Trade with Correlation
Online Appendix

Nelson Lind
Emory University

Natalia Ramondo
Boston University and NBER

June 2, 2022

O.1 Properties of Fréchet Random Variables

Definition O.1 (Fréchet). *A random variable, Z, has a Fréchet distribution if there exists a scale parameter $T > 0$ and shape parameter $\theta > 0$ such that $\mathbb{P}[Z \leq z] = e^{-Tz^{-\theta}}$.*

Lemma O.1. *If Z is Fréchet with shape $\theta > 0$, then its scale is $T \equiv -\ln \mathbb{P}[Z \leq 1]$.*

Proof. $-\ln \mathbb{P}[Z \leq 1] = -\ln e^{-T} = T$. □

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

Lemma O.2. *If Z is Fréchet with scale T and shape $\theta > 1$, then $\mathbb{E}[Z] = \Gamma(1 - 1/\theta)T^{1/\theta}$.*

Proof. $\mathbb{E}[Z] = \int_0^\infty z\theta T z^{-\theta-1}e^{-Tz^{-\theta}}\mathrm{d}z = \int_0^\infty t^{-1/\theta}e^{-t}\mathrm{d}t T^{1/\theta} = \Gamma(1-1/\theta)T^{1/\theta}$. □

Definition O.2 (Copula). *A function $C : [0,1]^N \to [0,1]$ is a copula if there exists a random vector $(U_1, \ldots, U_N)$ on $[0,1]^N$ such that $C(u_1, \ldots, u_N) = \mathbb{P}[U_1 \leq u_1, \ldots, U_N \leq u_N]$ for each $(u_1, \ldots, u_N) \in [0,1]^N$.*

Given a random vector $(Z_1, \ldots, Z_N)$, its copula is

$$C(u_1, \ldots, u_N) \equiv \mathbb{P}[F_1(Z_1) \leq u_1, \ldots, F_N(Z_N) \leq u_N]$$

where $F_i(z) \equiv \mathbb{P}[Z_i \leq z]$ for each $i = 1, \ldots, N$.

Definition O.3 (Max-Stable Copula). *A copula $C : [0,1]^N \to [0,1]$ is* max-stable *if $C(u_1, \ldots, u_N) = C(u_1^{1/m}, \ldots, u_N^{1/m})^m$ for any $m > 0$ and all $(u_1, \ldots, u_N) \in [0,1]^N$.*

Definition O.4 (Correlation Function). *A function $G : \mathbb{R}_+^N \to \mathbb{R}_+$ is a* correlation function *if* $\exp[-G(-\ln u_1, \ldots, u_N)]$ *is a max-stable copula.*

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

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

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

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

Recall that for a max-stable copula,

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

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

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

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

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

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

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

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

The sign-switching property simply follows from the non-negativity of joint probability density functions. If the mixed partial derivatives of G exist and are continuous up to order N, then for any integer $M \leq N$ and distinct integers n_m for $m = 1, \ldots, M$ we have

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

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

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

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

□

Definition O.5 (Max-Stable Multivariate Fréchet). *A random vector, $(Z_1, \ldots, Z_N)$, has a max-stable multivariate Fréchet distribution if there exists a shape parameter $\theta > 0$ such that for any $\alpha_i \geq 0$ with $i = 1, \ldots, N$ the random variable $\max_{i=1,\ldots,N} \alpha_i Z_i$ has a Fréchet distribution with shape parameter θ.*

Lemma O.4. *A vector $(Z_1, \ldots, Z_N)$ is max-stable multivariate Fréchet if and only if there exists $\theta > 0$, scale parameters $\{T_i\}_{i=1}^N$, and a correlation function, G, such that*

$$\mathbb{P}\left[Z_1 \leq z_1, \ldots, Z_N \leq z_N\right] = \exp\left[-G(T_1 z_1^{-\theta}, \ldots, T_N z_N^{-\theta})\right]. \tag{O.1}$$

Proof. (Sufficiency) Suppose $(Z_1, \ldots, Z_N)$ is max-stable multivariate Fréchet. Then by Definition O.1, there exists a θ such that, for each i, $Z_i = \max_{n=1,\ldots,N} \mathbf{1}\{i = n\} Z_n$ is Fréchet with shape θ. By Definition O.5, there is then $\{T_i\}_{i=1}^N$ such that $\mathbb{P}[Z_i \leq z] = \exp(-T_i z^{-\theta}) \equiv F_i(z)$ for each $i = 1, \ldots, N$.

Additionally, letting $z_i > 0$ for $i = 1, \ldots, N$, we have $\max_{i=1,\ldots,N} Z_i/z_i$ Fréchet with shape θ and some scale. In particular, by Lemma O.1, its scale is

$$-\ln \mathbb{P}\left[\max_{i=1,\ldots,N} Z_i/z_i \leq 1\right] = G(T_1 x_1^{-\theta}, \ldots, T_N z_N^{-\theta})$$

where

$$G(x_1,\ldots,x_N) \equiv -\ln \mathbb{P}\left[\max_{i=1,\ldots,N}(x_i/T_i)^{1/\theta} Z_i \le 1\right].$$

Therefore,

$$\begin{aligned}
\mathbb{P}\left[Z_1 \le z_1,\ldots,Z_N \le z_N\right] &= \mathbb{P}\left[Z_1/z_1 \le 1,\ldots,Z_N/z_N \le 1\right] \\
&= \mathbb{P}\left[\max_{i=1,\ldots,N} Z_i/z_i \le 1\right] \\
&= \mathbb{P}\left[\max_{i=1,\ldots,N} Z_i/z_i \le t\right]\Bigg|_{t=1} \\
&= e^{-G(T_1 z_1^{-\theta},\ldots,T_N z_N^{-\theta})t^{-\theta}}\Big|_{t=1} = e^{-G(T_1 z_1^{-\theta},\ldots,T_N z_N^{-\theta})}
\end{aligned}$$

We now show that G is a correlation function. To do so, we show that the copula of $(Z_1,\ldots,Z_N)$,

$$\begin{aligned}
C(u_1,\ldots,u_N) &\equiv \mathbb{P}\left[F_1(X_1) \le u_1,\ldots,F_N(X_N) \le u_N\right] \\
&= \mathbb{P}\left[X_1 \le F_1^{-1}(u_1),\ldots,X_N \le F_N^{-1}(u_N)\right] \\
&= \exp\left[-G(T_1 F_1^{-1}(u_1)^{-\theta},\ldots,T_N F_N^{-1}(u_N)^{-\theta})\right] \\
&= \exp\left[-G(-\ln u_1,\ldots,-\ln u_N)\right],
\end{aligned}$$

is max-stable.

Next, let $m > 0$ and fix a $(u_1,\ldots,u_N) \in [0,1]^N$. We have

$$\begin{aligned}
C(u_1^{1/m},\ldots,u_N^{1/m}) &= \exp\left[-G(m^{-1}\ln u_1^{-1},\ldots,-m^{-1}\ln u_N^{-1})\right] \\
&= \mathbb{P}\left[\max_{i=1,\ldots,N}(m^{-1}\ln u_i^{-1}/T_i)^{1/\theta} Z_i \le 1\right] \\
&= \mathbb{P}\left[\max_{i=1,\ldots,N}(\ln u_i^{-1}/T_i)^{1/\theta} Z_i \le m^{1/\theta}\right] \\
&= \exp\left[-G(-\ln u_1,\ldots,-\ln u_N)m^{-1}\right] \\
&= C(u_1,\ldots,u_N)^{1/m}.
\end{aligned}$$

where the second to last line follows from $\max_{i=1,\ldots,N}(\ln u_i^{-1}/T_i)^{1/\theta} Z_i$ being distributed Fréchet with scale $G(-\ln u_1,\ldots,-\ln u_N)$ and shape θ. Therefore, we have $C(u_1,\ldots,u_N) = C(u_1^{1/m},\ldots,u_N^{1/m})^m$, so C is a max-stable copula and G is a correlation function.

(Necessity) Let $T_i > 0$ for each $i = 1,\ldots,N$, and let $G : \mathbb{R}_+^N \to \mathbb{R}_+$ be a correlation function. Suppose that $(Z_1,\ldots,Z_N)$ satisfies

$$\mathbb{P}[Z_1 \le z_1,\ldots,Z_N \le z_N] = \exp\left[-G(T_1 z_1^{-\theta},\ldots,T_N z_N^{-\theta})\right].$$

Let $(\alpha_1, \ldots, \alpha_N) \in \mathbb{R}_+^N$ and consider the distribution of $\max_{i=1,\ldots,N} \alpha_i Z_i$,

$$\begin{aligned}
\mathbb{P}\left[\max_{i=1,\ldots,N} \alpha_i Z_i \leq t\right] &= \mathbb{P}[\alpha_1 Z_1 \leq z, \ldots, \alpha_N Z_N \leq t] \\
&= \mathbb{P}[Z_1 \leq t/\alpha_1, \ldots, Z_N \leq t/\alpha_N] \\
&= \exp\left[-G(T_1\alpha_1^\theta t^{-\theta}, \ldots, T_N\alpha_N^\theta t^{-\theta})\right] \\
&= \exp\left[-G(T_1\alpha_1^\theta, \ldots, T_{Nd}\alpha_N^\theta)t^{-\theta}\right],
\end{aligned}$$

where the last equality uses the homogeneity of G. Therefore, $\max_{i=1,\ldots,N} \alpha_i Z_i$ is Fréchet with scale parameter $G^d(T_1\alpha_1^\theta, \ldots, T_N\alpha_N^\theta)$ and shape θ. As a result, we conclude that $(Z_1, \ldots, Z_N)$ is max-stable multivariate Fréchet.

□

Lemma O.5. *Let $(Z_1, \ldots, Z_N)$ be max-stable multivariate Fréchet with scale parameters $\{A_i\}_{i=1}^N$, shape α, and correlation function $G : \mathbb{R}_+^N \to \mathbb{R}_+$. Then, for any $B_i \geq 0$ $i = 1, \ldots, N$ and $\beta > 0$, $\max_{i=1,\ldots,N} B_i Z_i^\beta$ is Fréchet with scale $G(A_1 B_1^{\alpha/\beta}, \ldots, A_N B_N^{\alpha/\beta})$, and shape α/β.*

Proof.

$$\begin{aligned}
\mathbb{P}\left[\max_{i=1,\ldots,N} B_i Z_i^\beta \leq t\right] &= \mathbb{P}\left[Z_1 \leq (t/B_1)^{1/\beta}, \ldots, Z_N \leq (t/B_N)^{1/\beta}\right] \\
&= \exp\left[-G(A_1(t/B_1)^{-\alpha/\beta}, \ldots, A_N(t/B_N)^{-\alpha/\beta})\right] \\
&= \exp\left[-G(A_1 B_1^{\alpha/\beta}, \ldots, A_N B_N^{\alpha/\beta})t^{-\alpha/\beta}\right].
\end{aligned}$$

□

Lemma O.6. *Let $(Z_1, \ldots, Z_N)$ be max-stable multivariate Fréchet with scale parameters $\{T_i\}_{i=1}^N$, shape θ, and continuously differentiable correlation function $G : \mathbb{R}_+^N \to \mathbb{R}_+$. Then*

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

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

$$\begin{aligned}
&\mathbb{P}\left[\max_{i'=1,\ldots,N} Z_{i'} \leq z \text{ and } Z_i = \max_{i'=1,\ldots,N} Z_{i'}\right] = \mathbb{P}\left[Z_j \leq z \text{ and } Z_i \leq Z_{i'}, \forall i' \neq i\right] \\
&= \int_0^z \frac{\partial}{\partial t}\, \mathbb{P}\left[Z_{i'} \leq z, \forall i' \neq i, \text{ and } Z_i \leq t\right]\big|_{z=t}\,\mathrm{d}t = \int_0^z \frac{\partial}{\partial z_i}\, e^{-G(T_1 z_1^{-\theta},\ldots,T_N z_N^{-\theta})}\Big|_{z_1=t,\ldots,z_N=t}\,\mathrm{d}t \\
&= \int_0^z e^{-G(T_1 z_1^{-\theta},\ldots,T_N z_N^{-\theta})} G_i(T_1 z_1^{-\theta},\ldots,T_N z_N^{-\theta}) T_i \theta z_i^{-\theta-1}\Big|_{z_1=t,\ldots,z_N=t}\,\mathrm{d}t \\
&= \int_0^z e^{-G(T_1,\ldots,T_N)t^{-\theta}} G_i(T_1,\ldots,T_N) T_i \theta t^{-\theta-1}\mathrm{d}t \\
&= \frac{T_i G_i(T_1,\ldots,T_N)}{G(T_1,\ldots,T_N)} \int_0^z e^{-G(T_1,\ldots,T_N)t^{-\theta}} G(T_1,\ldots,T_N)\theta t^{-\theta-1}\mathrm{d}t \\
&= \frac{T_i G_i(T_1,\ldots,T_N)}{G(T_1,\ldots,T_N)} e^{-G(T_1,\ldots,T_N)z^{-\theta}}.
\end{aligned}$$

Let $z \to \infty$ to get $\mathbb{P}\left[Z_i = \max_{i'=1,\ldots,N} Z_{i'}\right] = \frac{T_i G_i(T_1,\ldots,T_N)}{G(T_1,\ldots,T_N)}$.

Finally, part 2 follows from the previous results:

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

□

O.2 Models in the GEV Class

We present applications that extend the Ricardian model of trade in EK to many sectors (Caliendo and Parro, 2015), multinational production (Ramondo and Rodríguez-Clare, 2013), domestic geography (Ramondo et al., 2016), and global value chains (Antràs and de Gortari, 2020). We also present the case of mixed CES. All of these models deliver a GEV import demand system.

O.2.1 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 ϵ_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 η_s, where $\eta_s - 1 > \epsilon_s$. Given trade costs τ_{sod}, the share of destination d's expenditure on goods from origin o and sector s is

$$\pi_{sod} = \left(\frac{\tau_{sod} \frac{W_o}{A_{so}}}{P_{sd}} \right)^{-\epsilon_s} \left(\frac{P_{sd}}{P_d} \right)^{-\bar{\theta}}, \tag{O.2}$$

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 θ, and sector-level correlation function

$$G^{sd}(x_1, \ldots, x_N) = \left(\sum_{o=1}^{N} x_o^{1/(1-\rho_s)} \right)^{1-\rho_s}, \tag{O.3}$$

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

$$\pi_{sod} = \left(\frac{T_{sod}^{-1/\theta} W_o}{P_{sd}} \right)^{-\frac{\theta}{1-\rho_s}} \left(\frac{P_{sd}}{P_d} \right)^{-\theta}, \tag{O.4}$$

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 (O.2) 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 (O.4)—and (O.2)—-is expenditure within sector s and is CES with elasticity $\theta/(1-\rho_s)$ — ϵ_s in (O.2). The second term refers to between-sector expenditure and is also CES with elasticity θ — $\bar{\theta}$ in (O.2). 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.

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 σ_s^m and scale T_{sod}^m.

The sectoral expenditure shares are given by

$$\pi_{sod} = \frac{(P_{sod}^m)^{-\sigma_s^m}}{\sum_{o'=1}^N (P_{so'd}^m)^{\sigma_s^m}} \frac{\left[\sum_{o'=1}^N (P_{so'd}^m)^{-\sigma_s^m}\right]^{\frac{\theta}{\sigma_s^m}}}{\sum_{s'=1}^S \left[\sum_{o'=1}^N (P_{so'd}^m)^{-\sigma_{s'}^m}\right]^{\frac{\theta}{\sigma_{s'}^m}}} \quad \text{with} \quad P_{sod}^m \equiv (T_{sod}^m)^{-1/\theta^m} c_{so}. \tag{O.5}$$

Specializing (O.5) to the domestic pair, π_{sdd}, and after some algebra, we get the expression for the gains from trade in (31).

O.2.2 Multinational Production

Assume that productivity is specific to a good, a location of production, and the home country of the technology, j. Productivity for each home country j across locations of production o has a CES correlation function as in (O.3), with s replaced by j. In this application, the parameter ρ_j measures correlation across production locations for firms with home country j.

The expenditure share on goods produced in o for d with technologies from j is

$$\pi_{jod} = \left(\frac{P_{jod}}{P_{jd}}\right)^{-\frac{\theta}{1-\rho_j}} \left(\frac{P_{jd}}{P_d}\right)^{-\theta}, \tag{O.6}$$

where $P_{jod} \equiv T_{jod}^{-1/\theta} W_o$, and $P_{jd} \equiv (\sum_{o=1}^N P_{jod}^{-\frac{\theta}{1-\rho_j}})^{-\frac{1-\rho_j}{\theta}}$. The factor demand system in (O.6) matches the one in the model of multinational production in Ramondo and Rodríguez-

Clare (2013) for $\rho_j = \rho$ and $T_{jod}^{-1/\theta} = \tau_{od} h_{jo}/A_j$.

O.2.3 Multiple Regions

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

$$G^{nd}(x_1, \ldots, x_{R_n}) = \left(\sum_{r=1}^{R_n} x_r^{1/(1-\rho_n)} \right)^{1-\rho_n}. \tag{O.7}$$

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

$$\pi_{rnd} = \left(\frac{P_{rnd}}{P_{nd}} \right)^{-\frac{\theta}{1-\rho_n}} \frac{P_{nd}^{-\theta}}{\sum_{n'} P_{n'd}^{-\theta}} \quad \text{with} \quad P_{nd} = \left(\sum_{r'=1}^{R_n} P_{r'nd}^{-\frac{\theta}{1-\rho_n}} \right)^{-\frac{1-\rho_n}{\theta}}. \tag{O.8}$$

The first fraction on the right-hand side of (O.8) is the probability of importing from region r in country n conditional on importing from some region in country n, while the second fraction is the probability of importing from country n into d.

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

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

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

O.2.4 Global Value Chains

We now show that the model of global value chains in Antràs and de Gortari (2020) generates a GEV demand system. That is, it has the same macroeconomic implications as a model without global value chains, but where productivity follows a multivariate max-stable Fréchet distribution with an appropriately chosen correlation function.

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

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

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

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

$$\pi_{\ell d} = \frac{T_{\ell d} c_\ell^{-\theta}}{\sum_{\ell'} T_{\ell' d} c_{\ell'}^{-\theta}}. \tag{O.9}$$

This factor demand share matches the one in Antràs and de Gortari (2020) for $T_{\ell d} = \tau_{\ell(K),d}^{-\theta} T_{\ell(K)}^{1-\sum_{k=1}^{K-1} \alpha_k} \prod_{k=1}^{K-1} (\tau_{\ell(k),\ell(k+1)})^{-\theta \alpha_k} T_{\ell(k)}^{\alpha_k}$ where τ_{ij} is an iceberg cost of transporting goods from country i to country j, and T_i is a productivity index for country i. Aggregate trade shares from country o to d are obtained by summing $\pi_{\ell d}$ over production paths with last production stage in country o—i.e., $\ell(K) = o$.

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

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

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

O.2.5 Mixed CES

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

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

where F is a cumulative distribution function on $\mathbb{R}_+ \times \mathbb{R}^M$ and $\text{Geo}_{od} \in \mathbb{R}^M$ 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 $K \to \infty$:

$$G^d(x_1, \dots, x_N) = \sum_{k=1}^{\infty} \left(\sum_{o=1}^{N} (\omega_{kod} x_o)^{\frac{1}{1-\rho_k}} \right)^{1-\rho_k} \tag{O.10}$$

where for each $o = 1, \dots, N$ we have $\omega_{kod} \geq 0$ for each $k = 1, 2, \dots$ 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 θ, scales of $\{T_{od}\}_{o=1}^{N}$, and correlation function as in (O.10). 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 $k = 1, 2, \dots$ 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.

O.3 GEV Approximation

Proposition O.1 (GEV Approximation). *Let $\{Z_{od}(v)\}_{o=1}^{N}$ have any multivariate distribution whose marginals have finite moment of order $\eta - 1 > 0$. Denote the import demand system implied* *by the Ricardian model with productivity distributed the same as* $\{Z_{od}(v)\}_{o=1}^N$ *by* $\{\pi_{od}(P_{1d},\ldots,P_{Nd})\}_{o=1}^N$. *Then, for any compact* $K \subset \mathbb{R}_+^{N+1}$ *and any* $\epsilon > 0$*, there exists a GEV import demand system captured by a* θ *and* G^d *such that*

$$\sup_{(P_{1d},\ldots,P_{Nd})\in K} \left| \pi_{od}(P_{1d},\ldots,P_{Nd}) - \frac{P_{od}^{-\theta} G_o^d(P_{1d}^{-\theta},\ldots,P_{Nd}^{-\theta})}{G^d(P_{1d}^{-\theta},\ldots,P_{Nd}^{-\theta})} \right| < \epsilon \quad \forall o = 1,\ldots,N.$$

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

$$X_d(v) = \left(\frac{W_o/Z_{od}(v)}{P_d}\right)^{1-\eta} X_d.$$

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

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

so that we can write the expenditure share as

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

Define import price indices as

$$P_{od} = \left[\int_0^1 (W_{o'}/Z_{o'd}(v))^{1-\eta}\,\mathrm{d}v\right]^{\frac{1}{1-\eta}} = W_o/\bar{Z}_{od},$$

for $\bar{Z}_{od} \equiv \left[\int_0^1 Z_{o'd}(v)^{\eta-1}\mathrm{d}v\right]^{\frac{1}{\eta-1}}$. Then

$$\begin{aligned}
\pi_{od} &= \frac{\mathbb{E}\left[(P_{od}\bar{Z}_{od}/Z_{od}(v))^{1-\eta} \mathbf{1}\left\{P_{od}\bar{Z}_{od}/Z_{od}(v) = \min_{o'} P_{o'd}\bar{Z}_{o'd}/Z_{o'd}(v)\right\}\right]}{\mathbb{E}\left[\left(\min_{o'} P_{o'd}\bar{Z}_{o'd}/Z_{o'd}(v)\right)^{1-\eta}\right]} \\
&\equiv \pi_{od}(P_{1d},\ldots,P_{Nd}).
\end{aligned}$$

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

Consider an approximating GEV import demand system that comes from multiplying productivity by independent Fréchet noise. Specifically, replace productivity by the random vector $\{\gamma^{-1}Z_{od}(v)U_{od}(v)\}_{o=1}^{N}$ where $U_{od}(v)$ is some θ-Fréchet noise with unit scale that is independent across o and independent of $\{Z_{od}(v)\}_{o=1}^{N}$ with $\gamma = \Gamma\left(\frac{\theta+1-\eta}{\theta}\right)^{\frac{\theta}{1-\eta}}$. Under this modified productivity distribution, potential import prices are $P_{od}(v) = \gamma W_o/(Z_{od}(v)U_{od}(v))$ and $\{P_{od}(v)^{1-\eta}\}_{o=1}^{N} \mid \{Z_{od}\}_{o=1}^{N}$ is $\theta/(\eta-1)$-Fréchet with scale $(\gamma^{-1}Z_{od}(v)/W_o)^{\theta}$ and independent across o by Lemma O.2. As a consequence, $P_d(v)^{1-\eta} \mid \{Z_{od}\}_{o=1}^{N}$ is $\theta/(\eta-1)$-Fréchet and has scale $\sum_{o=1}^{N}(\gamma^{-1}Z_{od}(v)/W_o)^{\theta}$. The associated price level is

$$
\begin{aligned}
P_d &= \left\{\mathbb{E}\left[\mathbb{E}\left(P_d(v)^{1-\eta} \mid \{Z_{od}(v)\}_{o=1}^{N}\right)\right]\right\}^{\frac{1}{1-\eta}} \\
&= \mathbb{E}\left[\Gamma\left(\frac{\theta+1-\eta}{\theta}\right)\left(\sum_{o=1}^{N}(\gamma^{-1}Z_{od}(v)/W_o)^{\theta}\right)^{\frac{\eta-1}{\theta}}\right]^{\frac{1}{1-\eta}} \\
&= \mathbb{E}\left[\left(\sum_{o=1}^{N}(Z_{od}(v)/\bar{Z}_{od})^{-\theta}P_{od}^{-\theta}\right)^{\frac{\eta-1}{\theta}}\right]^{\frac{1}{1-\eta}} \\
&= G^{d}(P_{1d}^{-\theta},\dots,P_{Nd}^{-\theta};\theta)^{-\frac{1}{\theta}},
\end{aligned}
$$

for

$$
G^{d}(x_1,\dots,x_N;\theta) \equiv \left[\mathbb{E}\left(\sum_{o}(Z_{od}(v)/\bar{Z}_{od})^{\theta}x_o\right)^{\frac{\eta-1}{\theta}}\right]^{\frac{\theta}{\eta-1}}.
$$

Note that $G^{d}(0,\dots,0,1,0,\dots,0;\theta) = \left[\mathbb{E}(Z_{od}(v)/\bar{Z}_{od})^{\eta-1}\right]^{\frac{\theta}{\eta-1}} = 1$.

This price level is identical to assuming that productivity is θ-Fréchet with scale $\bar{Z}_{od}^{\theta}$ and correlation function G^{d}. It also approximates the true price level. In particular,

$$
\begin{aligned}
P_d &= \left[\mathbb{E}\left(\sum_{o}(Z_{od}(v)/W_o)^{\theta}\right)^{\frac{\eta-1}{\theta}}\right]^{\frac{1}{1-\eta}} \\
&\overset{\theta\to\infty}{\to} \left[\mathbb{E}\left(\max_{o} Z_{od}(v)/W_o\right)^{\eta-1}\right]^{\frac{1}{1-\eta}} = \left[\mathbb{E}\left(\min_{o} W_o/Z_{od}(v)\right)^{1-\eta}\right]^{\frac{1}{1-\eta}}.
\end{aligned}
$$

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

The implied GEV import demand system is

$$\frac{P_{od}^{-\theta} G_o^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}; \theta)}{G^d(P_{1d}^{-\theta}, \ldots, P_{Nd}^{-\theta}; \theta)} = \frac{\mathbb{E}\left[\left(\sum_{o'} (Z_{o'd}(v)/W_{o'})^\theta\right)^{\frac{\eta-1}{\theta}-1} (Z_{od}(v)/W_o)^\theta\right]}{\mathbb{E}\left(\sum_{o'} (Z_{o'd}(v)/W_{o'})^\theta\right)^{\frac{\eta-1}{\theta}}}$$
$$\overset{\theta\to\infty}{\to} \frac{\mathbb{E}\left[(W_o/Z_{od}(v))^{1-\eta} \mathbf{1}\left\{W_o/Z_{od}(v) = \min_{o'} W_{o'}/Z_{o'd}(v)\right\}\right]}{\mathbb{E}\left[\left(\min_{o'} W_{o'}/Z_{o'd}(v)\right)^{1-\eta}\right]}$$
$$= \frac{\mathbb{E}\left[(W_o/Z_{od}(v))^{1-\eta} \mathbf{1}\left\{P_{od}\bar{Z}_{od}/Z_{od}(v) = \min_{o'} P_{o'd}\bar{Z}_{o'd}/Z_{o'd}(v)\right\}\right]}{\mathbb{E}\left[\left(\min_{o'} P_{o'd}\bar{Z}_{o'd}/Z_{o'd}(v)\right)^{1-\eta}\right]}$$
$$= \pi_{od}(P_{1d}, \ldots, P_{Nd}).$$

That is, the implied GEV import demand system converges point-wise to the true demand system. The remainder of the proof is identical to Dagsvik (1995). To establish uniform convergence on $K \subset \mathbb{R}_+^{N+1}$ compact, note that if $(P_{1d}, \ldots, P_{Nd}) \mapsto \frac{P_{od}^{-\theta_j} G_o^d(P_{1d}^{-\theta_j}, \ldots, P_{Nd}^{-\theta_j}; \theta_j)}{G^d(P_{1d}^{-\theta_j}, \ldots, P_{Nd}^{-\theta_j}; \theta_j)}$ converges as $j \to \infty$ then there exists a positive sequence $\{\theta_k\}_{k=1}^\infty$ that diverges such that $(P_{1d}, \ldots, P_{Nd}) \mapsto \frac{P_{od}^{-\theta_k} G_o^d(P_{1d}^{-\theta_k}, \ldots, P_{Nd}^{-\theta_k}; \theta_k)}{G^d(P_{1d}^{-\theta_k}, \ldots, P_{Nd}^{-\theta_k}; \theta_k)}$ is monotone and converges as $k \to \infty$. Then, since $\pi_{od}(P_{1d}, \ldots, P_{Nd})$ is continuous, we can apply Theorem 7.13 in Rudin et al. (1964) to establish uniform convergence. □

O.4 Exact Hat-Algebra

We now show how to apply exact hat-algebra methods to solve for a change from the current (observed) equilibrium to any counterfactual equilibrium. First, we define the model's equilibrium.

Definition O.6 (Competitive Equilibrium). *Given endowments $\{L_o\}_{o=1}^N$, trade imbalances $\{TB_d\}_{d=1}^N$, scale parameters $\{T_{od}\}_{o,d=1}^N$, and correlation functions $\{G^d\}_{d=1}^N$, a competitive equilibrium consists of wages $\{W_o\}_{o=1}^N$, and expenditure shares $\{\pi_{od}\}_{o,d=1}^N$ such that*

1. *Expenditure shares satisfy*

$$\pi_{od} = \frac{T_{od} W_o^{-\theta} G_o^d(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta})}{G^d(T_{1d} W_1^{-\theta}, \ldots, T_{Nd} W_N^{-\theta})};$$

2. *The labor market in o clears*

$$W_o L_o = \sum_{d=1}^N \pi_{od} X_d; \text{ and}$$

3. *The resource constraint in each destination d holds*

$$W_d L_d \equiv Y_d = X_d + TB_d.$$

GEV factor demand systems satisfy strict gross substitutability. As a result, the existence and uniqueness of the equilibrium follows from standard results in general equilibrium theory, as proved next.

Proposition O.2 (Existence and Uniqueness). *Assume that expenditure in each country is always strictly positive, productivity is distributed multivariate θ-Fréchet, and markets are perfectly competitive. Then, there exists a competitive equilibrium. The equilibrium is unique up to the choice of numeraire.*

Proof. Define the excess demand function $E : \mathbb{R}_+^N \to \mathbb{R}^N$ as satisfying

$$E_o(\mathbf{W}) = -W_o L_o + \sum_{d=1}^{N} \pi_{od}(W_d L_d - TB_d) \quad \text{for each} \quad o = 1, \ldots, N.$$

Since productivity is multivariate θ-Fréchet, by Lemma O.3, G^d is a correlation function. As a result, we must have $\pi_{od} > 0$ for any finite competitiveness indices.

The implication is that the excess demand system satisfies strict gross substitutability. For each $o = 1, \ldots, N$ and each $n \neq o$ we have

$$\begin{aligned}
\frac{\partial E_o(\mathbf{W})}{\partial W_n} &= \sum_{d=1}^{N} \frac{\partial}{\partial W_n} \frac{T_{od} W_o^{-\theta} G_o^d}{G^d} (W_d L_d - TB_d) \\
&= \sum_{d=1}^{N} \frac{T_{od} W_o^{-\theta}}{G^d} \underbrace{\left(G_{on}^d - \frac{G_o^d G_n^d}{G^d} \right)}_{\leq 0} \underbrace{(W_d L_d - TB_d)}_{=X_d > 0} \underbrace{\frac{\partial P_{Nd}^{-\theta}}{\partial W_n}}_{<0} + \frac{T_{on} W_o^{-\theta} G_o^n}{G^d} L_n \\
&\geq \frac{P_{on}^{-\theta} G_o^n}{G^d} L_n = \pi_{on} L_n > 0.
\end{aligned}$$

The first inequality in the second line follows from the differentiability restriction on the correlation function. The final strict inequality follows from $\pi_{on} > 0$. Given that the excess demand function is homogenous of degree one and satisfies strict gross substitutability, we can apply Proposition 17.F.3 of Mas-Collell et al. (1995) to establish existence and uniqueness. □

Next, we solve for the equilibrium using exact hat-algebra methods (see Costinot and Rodrìguez-Clare, 2014). As a first step, we use the results in Section 3.2 in the paper to solve for correlation-adjusted trade shares, given the structure of G^d and data on bilateral expenditure,

$$\pi_{od} = \widetilde{\pi}_{od} G_o^d(\widetilde{\pi}_{1d}, \ldots, \widetilde{\pi}_{Nd}).$$

In the second step, for a given counterfactual shock, such as $\{\widehat{T}_{od}\}_{o,d=1}^N$, for $d \neq o$, we solve for $\{\widehat{W}_o\}_{o=1}^N$ from

$$\widehat{W}_o Y_o = \sum_{d=1}^N \widehat{\pi}_{od} \pi_{od} (\widehat{W}_d Y_d - TB_d) \quad \text{for each} \quad o = 1, \ldots, N,$$

where

$$\widehat{\pi}_{od} \pi_{od} = \frac{\widehat{T}_{od} \widehat{W}_o^{-\theta} \widetilde{\pi}_{od} G_o^d(\widehat{T}_{1d} \widehat{W}_1^{-\theta} \widetilde{\pi}_{1d}, \ldots, \widehat{T}_{Nd} \widehat{W}_N^{-\theta} \widetilde{\pi}_{Nd})}{G^d(\widehat{T}_{1d} \widehat{W}_1^{-\theta} \widetilde{\pi}_{1d}, \ldots, \widehat{T}_{Nd} \widehat{W}_N^{-\theta} \widetilde{\pi}_{Nd})},$$

and $\widehat{L}_o = \widehat{TB}_d = 1$. Since the equilibrium is unique (up a normalization), we can use a *tâtonnement* process to solve for the equilibrium following Alvarez and Lucas (2007).

After solving for the equilibrium change in wages, we can directly compute the equilibrium change in the price level as

$$\hat{P}_d = \frac{\gamma G^d(\widehat{T}_{1d} \widehat{W}_1^{-\theta} T_{1d} W_1^{-\theta}, \ldots, \widehat{T}_{Nd} \widehat{W}_N^{-\theta} T_{Nd} W_N^{-\theta})^{-\frac{1}{\theta}}}{P_d}.$$

Since $\widetilde{\pi}_{od} = T_{od}(\gamma W_o / P_d)^{-\theta}$ and G^d is homogenous of degree one,

$$\hat{P}_d = G^d(\widehat{T}_{1d} \widehat{W}_1^{-\theta} \widetilde{\pi}_{1d}, \ldots, \widehat{T}_{Nd} \widehat{W}_N^{-\theta} \widetilde{\pi}_{Nd})^{-\frac{1}{\theta}}.$$

O.5 Homothetic Expenditure

As an alternative to CES expenditure shares, consider preferences that imply homothetic expenditure-share functionals, $P_d(\cdot) \mapsto \pi_d(P_d(\cdot); v)$ for each $v \in [0,1]$. Then,

$$\forall v \in [0,1] \;\; X_d(v) = \pi_d(P_d(\cdot); v) X_d,$$

for given deterministic total expenditure, X_d. This specification assumes that consumers in d have homothetic preferences, and they do not have preferences over where they purchase goods. Each expenditure-share functional must satisfy

$$\int_0^1 \pi_d(P_d(\cdot); v) \mathrm{d}v = 1 \quad \text{and} \quad \forall v \in [0,1] \;\; \pi_d(P_d(\cdot); v) \geq 0,$$

for any realization of $P_d(\cdot)$. Additionally, we need to restrict the collection of expenditure-share functionals such that a law of large numbers holds on subsets of the continuum. Formally, for each Borel $B \subset [0,1]$ with positive measure, we need to have that

$$\int_B \pi_d(P_d(\cdot);v)\mathrm{d}v = \mathbb{E}\left[\int_B \pi_d(P_d(\cdot);v)\mathrm{d}v\right].$$

This assumption states that the left-hand side is deterministic. For example, it holds if consumers have unrestricted preferences over CES bundles of varieties.

Then, aggregate expenditure shares only reflect import probabilities. Letting $P_{od}(v) \equiv W_o/Z_{od}(v)$, we have

$$\begin{aligned}
X_{od} &= \int_{V_{od}} X_d(v)\mathrm{d}v = \int_{V_{od}} \pi_d(P_d(\cdot);v)X_d\mathrm{d}v = \mathbb{E}\left[\int_{V_{od}} \pi_d(P_d(\cdot);v)\mathrm{d}v\right] X_d \\
&= \mathbb{E}\left[\int_0^1 \pi_d(P_d(\cdot);v)\mathbf{1}\{v \in V_{od}\}\mathrm{d}v\right] X_d = \int_0^1 \mathbb{E}\left[\pi_d(P_d(\cdot);v)\mathbf{1}\{v \in V_{od}\}\right]\mathrm{d}vX_d \\
&= \int_0^1 \mathbb{E}\left[\pi_d\left(\min_{o'} P_{o'd}(\cdot);v\right)\mathbf{1}\left\{P_{od}(v) = \min_{o'} P_{o'd}(v)\right\}\right]\mathrm{d}vX_d \\
&= \int_0^1 \mathbb{E}\left[\pi_d\left(\min_{o'} P_{o'd}(\cdot);v\right) \mid P_{od}(v) = \min_{o'} P_{o'd}(v)\right]\mathbb{P}\left[P_{od}(v) = \min_{o'} P_{o'd}(v)\right]\mathrm{d}vX_d \\
&= \int_0^1 \mathbb{E}\left[\pi_d\left(\min_{o'} P_{o'd}(\cdot);v\right) \mid P_{od}(v) = \min_{o'} P_{o'd}(v)\right]\mathrm{d}v\mathbb{P}\left[P_{od}(v) = \min_{o'} P_{o'd}(v)\right]X_d \\
&= \int_0^1 \mathbb{E}\left[\pi_d\left(\min_{o'} P_{o'd}(\cdot);v\right)\right]\mathrm{d}v\mathbb{P}\left[P_{od}(v) = \min_{o'} P_{o'd}(v)\right]X_d \\
&= \mathbb{E}\left[\int_0^1 \pi_d(P_d(\cdot);v)\mathrm{d}v\right]\mathbb{P}\left[P_{od}(v) = \min_{o'} P_{o'd}(v)\right]X_d = \mathbb{P}\left[P_{od}(v) = \min_{o'} P_{o'd}(v)\right]X_d,
\end{aligned}$$

where we use the fact that $\pi_d(P_d(\cdot);v) \geq 0$ and Tonelli's Theorem to justify interchanging the integration and expectation operators. The key step in this derivation is moving $\mathbb{P}\left[P_{od}(v) = \min_{o'} P_{o'd}(v)\right]$ outside of the integral over v. This step is justified because $P_{od}(v)$ is i.i.d. over v (so that the import probability is the same across varieties). If the distribution of productivity were not i.i.d. over v, then the composition of demand over sub-intervals would matter. For instance, if $P_{od}(v)$ were i.i.d. within $v \in V_s$ where $\{V_s\}_{s=1}^N$ is a partition of $[0,1]$, we would end up with an expression of the form $X_{od} = \sum_s \mathbb{P}\left[P_{od}(v) = \min_{o'} P_{o'd}(v) \mid V \in V_s\right]\mu_{sd}X_d$ where $\mu_{sd} \equiv \int_{V_s} \pi_d(P_d(\cdot);v)\mathrm{d}v$. In this way, complementarity coming from preferences would lead to complementarity in the factor demand system between, but not within, groups.

O.6 Evidence on Departures from IIA

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.

To estimate these gravity specifications, we use more aggregate sectoral categories. Rather than 4-digit SITC sectors (denoted by s), as for the LFM estimates in the paper, we use 14 aggregate sectoral categories from the World Input-Output Database (WIOD), denoted by j. For comparison purposes, when needed, we can always aggregate our LFM estimates at 4-digit SITC to the WIOD sectoral aggregates.

Adding a time subscript t to the denote years in the period 1999-2007, we estimate the following specification:

$$\pi_{jodt} \equiv \frac{X_{jodt}}{X_{dt}} = \exp\left[D^1_{jot} + D^2_{jdt} + D^3_{jod} + \left(\beta_j + \alpha' Geo_{od}\right) \ln t_{jodt} + \delta' I_{jodt}\right] \nu_{jodt}. \quad \text{(O.11)}$$

The variable t_{jodt} is a tariff index for sector j.[1] 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.

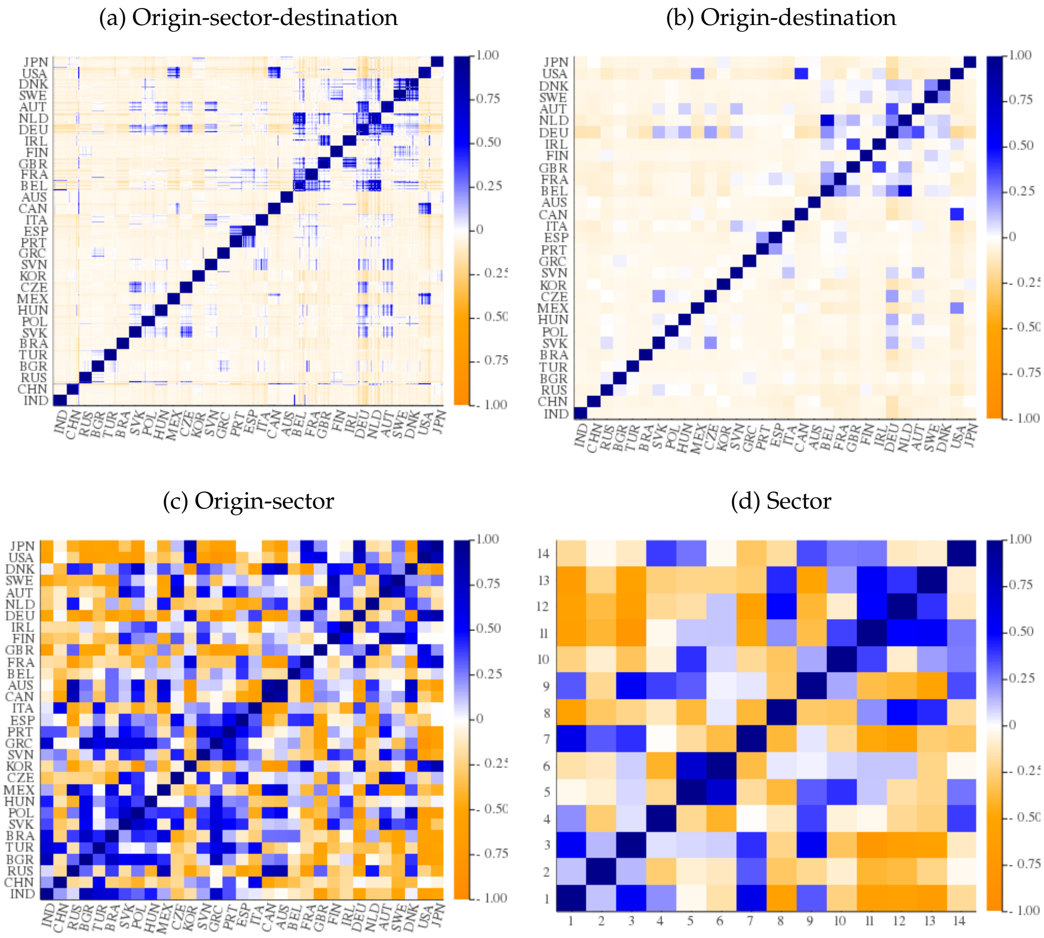
To construct the indices, we use correlation patterns of expenditure observed in the data, shown in Figures O.1b, O.1c, and O.1d. The figures use sector-origin-destination expenditure shares averaged over time, and sort sectors by WIOD classification code and countries by GDP per capita.

Figure O.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

[1] 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.

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



Notes: Each entry shows expenditure correlations: (O.1a) across destinations between a sector-origin pair; (O.1b) across destinations between two origins, $C_{oo'}^{\text{Origin - Geo}}$; (O.1c) across sectors between two origins, $C_{oo'}^{\text{Origin-Sectors}}$; and (O.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.

(to remove correlation induced by sectoral export patterns).

$$I_{jodt}^{\text{Origin-Geo}} = \sum_{o' \neq o} C_{oo'}^{\text{Origin - Geo}} \ln t_{jo'dt}. \tag{O.12}$$

$C_{oo'}^{\text{Origin - Geo}}$ are the entries in Figure O.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 O.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).

$$I_{jodt}^{\text{Origin-Sector}} = \sum_{o' \neq o} C_{oo'}^{\text{Origin-Sector}} \ln t_{jo'dt}. \tag{O.13}$$

$C_{oo'}^{Origin-Sector}$ are the entries in Figure O.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.

$$I_{jodt}^{\text{Sector}} = \sum_{j' \neq j} C_{jj'}^{\text{Sector}} \ln t_{j'odt}. \tag{O.14}$$

$C_{jj'}^{Sector}$ denotes the entries in Figure O.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 O.1 presents PPML estimates of (O.11). If the SGM is correctly specified, we should find that $\alpha = \delta = 0$ in (O.11). If the CES model is correctly specified, we should further find that elasticities are the same across sectors.

In column 1, we restrict 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 equation (19) in the paper. In this case, $\alpha = \delta = 0$ and $\beta_j = -\theta = -2.63$ for all j.[2]

In columns 2 to 5, we allow for β_j to be heterogenous across sectors. The estimates in column 2 correspond to structural estimates of the SGM where $\beta_j = -\sigma_j$, with $\sigma_j \neq \sigma_{j'}$, 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.[3] The Wald test strongly rejects that elasticities are equal across sectors.

Columns 3-5 add tariff interactions and our indices. In this case, t_{jodt} is deflated by the sector mean because the inclusion of sector-destination-time fixed effects absorbs that variation.

[2] 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).

[3] The sectoral estimates are in the range of the sectoral elasticities estimated in Caliendo and Parro (2015).

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

Dep. variable	$\pi_{jodt} \equiv X_{jodt}/X_{dt}$				
	(1)	(2)	(3)	(4)	(5)
β	-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)
$\lvert \ln Y_{ot} - \ln Y_{dt} \rvert \times \bar{t}_{jodt}$			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}^{Sector}				1.13***	0.79***
				(0.175)	(0.173)
$\lvert \ln Y_{ot} - \ln Y_{dt} \rvert$	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		$\beta_j = \beta$	$\alpha = 0$	$\delta = 0$	$\alpha = \delta = 0$
χ^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 (O.11). 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}^{Sector} are defined in (O.12), (O.13), and (O.14). 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 Table O.5. †: 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.

The coefficients on the tariff interactions are interpreted relative to the sectoral average. 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.

The results in Table O.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-origin substitution within a sector rather than with within-origin cross-sector 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 expenditure 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 LFM identifying assumption does — countries load on sectors through common weights λ_{sk}.

We provide further reduced-form support for the LFM assumption in equation (21) in the paper by performing a principal-component analysis that predicts π_{jodt} based on decompositions of the average expenditure share across destinations d. The principal-component structures are: $\pi_{jodt}^{\text{OPC}} = \sum_{k=1}^{K} \lambda_{ok}^{\text{OPC}} \phi_{kjdt}^{\text{OPC}}$ (origins load on latent sector-destination specific factors); and $\pi_{jodt}^{\text{SPC}} = \sum_{k=1}^{K} \lambda_{jk}^{\text{SPC}} \phi_{kodt}^{\text{SPC}}$ (sectors load on latent origin-destination specific factors). The factor weights $\lambda_{ok}^{\text{OPC}}$ are the right eigenvalues of the matrix of average expenditure across destinations, while $\lambda_{jk}^{\text{SPC}}$ 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 ϕ_{kjdt}^{OPC} and ϕ_{kodt}^{SPC}. Their predicted values explain, respectively, 19.9 and 90.7 percent of the variation in π_{jodt}. This analysis reveals that a structure where

Table O.2: LFM Selection: Likelihood Ratio Test. Extended results.

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
χ^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 (29) 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 restricted factor weights as in the sectoral gravity model (SGM). †: Model's degrees of freedom. Last panel shows likelihood ratio tests comparing specifications across columns.

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 sector-destination-time specific latent factors.

O.7 Fit and Validation Tests for LFM

We provide additional evidence on the fit of the LFM. Our estimated model fits the expenditure data very well if we aggregate sectors from 4-digit SITC to the WIOD level, as shown in Table O.2. To compare the fit of LFM with the SGM, we estimate an LFM model with $K = 14$, which are the same number of sectors used in Table O.1 for SGM. The likelihood-ratio test shows that this model is not significantly different from our baseline estimate of LFM with $K = 7$. That said, although $K = 14$ is statistically indistinguishable from the LFM with $K = 7$, it is significantly different from LFM with $K = 14$ *plus* Λ constrained to match the restrictions of SGM at the WIOD sector level, which we denote by $K = 14^r$.[4] 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 $K = 7$.

As a validation test of our LFM estimates, in Table O.3, we go back to the same gravity-type regressions as in Table O.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?

[4]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 O.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^{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	$\alpha = \delta = 0$	$\alpha = 0$	$\delta = 0$	$\alpha = \delta = 0$	$\alpha = \delta = 0$
χ^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 (O.11) augmented by LFM predictions. Number of observations = 121,086. j refers to a WIOD sectoral category. Column 2 corresponds to column 5 in Table O.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^{Sector}_{jodt} are defined in (O.12), (O.13), and (O.14). 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.

Overall, LFM succeeds in capturing those patterns: Both tariff interactions as well as the origin-based indices are not significant, while the magnitude of the effect of the index capturing cross-sector correlation, I^{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.

O.8 Alternative Estimation for the Parameter θ

The alternative estimation of the parameter θ 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

Table O.4: Alternative Estimates of the Parameter θ. PPML.

Dep. variable	Between-factor π^B_{kodt}		
	(1)	(2)	(3)
$\ln t^*_{kodt}$	-1.168	-0.939*	-0.935*
	(1.763)	(0.416)	(0.416)
$\ln \hat{Z}_{kodt}$	Yes	Yes	No
$k \times \ln \hat{Z}_{kodt}$	No	No	Yes
$k \times o \times t$	Yes	Yes	Yes
$d \times t$	Yes	Yes	Yes
$o \times d$	Yes	No	No
$k \times o \times d$	No	Yes	Yes
Observations	60,542	60,542	60,542
Degrees of freedom	57,347	58,307	58,301
Deviance	1,632	122.3	122.2
χ^2	0.20	1.83	1.81
P-value	0.65	0.17	0.17

Notes: Estimates of (O.16). Robust standard errors in parenthesis, clustered by $k \times d$, 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 θ.

and inferred within-factor relative prices from the LFM estimates.

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

$$\pi^B_{kdt} = \frac{\left[\sum_{o'=1}^{N} (t^*_{ko'dt} W_{o't}/A_{ko'dt})^{-\sigma_k}\right]^{\frac{\theta}{\sigma_k}}}{\sum_{k'=1}^{K} \left[\sum_{o'=1}^{N} (t^*_{k'o'dt} W_{o't}/A_{k'o'dt})^{-\sigma_{k'}}\right]^{\frac{\theta}{\sigma_{k'}}}} \equiv \left(\frac{P^*_{kdt}}{P^*_{dt}}\right)^{-\theta}. \tag{O.15}$$

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 θ from the coefficient on the tariff index t^*_{kodt} in

$$\pi^B_{kdt} = \exp\left(-\theta \ln t^*_{kodt} + D^1_{kot} + D^2_{dt} + D^3_{kod} - \theta \ln \hat{Z}^*_{kodt}\right) u_{kodt}, \tag{O.16}$$

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 O.4. The Wald test in the last row indicates that estimates are statistically indistinguishable from our baseline estimate of

$\theta = 0.375$.

O.9 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).

O.9.1 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).[5] 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,

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

Table O.5: WIOD sectoral categories, SGM elasticities, and sectoral expenditure.

Code	Name	Sectoral elasticity	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. Sectoral elasticity from estimating by PPML a sectoral gravity equation using WIOD sectoral aggregation (see Online Appendix O.6). 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 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.

O.9.2 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.Finally, we keep the 14 WIOD aggregates that correspond to traded goods. Table O.5 lists the sectors and their code.

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.

O.9.3 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.

O.10 Latent-Factor Model Estimation: Algorithm

We do not observe all sectors in (28). Additionally, we need to account for observed tariffs, and simultaneously estimate of σ_k for $k = 1, \ldots, K$. The presence of missing data requires to use an adjusted version of (29), which we describe in Section O.10.2. 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 O.10.3.

O.10.1 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 (Λ, Φ^*) 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 Λ and the rows of Φ^*. That is, the matrices of any other factorization can be written as ΛR^{-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 Λ (or columns of Φ^*), 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 Λ', $\text{cone}(\Lambda') = \{\Lambda' x \mid x \in \mathbb{R}_+^S\}$, which is the subset of $\mathbb{R}_+^K$ consisting of positive linear combinations of the rows of Λ. 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 Λ correspond to how each sector, s, loads on each factor, k. In particular, it assumes that factors do not share sectors, forcing cone(Λ') to entirely fill the positive orthant.

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

First, the diagonal of the orthogonality constraint normalizes the scale of each column of Λ, removing the scale indeterminacy of the factorization. Second, the off-diagonal entries force the columns of Λ to be mutually orthogonal. Since these columns have only non-negative entries, there can never be an s such that λ_{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 Λ 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 Λ lies along an axis of $\mathbb{R}^K_+$—it is a scaled standard basis vector. Geometrically, this means that the rows of Λ are maximally spread out in $\mathbb{R}^K_+$, implying that 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 Λ 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 cone$(\Lambda') = \mathbb{R}^K_+$ without requiring all rows of Λ 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 Λ being "spread out" in $\mathbb{R}^K_+$.

Definition O.7 (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*(Λ').

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 Λ (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 Λ that lie outside of $\mathcal{C}$, not all rotations become possible as they will eventually hit the facets of $\mathbb{R}_+^K$. When Λ is sufficiently scattered, the rows of Λ 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(Λ') that intersect the faces of $\mathbb{R}_+^K$ (ruling our small rotations), while the second is a regularity condition that means that cone(Λ') 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 O.2 (Huang et al. (2014)). *If Λ and $\Phi^{*'}$ are sufficiently scattered, then (Λ, Φ^*) is identified.*

If we view the rows of Λ (columns of Φ^*) 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(Λ') 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 Λ and Φ^* 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, $\mathbb{R}_+^K$.

O.10.2 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.

O.10.3 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

$$\mathcal{L} = \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{\ell(Y_{it}, \hat{Y}_{it})}{N_{it}}, \tag{O.17}$$

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

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

We can therefore write the gradient in Λ_k as

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

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

$$\Lambda_k \leftarrow \Lambda_k \odot \frac{\left(\frac{\mathbf{Y}}{\hat{\mathbf{Y}}} \odot \mathbf{Z}^{-\sigma_k}\right) \Phi_k^*}{(\mathbf{Z}^{-\sigma_k}) \Phi_k^*}. \tag{O.18}$$

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

Similarly, we get an updating rule for Φ_k^* given by

$$\Phi_k^* \leftarrow \Phi_k^* \odot \frac{\left(\frac{\mathbf{Y}}{\hat{\mathbf{Y}}} \odot \mathbf{Z}^{-\sigma_k}\right)' \Lambda_k}{(\mathbf{Z}^{-\sigma_k})' \Lambda_k}. \tag{O.19}$$

Finally, the derivative in σ_k is

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

The implied updating rule is

$$\sigma_k \leftarrow \sigma_k \odot \frac{\mathbf{1}'[\mathbf{Z}^{-\sigma_k} \odot (\Lambda_k \Phi_{k'}^*) \odot \ln \mathbf{Z}]\mathbf{1}}{\mathbf{1}' \left[\frac{\mathbf{Y}}{\hat{\mathbf{Y}}} \odot \mathbf{Z}^{-\sigma_k} \odot (\Lambda_k \Phi_{k'}^*) \odot \ln \mathbf{Z}\right] \mathbf{1}}. \tag{O.20}$$

Using the proof technique in Lee and Seung (2001), one can show that (O.17) is monotonically decreasing in any of (O.18), (O.19), and (O.20). 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 (O.17) as our estimate.

O.11 Additional Quantitative Results

Table O.6: Factor Weights: Sharing and Similarity.

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

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

Table O.7: Elasticity estimates: LFM with different number of factors.

	Number of factors K							
	1	2	3	4	5	6	7	8
σ_1	3.003	3.933	3.300	4.814	3.929	7.866	5.175	9.944
σ_2		2.767	2.638	3.342	3.780	3.536	4.868	5.471
σ_3			1.592	2.614	3.573	2.559	4.624	4.417
σ_4				1.223	0.806	0.804	1.481	3.435
σ_5					0.418	0.574	0.670	1.884
σ_6						0.163	0.390	1.594
σ_7							0.375	0.111
σ_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 σ_k	3.003	3.350	2.510	2.998	2.501	2.584	2.512	3.371

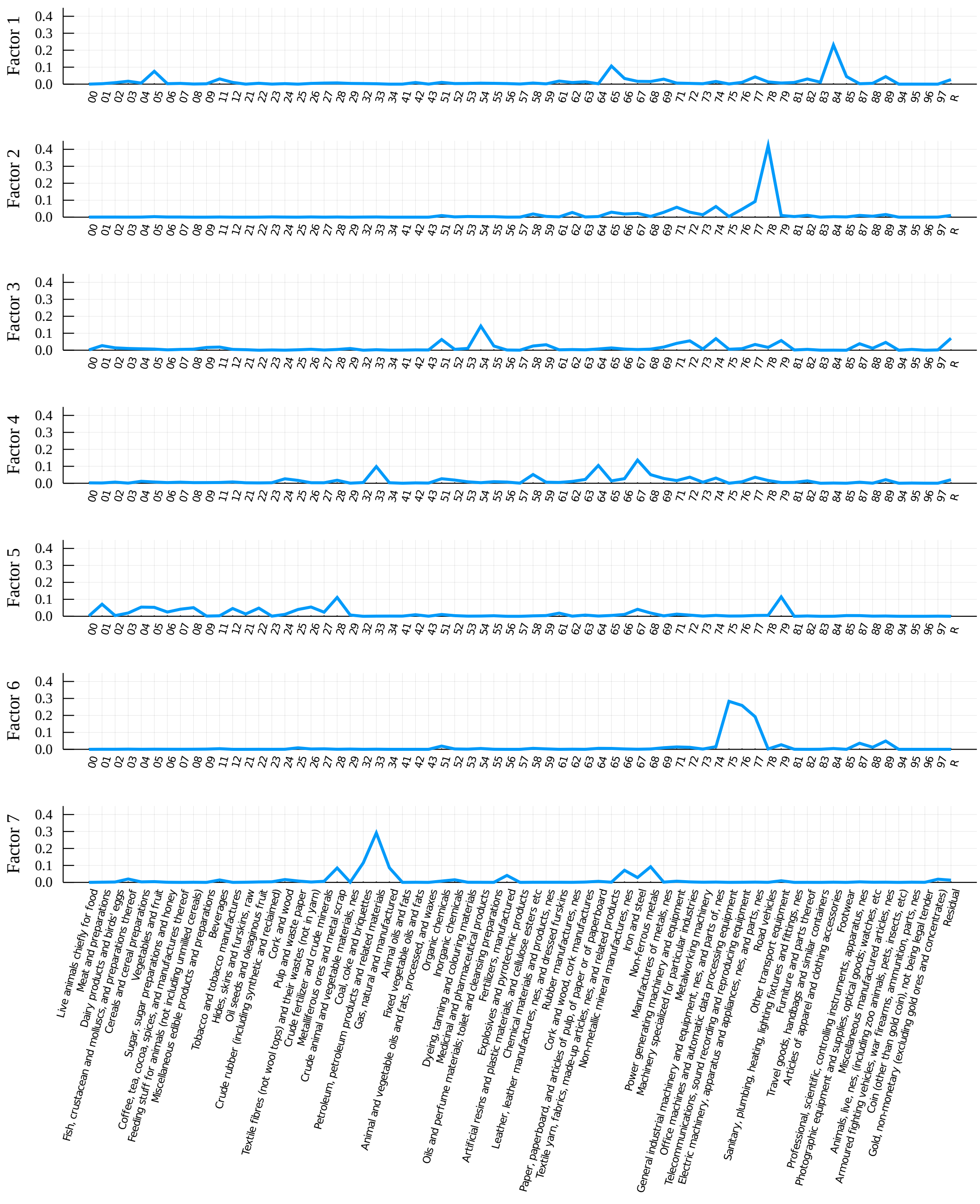
Notes: Estimates of factor-level elasticities, σ_k, for latent-factor models (LFM) with K = 1,. . . , 8. In each case, $F1$ is the factor with the highest elasticity, while FK is the one with the lowest, with $\theta = \sigma_K$.

Table O.8: 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 O.2: Factor Weights: Two-Digit SITC Sectors.



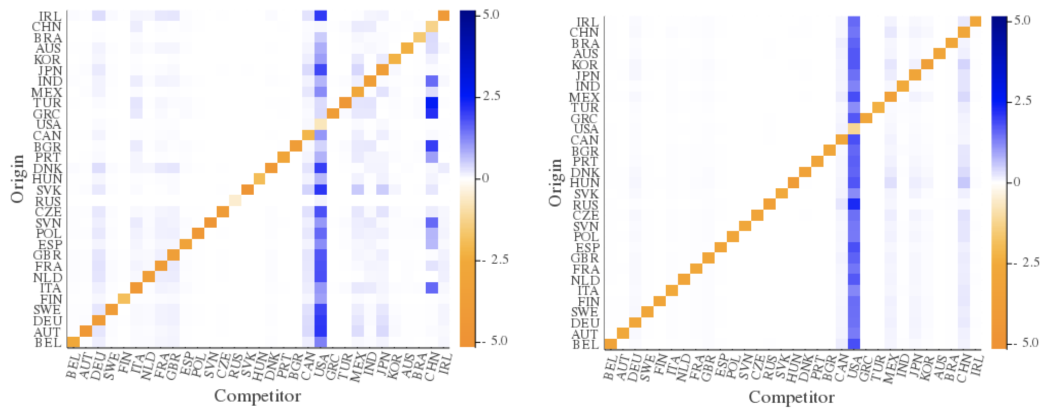
Notes: Estimates of factor weights across 4-digit SITC sectors, aggregated to 2-digit SITC sectoral level.

Figure O.3: Similarity of Factor Use Across Two-Digit SITC Sectors.

Notes: Similarity refers to $\sum_k \lambda_{hk}\lambda_{h'k}/\sqrt{\sum_k \lambda_{hk}^2 \sum_k \lambda_{h'k}^2}$ and λ_{hk} is constructed from estimates of λ_{sk}. h refers to a 2-digit SITC sector, and s refers to a 4-digit SITC sector.

Figure O.4: 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 (30) and estimates from the latent-factor model (LFM) and sectoral gravity model (SGM). Year 2007.

Figure O.5: Tariff effects, densities.

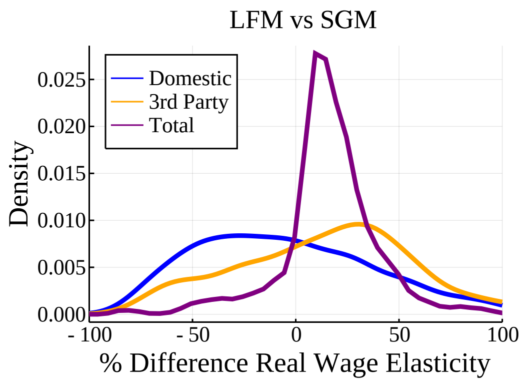


Table O.9: Tariff effects, moments.

%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 O.5 shows density plots of the percent difference in the components of (32) 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 O.9 shows moments of these densities. Year 2007.

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.

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

Antràs, P. and A. de Gortari (2020). On the geography of global value chains. *Econometrica 84*(4), 1553–1598.

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.

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

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

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.

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.

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

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.

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

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.

Mas-Collell, A., M. Whinston, and J. R. Green (1995). *Microeconomic Theory*. Oxford University Press.

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.

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

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.