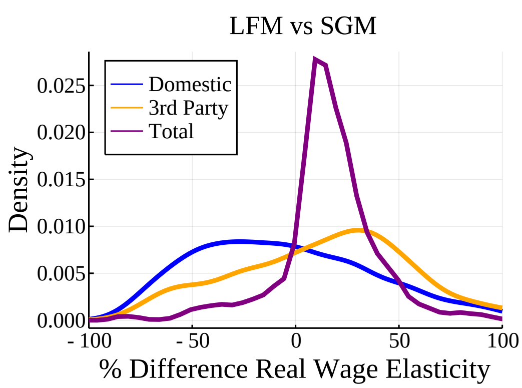



Table F.3: Tariff effects, moments.

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

Notes: Figure F.2 shows density plots of the percent difference in the components of (36) between the latent factor model (LFM) and sectoral gravity model (SGM). Blue corresponds to the domestic wage effect, orange corresponds to the third party effect, and purple shows the full effect. The direct tariff effect is identical between the two models. Table F.3 shows moments of these densities. Year 2007.

# G Alternative Estimation for the Parameter $\theta$

The alternative estimation of the parameter $\theta$ exploits the structure of the import demand system at the latent-factor level, and uses the estimates of those expenditure shares from the LFM estimation procedure. We use a specification that relies on variation across factors and inferred within-factor relative prices from the LFM estimates.

Summing over origins $o$ in (25) yields the between-factor expenditure share

$$\pi^B_{kdt} = \frac{\left[\sum_{o'=1}^{N}(t^*_{ko'dt}W_{o't}/A_{ko'dt})^{-\sigma_k}\right]^{\frac{\theta}{\sigma_k}}}{\sum_{k'=1}^{K}\left[\sum_{o'=1}^{N}(t^*_{k'o'dt}W_{o't}/A_{k'o'dt})^{-\sigma_{k'}}\right]^{\frac{\theta}{\sigma_{k'}}}} \equiv \left(\frac{P^*_{kdt}}{P^*_{dt}}\right)^{-\theta}, \tag{G.1}$$

Multiplying and dividing by $(P^*_{kodt})^{-\theta}$ with $P^*_{kodt} \equiv t^*_{kodt}W_{ot}/A_{kodt}$ yields

$$\pi^B_{kdt} = \left(\frac{t^*_{kodt}W_{ot}/A_{kodt}}{P_{dt}}\right)^{-\theta}\left(\pi^W_{ko'dt}\right)^{-\frac{\theta}{\sigma_k}},$$

where $\pi^W_{kodt} = (P^*_{kodt}/P^*_{kdt})^{-\sigma_k}$.

We estimate the parameter $\theta$ from the coefficient on the tariff index $t^*_{kodt}$ in

$$\pi^B_{kdt} = \exp\left(-\theta \ln t^*_{kodt} + D^1_{kot} + D^2_{dt} + D^3_{kod} - \theta \ln \hat{Z}^*_{kodt}\right) u_{kodt}, \tag{G.2}$$

where $\hat{Z}^*_{kodt} \equiv \left(\hat{\pi}^W_{kodt}\right)^{-1/\hat{\sigma}_k}$, and $D^l$, for $l = 1, 2, 3$, are fixed effects. Identification comes from controlling for within-factor expenditure using our LFM estimates. The identification assumption is that the error term (e.g. unobserved component of trade costs) is orthogonal to the latent-factor tariff index conditional on the other covariates. We estimate this equation by PPML. Results are gathered in columns 1-3 of Table G.1. The Wald test in the last row indicates that estimates are statistically indistinguishable from our baseline estimate of $\theta = 0.375$.

The analogous procedure can be applied to estimating $\theta$ in the context of the sectoral gravity model (SGM) in which factors are specific to sectors. In that case, an equation analogous to (G.2) can be estimated using the observed sectoral data on expenditure and tariffs. Results are shown in columns 4-6 of Appendix Table G.1. These estimates are statistically the same as our baseline estimate of $\theta$.

Table G.1: Alternative Estimates of the Parameter $\theta$. PPML.

| Dep. variable | Between-factor $\pi^B_{kodt}$ | | | Dep. variable | Between-sector $\pi^B_{jodt}$ | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | | (4) | (5) | (6) |
| $\ln t^*_{kodt}$ | -1.168 | -0.939* | -0.935* | $\ln t_{jodt}$ | -0.164 | -0.310* | -0.305* |
| | (1.763) | (0.416) | (0.416) | | (0.373) | (0.146) | (0.144) |
| $\ln \hat{Z}_{kodt}$ | Yes | Yes | No | $\ln \hat{Z}_{jodt}$ | Yes | Yes | No |
| $k \times \ln \hat{Z}_{kodt}$ | No | No | Yes | $j \times \ln \hat{Z}_{jodt}$ | No | No | Yes |
| $k \times o \times t$ | Yes | Yes | Yes | $j \times o \times t$ | Yes | Yes | Yes |
| $d \times t$ | Yes | Yes | Yes | $d \times t$ | Yes | Yes | Yes |
| $o \times d$ | Yes | No | No | $o \times d$ | Yes | No | No |
| $k \times o \times d$ | No | Yes | Yes | $j \times o \times d$ | No | Yes | Yes |
| Observations | 60,542 | 60,542 | 60,542 | Observations | 121,010 | 121,010 | 121,010 |
| Degrees of freedom | 57,347 | 58,307 | 58,301 | Degrees of freedom | 115,862 | 116,822 | 116,809 |
| Deviance | 1,632 | 122.3 | 122.2 | Deviance | 1,002 | 68.25 | 67.55 |
| $\chi^2$ | 0.20 | 1.83 | 1.81 | $\chi^2$ | 0.32 | 0.20 | 0.23 |
| P-value | 0.65 | 0.17 | 0.17 | P-value | 0.57 | 0.65 | 0.62 |

Notes: Estimates of (G.2). In columns 4-5, $j = 1, \ldots, 14$ denotes WIOD aggregate sectors. Robust standard errors in parenthesis, clustered by $k \times d$ ($j \times d$), with levels of significance denoted by *** p < 0.001, and ** p < 0.01 and * p<0.05. Last row reports Wald test of the null hypothesis that estimates are not significantly different from 0.375, the baseline LFM estimate of $\theta$.