

Matching to Suppliers in the Production Network: a Quantitative Framework

By ALONSO ALFARO-UREÑA AND PAOLO ZACCHIA*

ONLINE APPENDIX

<i>List of appendices</i>	<i>Page</i>
A. Proposition proofs	<i>A-2</i>
B. Extended analysis of the model	<i>A-12</i>
C. Data and supporting stylized facts	<i>A-23</i>
D. Supplement to the empirical application	<i>A-34</i>
References	<i>A-45</i>

* Alfaro-Ureña: Banco Central de Costa Rica and University of Costa Rica (alfaroua@bccr.fi.cr); Zacchia: CERGE-EI (joint workplace of Charles University, Prague and the Czech Academy of Sciences), Ca' Foscari University of Venice, and CEPR (mail@paolozacchia.com). The processing of data used for this paper's statistical elaborations was conducted at the Banco Central de Costa Rica (BCCR) through a secure technology that guarantees integrity and confidentiality of the information. All results were thoroughly reviewed by the BCCR, ensuring no disclosure of confidential information. All views and mistakes expressed in this paper are solely the authors' and do not extend to the BCCR.

A. Proposition proofs

This appendix provides proofs of the propositions from Sections I and II in the text.

Proof of Proposition 1

We start from characterizing the probability mass function of the production network $\mathbf{G}$ conditional on the sequence of seller fixed effects γ_j and the dyadic characteristics $\mathbf{z}_{ij}$. In general, models of network formation feature multiple equilibria, which prevent specifying a unique joint mass function (let alone conditional). However, Assumption 3 ensures equilibrium uniqueness and consequently, a well-specified joint distribution. Let

$$\begin{aligned}\boldsymbol{\gamma} &= \{\gamma_j\}_{j \in \mathcal{I}}, \\ \boldsymbol{\gamma}_{(s)} &= \{\gamma_j\}_{j: s(j) \leq s}, \\ \mathbf{G}_{(s)} &= \{g_{ij}\}_{i \in \mathcal{S}_s, j \in \mathcal{I}} \quad \text{and} \\ \mathbf{Z}_{(s)} &= \{\mathbf{z}_{ij}\}_{i \in \mathcal{S}_s, j \in \mathcal{I}}.\end{aligned}$$

It holds:

$$(A.1) \quad f_G(\mathbf{G} | \boldsymbol{\gamma}, \mathbf{Z}) = \prod_{s=2}^S f_{G_s}(\mathbf{G}_{(s)} | \boldsymbol{\gamma}_{(s-1)}, \mathbf{Z}_{(s)})$$

a conditional probability factorization that is contingent upon the acyclical structure of the network (given $\boldsymbol{\gamma}_{(s-1)}$, the choices of buyers from sector s do not depend upon

$$(\mathbf{G}_{(1)}, \dots, \mathbf{G}_{(s-1)}),$$

and cannot reverse-cause $\boldsymbol{\gamma}_{(s-1)}$). In particular, for $s = 2, \dots, S$:

$$\begin{aligned}f_{G_s}(\mathbf{G}_{(s)} | \boldsymbol{\gamma}_{(s-1)}, \mathbf{Z}_{(s)}) &= \prod_{i \in \mathcal{S}_s} \prod_{k \in \mathcal{K}_i} \prod_{j \in \mathcal{I}} \left[\mathbb{P}(j^*(i, k) = j | \boldsymbol{\gamma}_{(s-1)}, \mathbf{Z}_{(s)}) \right]^{\mathbb{1}_{[j^*(i, k) = j]}} \\ &= \prod_{i \in \mathcal{S}_s} \prod_{k \in \mathcal{K}_i} \prod_{j \in \mathcal{I}} \left(\frac{\exp(\gamma_j + \mathbf{z}'_{ij} \boldsymbol{\beta})}{\sum_{\ell \in \mathcal{S}_{z(k)}} \exp(\gamma_\ell + \mathbf{z}'_{i\ell} \boldsymbol{\beta})} \right)^{\mathbb{1}_{[j^*(i, k) = j]}} \\ &= \prod_{i \in \mathcal{S}_s} \prod_{j \in \mathcal{I}} \left(\frac{\exp(\gamma_j + \mathbf{z}'_{ij} \boldsymbol{\beta})}{\sum_{\ell \in \mathcal{S}_{s(j)}} \exp(\gamma_\ell + \mathbf{z}'_{i\ell} \boldsymbol{\beta})} \right)^{g_{ij}},\end{aligned}$$

where, upon conditioning on both $\boldsymbol{\gamma}_{(s-1)}$ and $\mathbf{Z}_{(s)}$, the second equality follows from Assumptions 4-5, while the third one follows from Assumptions 1-2 and the

definition of g_{ij} . Collecting terms, the conditional mass function of the network is as follows:

$$(A.2) \quad f_G(\mathbf{G}|\boldsymbol{\gamma}, \mathbf{Z}) = \prod_{i \in \mathcal{I}} \prod_{j \in \mathcal{I}} \left(\frac{\exp(\gamma_j + \mathbf{z}'_{ij} \boldsymbol{\beta})}{\sum_{\ell \in \mathcal{S}_{s(j)}} \exp(\gamma_\ell + \mathbf{z}'_{i\ell} \boldsymbol{\beta})} \right)^{g_{ij}}.$$

From (A.2) one can derive the mass function of the out-degree sequence $\mathbf{d}$:

$$(A.3) \quad \begin{aligned} f_d(\mathbf{d}|\boldsymbol{\gamma}, \mathbf{Z}) &= \sum_{\mathbf{H} \in \mathcal{H}} f_G(\mathbf{H}|\boldsymbol{\gamma}, \mathbf{Z}) \\ &= \sum_{\mathbf{H} \in \mathcal{H}} \left[\prod_{i \in \mathcal{I}} \prod_{j \in \mathcal{I}} \left(\frac{\exp(\gamma_j + \mathbf{z}'_{ij} \boldsymbol{\beta})}{\sum_{\ell \in \mathcal{S}_{s(j)}} \exp(\gamma_\ell + \mathbf{z}'_{i\ell} \boldsymbol{\beta})} \right)^{h_{ij}} \right], \end{aligned}$$

where $\mathcal{H}$ is defined as the set of networks that feature the out-degree sequence $\mathbf{d}$, and that are consistent with the assumptions of our model:

$$\mathcal{H} \equiv \left\{ \mathbf{H} : \sum_{i \in \mathcal{I}} (h_{ij} - g_{ij}) = 0 \ \forall j \in \mathcal{I}, \sum_{j \in \mathcal{S}_s} (h_{ij} - g_{ij}) = 0 \ \forall i \in \mathcal{I} \wedge s = 1, \dots, S \right\}.$$

Every matrix $\mathbf{H} \in \mathcal{H}$ corresponds bijectively with a unique collection $\{\mathbf{H}_s\}_{s=1}^S$ where $\mathbf{H}_s \in \mathcal{H}_s$ for $s = 1, \dots, S$, with $\mathbf{H}$ resulting from (re-)joining the columns of all $\mathbf{H}_s$ matrices according to their appropriate ordering along $\mathcal{I}$. Thus, the Cartesian product $\times_{s=1}^S \mathcal{H}_s$ maps $\mathcal{H}$ one-to-one and onto. Armed with this observation, one obtains the joint mass function of the network, conditional on the out-degree sequence, as:

$$(A.4) \quad \begin{aligned} f_{G|\mathbf{d}}(\mathbf{G}|\mathbf{d}, \boldsymbol{\gamma}, \mathbf{Z}) &= \frac{f_{(G, \mathbf{d})}(\mathbf{G}, \mathbf{d}|\boldsymbol{\gamma}, \mathbf{Z})}{f_d(\mathbf{d}|\boldsymbol{\gamma}, \mathbf{Z})} \\ &= \frac{f_G(\mathbf{G}|\boldsymbol{\gamma}, \mathbf{Z})}{f_d(\mathbf{d}|\boldsymbol{\gamma}, \mathbf{Z})} \\ &= \frac{\exp\left(\sum_{j \in \mathcal{I}} D_j \gamma_j\right) \exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}} g_{ij} \mathbf{z}'_{ij} \boldsymbol{\beta}\right)}{\sum_{\mathbf{H} \in \mathcal{H}} \left[\exp\left(\sum_{j \in \mathcal{I}} D_j \gamma_j\right) \exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}} h_{ij} \mathbf{z}'_{ij} \boldsymbol{\beta}\right) \right]} \\ &= \frac{\prod_{s=1}^S \exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} g_{ij} \mathbf{z}'_{ij} \boldsymbol{\beta}\right)}{\sum_{\mathbf{H} \in \mathcal{H}} \left[\prod_{s=1}^S \exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} h_{ij} \mathbf{z}'_{ij} \boldsymbol{\beta}\right) \right]} \\ &= \prod_{s=1}^S \left(\frac{\exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} g_{ij} \mathbf{z}'_{ij} \boldsymbol{\beta}\right)}{\sum_{\mathbf{H}_s \in \mathcal{H}_s} \exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} h_{ij} \mathbf{z}'_{ij} \boldsymbol{\beta}\right)} \right) \end{aligned}$$

where: the second equality follows from the fact that $\mathbf{G}$ implies $\mathbf{d}$, hence the probabilities in the numerators must coincide; the third equality follows from the definition of D_j and the sectoral structure of buyers' choice as prescribed by Assumption 2 and embedded in the definition of $\mathcal{H}$, which allows one to drop the denominators of (A.2) and (A.3); the fourth equality drops the seller fixed effects; lastly, the manipulation of the denominator in the fifth equality follows from the previous observation about $\mathcal{H}$. The likelihood function in (7) derives from (A.4); note that we drop $\boldsymbol{\gamma}$ from the list of conditioned variables as it is implicit from conditioning on $\mathbf{d}$.

Proof of Proposition 2

This is largely an application of existing results to our particular case. Here we focus on proving consistency under sampling schemes of the alternatives with replacement; this extends the original results by McFadden (1978) which was restricted to sampled subsets of the full set of alternatives. To proceed, it is useful to simplify some notation. Recognizing that both (7) and the maximand in (9) represent likelihood functions for the event that some collection of subnetworks is observed, we write for $s = 1, \dots, S$:

$$\Phi(\mathbf{G}_s | \tilde{\mathcal{H}}_s; \mathbf{Z}_s; \boldsymbol{\beta}) = \frac{\exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} g_{ij} \mathbf{z}'_{ij} \boldsymbol{\beta}\right)}{\sum_{\mathbf{H}_s \in \tilde{\mathcal{H}}_s} \exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} h_{ij} \mathbf{z}'_{ij} \boldsymbol{\beta}\right)}$$

where $\tilde{\mathcal{H}}_s$ is some collection of, possibly repeated, elements of $\mathcal{H}_s$ (such as $\mathcal{H}_s$, or $\mathcal{H}_s^*$), while $\mathbf{Z}_s$ is the subset of $\mathbf{Z}$ restricted to those elements $\mathbf{z}_{ij}$ such that $s(j) = s$. The RSL estimator thus maximizes the following log-likelihood function:

$$(A.5) \quad \log \mathcal{L}^*(\boldsymbol{\beta} | \mathbf{d}, \mathbf{G}, \mathbf{Z}) = \frac{1}{S} \sum_{s=1}^S \log \Phi(\mathbf{G}_s | \mathcal{H}_s^*; \mathbf{Z}_s; \boldsymbol{\beta}).$$

We show that the maximum of (A.5) converges to the true value of $\boldsymbol{\beta}$, which we write as $\boldsymbol{\beta}_0$ (not to be mistaken for β_0), as $S \rightarrow \infty$.

Following McFadden (1978), the probability limit of (A.5) for $S \rightarrow \infty$, accounting for $\mathcal{H}_s^*$ being random and for sectors not being identically distributed, is:

$$(A.6) \quad \text{plim} \log \mathcal{L}^*(\boldsymbol{\beta} | \mathbf{d}, \mathbf{G}, \mathbf{Z}) = \lim_{S \rightarrow \infty} \frac{1}{S} \sum_{s=1}^S L(\mathbf{G}_s, \mathcal{H}_s, \mathcal{H}_s^*, \mathbf{Z}_s; \boldsymbol{\beta}),$$

where, for $s = 1, \dots, S$:

$$L(\mathbf{G}_s, \mathcal{H}_s, \mathcal{H}_s^*, \mathbf{Z}_s; \boldsymbol{\beta}) \equiv$$

$$\begin{aligned}
&\equiv \int_{\mathbb{Z}_s} \left[\sum_{\mathcal{H}_s^* \in \mathbb{H}_s} \sum_{\mathbf{G}_s \in \mathcal{H}_s} \mathbb{P}(\mathcal{H}_s^* | \mathbf{G}_s; \mathbf{Z}_s) \Phi(\mathbf{G}_s | \mathcal{H}_s; \mathbf{Z}_s; \boldsymbol{\beta}_o) \log \Phi(\mathbf{G}_s | \mathcal{H}_s^*; \mathbf{Z}_s; \boldsymbol{\beta}) \right] \\
&\quad \cdot dF_Z(\mathbf{Z}_s) \\
&\quad (A.7) \\
&= \int_{\mathbb{Z}_s} \left[\sum_{\mathcal{H}_s^* \in \mathbb{H}_s} \frac{\sum_{\mathbf{H}_s \in \mathcal{H}_s^*} \exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} h_{ij} z'_{ij} \boldsymbol{\beta}_o\right)}{\sum_{\mathbf{H}_s \in \mathcal{H}_s} \exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} h_{ij} z'_{ij} \boldsymbol{\beta}_o\right)} \cdot \right. \\
&\quad \left. \cdot \sum_{\mathbf{G}_s \in \mathcal{H}_s} \mathbb{P}(\mathcal{H}_s^* | \mathbf{G}_s; \mathbf{Z}_s) \Phi(\mathbf{G}_s | \mathcal{H}_s^*; \mathbf{Z}_s; \boldsymbol{\beta}_o) \log \Phi(\mathbf{G}_s | \mathcal{H}_s^*; \mathbf{Z}_s; \boldsymbol{\beta}) \right] dF_Z(\mathbf{Z}_s),
\end{aligned}$$

where $\mathbb{Z}_s$ is the support of $\mathbf{Z}_s$; $F_Z(\mathbf{Z}_s)$ is the cumulative distribution function of $\mathbf{Z}_s$; while $\mathbb{H}_s$ is the collection of all eligible $\mathcal{H}_s^*$ sets. Importantly, $\mathbb{H}_s$ is designed by the econometrician and allows for sampling with replacement. If the uniform conditioning property (10) holds, and regardless of whether subnetwork sampling is with or without replacement, the $\mathbb{P}(\mathcal{H}_s^* | \mathbf{G}_s; \mathbf{Z}_s)$ terms reduce to a constant; thus, (A.6) is understood as a weighted average of terms of the kind:

$$\sum_{\mathbf{G}_s \in \mathcal{H}_s} \Phi(\mathbf{G}_s | \cdot; \boldsymbol{\beta}_o) \log \Phi(\mathbf{G}_s | \cdot; \boldsymbol{\beta})$$

where $\sum_{\mathbf{G}_s \in \mathcal{H}_s} \Phi(\mathbf{G}_s | \cdot; \boldsymbol{\beta}) = 1$. Hence, the maximum of (A.6) is unique and equals $\boldsymbol{\beta}_o$; moreover, by standard arguments the maximum of (A.5) converges in probability to that of (A.6); therefore, the RSL estimator is consistent. This derivation is easily extended to the case in which uniform conditioning does not hold but $\mathbb{P}(\mathcal{H}_s^* | \mathbf{G}_s; \mathbf{Z}_s)$ is known by the econometrician, as in the original analysis by McFadden (1978).

The other statements of the proposition are easily shown by the same arguments as in D'Haultfœuille and Iaria (2016). Since this RSL estimator is a maximum likelihood estimator, it is asymptotically normal with an asymptotic variance-covariance given by the inverse of the information matrix, and a rate of convergence equal to the square root of the number of addends of the log-likelihood function. By the standard algebra of the multinomial logit, the RSL information matrix conditional on $\mathbf{Z}$ is:

$$\begin{aligned}
(A.8) \quad \mathcal{J}_S(\boldsymbol{\beta}_o | \mathbf{Z}) &= \frac{1}{S} \sum_{s=1}^S \mathbb{E}_{\mathbb{H}_s} \left[\sum_{\mathbf{H}_s \in \mathcal{H}_s} \Phi(\mathbf{H}_s | \mathcal{H}_s; \mathbf{Z}_s; \boldsymbol{\beta}_o) \tilde{\mathbf{Z}}_s(\mathbf{H}_s) \tilde{\mathbf{Z}}_s'(\mathbf{H}_s) - \right. \\
&\quad \left. - \sum_{(\mathbf{H}_s, \mathbf{K}_s) \in \mathcal{H}_s^2} \Phi(\mathbf{H}_s | \mathcal{H}_s^*; \mathbf{Z}_s; \boldsymbol{\beta}_o) \Phi(\mathbf{K}_s | \mathcal{H}_s^*; \mathbf{Z}_s; \boldsymbol{\beta}_o) \tilde{\mathbf{Z}}_s(\mathbf{H}_s) \tilde{\mathbf{Z}}_s'(\mathbf{K}_s) \right],
\end{aligned}$$

where $\tilde{\mathbf{Z}}_s(\mathbf{H}_s) \equiv \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} h_{ij} \mathbf{z}_{ij}$ and all S expectations are taken over the $\mathbb{H}_s$ sets. The above is derived from (9) noting that the $\Phi(\mathbf{H}_s | \cdot)$ probabilities are not identically distributed across the S subnetworks, but are conditionally (on $\mathbf{Z}$) independent per Assumption 4. Since the information matrix of the unfeasible maximizer of (7) also follows (A.8) but with singleton $\mathbb{H}_s$ sets which only contain $\mathcal{H}_s$ for $s = 1, \dots, S$, to show that such an estimator is more efficient than any RSL estimator one shall show that the matrices defined as:

$$\mathbf{M}_s(\mathbf{Z}_s) \equiv \sum_{(\mathbf{H}_s, \mathbf{K}_s) \in \mathcal{H}_s^2} \tilde{\mathbf{Z}}_s(\mathbf{H}_s) \tilde{\mathbf{Z}}_s'(\mathbf{K}_s) \Pi(\mathbf{H}_s, \mathbf{K}_s | \mathcal{H}_s^*; \mathbf{Z}_s; \beta_o),$$

where:

$$(A.9) \quad \Pi(\mathbf{H}_s, \mathbf{K}_s | \mathcal{H}_s^*; \mathbf{Z}_s; \beta_o) \equiv \mathbb{E}_{\mathbb{H}_s} \left[\Phi(\mathbf{H}_s | \mathcal{H}_s^*; \mathbf{Z}_s; \beta_o) \Phi(\mathbf{K}_s | \mathcal{H}_s^*; \mathbf{Z}_s; \beta_o) \right] - \Phi(\mathbf{H}_s | \mathcal{H}_s; \mathbf{Z}_s; \beta_o) \Phi(\mathbf{K}_s | \mathcal{H}_s; \mathbf{Z}_s; \beta_o),$$

are symmetric positive matrices for $s = 1, \dots, S$. This follows from some algebraic analysis as the (A.9) terms themselves can be arranged in symmetric positive matrices with the same size as their corresponding $\mathbb{H}_s$ sets; see the supplementary material by D'Haultfoeulle and Iaria (2016) for a more detailed discussion of an analogous case.

Proof of Proposition 3

This is an extension of Proposition 1 which accounts for the nested logit structure of the longitudinal version of the model. Because of possible time dependence in the ε_{ijkt} errors, we separately derive the conditional mass functions $f_{G|b,d}(\mathbf{G}_t^n | \mathbf{b}_t^n, \mathbf{d}_t^n, \boldsymbol{\gamma}_t^n, \mathbf{Z}_t^n)$ for $n = d, f$ and $t = 1, \dots, T$. In doing so, we omit for brevity dependence upon past realizations of the network(s) $\mathbf{G}_t^n$, the fixed effects $\boldsymbol{\gamma}_t^n$, and the dyadic characteristics $\mathbf{Z}_t^n$; in this setting, this is motivated by Assumption 5a (exogenous errors) and 6 (no structural dependence). As we endow the RSL estimator with a QMLE interpretation, we refrain from specifying the full-fledged likelihood function.

Before proceeding, it is useful to establish some additional ancillary notation. For all firms $j \in (\mathcal{I}^d \cup \mathcal{I}^f)$, let $\kappa_{jt} = \gamma_{jt}/\rho_n$, where $n = d$ if $j \in \mathcal{I}^d$ while $n = f$ if $j \in \mathcal{I}^f$. Let $\tilde{m}_{is(k)t} \equiv \beta_0 - \log W_t + m_{is(k)t}$, and let $K_{is} = |\{k \in \mathcal{K}_i : s(k) = s\}|$ be the number of tasks that firm i can only source from outside suppliers of sector s . By the nested logit structure of the model, the probability that a domestic firm i chooses a supplier $j \in \mathcal{I}^n$ for task $k \in \mathcal{K}_i$ at time t for $n = d, f$, conditional on $\mathbf{z}_{ijt}$, κ_{jt} and $\tilde{m}_{is(k)t}$ (all realized at time t), is obtained by Assumptions 4 and 5a

as:

$$\mathbb{P} \left(j^*(i, k, t) = j, j \in \mathcal{I}^n \mid \{(\kappa_\ell, \mathbf{z}_{i\ell t})\}_{\ell \in (\mathcal{S}_{z(k)t}^d \cup \mathcal{S}_{z(k)t}^f)}, \tilde{m}_{is(k)t} \right) = \Psi_{iktn} \Psi_{ijktn|n},$$

where:

$$\begin{aligned} \Psi_{iktn} &\equiv \mathbb{P} \left(j^*(i, k, t) \in \mathcal{I}^n \mid \cdot \right) = \\ &= \frac{\left[\sum_{\ell \in \mathcal{S}_{z(k)t}^n} \exp(\kappa_{\ell t} + \mathbf{z}'_{i\ell t} \boldsymbol{\varphi}^n) \right]^{\rho_n}}{\exp(\tilde{m}_{is(k)t}) + \sum_{\tilde{n} \in \{d, f\}} \left[\sum_{\ell \in \mathcal{S}_{z(k)t}^{\tilde{n}}} \exp(\kappa_{\ell t} + \mathbf{z}'_{i\ell t} \boldsymbol{\varphi}^{\tilde{n}}) \right]^{\rho_{\tilde{n}}}} \end{aligned}$$

is the total probability that firm i chooses *any* supplier in $\mathcal{I}^n$; and:

$$\begin{aligned} \Psi_{ijktn|n} &\equiv \mathbb{P} \left(j^*(i, k, t) = j \mid j^*(i, k, t) \in \mathcal{I}^n; \cdot \right) = \\ &= \frac{\exp(\kappa_{jt} + \mathbf{z}'_{ijt} \boldsymbol{\varphi}^n)}{\sum_{\ell \in \mathcal{S}_{z(k)t}^n} \exp(\kappa_{\ell t} + \mathbf{z}'_{i\ell t} \boldsymbol{\varphi}^n)} \end{aligned}$$

is the probability that j is chosen *conditional* on choosing within $\mathcal{I}^n$. Analogously,

$$\begin{aligned} \Psi_{iktm} &\equiv \mathbb{P} \left(j^*(i, k, t) = \emptyset \mid \cdot \right) = \\ &= \frac{\exp(\tilde{m}_{is(k)t})}{\exp(\tilde{m}_{is(k)t}) + \sum_{\tilde{n} \in \{d, f\}} \left[\sum_{\ell \in \mathcal{S}_{z(k)t}^{\tilde{n}}} \exp(\kappa_{\ell t} + \mathbf{z}'_{i\ell t} \boldsymbol{\varphi}^{\tilde{n}}) \right]^{\rho_{\tilde{n}}}} \end{aligned}$$

is probability of the “make” choice. The expressions of the Ψ_{iktd} , Ψ_{iktf} and Ψ_{iktm} kind are, for any given i and t , identical across all task pairs (k, k') such that $s(k) = s(k')$, and are thus equivalently indexed as Ψ_{istd} , Ψ_{istf} and Ψ_{istm} .

By extending (A.2), one obtains the joint mass function of the networks (domestic and foreign) at $t = 1, \dots, T$, for

$$\begin{aligned} \boldsymbol{\kappa}_t^n &= \{\kappa_{jt}\}_{j \in \mathcal{I}_t^n} \quad \text{and} \\ \mathbf{m}_t &= \{\tilde{m}_{i1t}, \dots, \tilde{m}_{iSt}\}_{i \in \mathcal{I}_t^d}. \end{aligned}$$

As in the case of Proposition 1, such a derivation is upheld by network acyclicity,

as per Assumption 3. Through a factorization analogous to (A.1) one obtains:

$$\begin{aligned}
 (A.10) \quad & f_{G,G} \left(\mathbf{G}_t^d, \mathbf{G}_t^f \mid \boldsymbol{\kappa}_t^d, \boldsymbol{\kappa}_t^f, \mathbf{Z}_t^d, \mathbf{Z}_t^f, \mathbf{m}_t \right) = \\
 & = \prod_{i \in \mathcal{I}_t^d} \prod_{k \in \mathcal{K}_i} \left[(\Psi_{iktm})^{\mathbb{1}[j^* = \emptyset]} \prod_{j \in \mathcal{I}_t^d} (\Psi_{iktd} \Psi_{ijkt|d})^{\mathbb{1}[j^* = j]} \prod_{j' \in \mathcal{I}_t^f} (\Psi_{iktf} \Psi_{ij'kt|f})^{\mathbb{1}[j^* = j']} \right] \\
 & = \prod_{i \in \mathcal{I}_t^d} \prod_{j \in \mathcal{I}_t^d} \Psi_{ijkt|d}^{g_{ijt}} \prod_{i \in \mathcal{I}_t^d} \prod_{j' \in \mathcal{I}_t^f} \Psi_{ij'kt|f}^{g_{ij't}} \cdot \\
 & \quad \cdot \prod_{s=1}^S \prod_{i \in \mathcal{I}_t^d} (\Psi_{istm})^{K_{is} - B_{ist}^d - B_{ist}^f} (\Psi_{istd})^{B_{ist}^d} (\Psi_{istf})^{B_{ist}^f},
 \end{aligned}$$

where j^* is shorthand for $j^*(i, k, t)$. Note that it is not necessary to specify a random vector which identifies the sequence of “make” choices for each firm in each sector of the economy, since these are determined residually from the K_{is} exogenous constants. The first line of (A.10) follows from Assumption 5a, and in particular on the ε_{ikt} random vectors being independent across buyers i and tasks k given t . The decomposition in the second line follows from the observation that all Ψ_{iktn} terms, for $n = d, f, m$ and given i and t , are identical across tasks within a sector (hence the replacement of the k subscript with the s one) as well as the fact that per Assumption 1a, $j^*(i, k, t)$ is a function (tasks can be fulfilled by at most one supplier). Similarly, the joint mass function of the in-degree sequences $\mathbf{b}_t^d$ and $\mathbf{b}_t^f$, for $n = d, f$ and $t = 1, \dots, T$, is:

$$\begin{aligned}
 (A.11) \quad & f_{b,b} \left(\mathbf{b}_t^d, \mathbf{b}_t^f \mid \boldsymbol{\kappa}_t^d, \boldsymbol{\kappa}_t^f, \mathbf{Z}_t^d, \mathbf{Z}_t^f, \mathbf{m}_t \right) = \\
 & = \prod_{i \in \mathcal{I}_t^d} \prod_{k \in \mathcal{K}_i} (\Psi_{iktm})^{\mathbb{1}[j^* = \emptyset]} (\Psi_{iktd})^{\mathbb{1}[j^* \in \mathcal{I}^d]} (\Psi_{iktf})^{\mathbb{1}[j^* \in \mathcal{I}^f]} \\
 & = \prod_{s=1}^S \prod_{i \in \mathcal{I}_t^d} (\Psi_{istm})^{K_{is} - B_{ist}^d - B_{ist}^f} (\Psi_{istd})^{B_{ist}^d} (\Psi_{istf})^{B_{ist}^f}.
 \end{aligned}$$

Hence, the joint mass function of the networks, conditional on the in-degree sequences, can be obtained as the ratio between (A.10) and (A.11), which can then be factorized as follows, as the two networks of type d and f are conditionally independent:

$$f_{G,G|b,b} \left(\mathbf{G}_t^d, \mathbf{G}_t^f \mid \mathbf{b}_t^d, \mathbf{b}_t^f, \boldsymbol{\kappa}_t^d, \boldsymbol{\kappa}_t^f, \mathbf{Z}_t^d, \mathbf{Z}_t^f, \mathbf{m}_t \right) = \frac{f_{G,G} \left(\mathbf{G}_t^d, \mathbf{G}_t^f \mid \cdot \right)}{f_{b,b} \left(\mathbf{b}_t^d, \mathbf{b}_t^f \mid \cdot \right)}$$

$$\frac{f_{G,G}(\mathbf{G}_t^d, \mathbf{G}_t^f | \cdot)}{f_{b,b}(\mathbf{b}_t^d, \mathbf{b}_t^f | \cdot)} = \prod_{n \in \{d,f\}} f_{G|b}(\mathbf{G}_t^n | \mathbf{b}_t^n, \boldsymbol{\kappa}_t^n, \mathbf{Z}_t^n),$$

where, for $n = d, f$ and $t = 1, \dots, T$:

$$(A.12) \quad f_{G|b}(\mathbf{G}_t^n | \mathbf{b}_t^n, \boldsymbol{\kappa}_t^n, \mathbf{Z}_t^n) = \prod_{i \in \mathcal{I}_t^n} \prod_{j \in \mathcal{I}_t^n} \left(\frac{\exp(\kappa_{jt} + \mathbf{z}'_{ijt} \boldsymbol{\varphi}^n)}{\sum_{\ell \in \mathcal{S}_{s(j)t}^n} \exp(\kappa_{\ell t} + \mathbf{z}'_{i\ell t} \boldsymbol{\varphi}^n)} \right)^{g_{ijt}}.$$

Note that the set of conditioned variables on the left-hand side of (A.12) only includes variables of type n . This is admissible as one can show that for $n = d$, the $\mathbf{b}_t^d$ vector is sufficient for variables of the f type, and vice versa.

The derivation then proceeds as in the proof of Proposition 1: one first obtains the mass function of the out-degree sequence $\mathbf{d}_t^n$ of type n , which is a sufficient statistic for the vector of rescaled fixed effects $\boldsymbol{\kappa}_t^n$, for $n = d, f$; hence, given a proper definition of $\mathcal{H}_{st}^n$, (15) is obtained from a derivation specular to (A.4); the details are omitted.

Proof of Proposition 4

We establish consistency and asymptotic normality for the RSL estimator as $S \rightarrow \infty$ or $T \rightarrow \infty$ in two ways. The first is to acknowledge that the model yields a categorical distribution representation of the probability to draw a specific subnetwork $\mathbf{H}_{st} \in \mathcal{H}_{st}^n$ for any $\mathcal{H}_{st}^n$ and for $n = d, f$, with the property that, given $\boldsymbol{\varphi}_o^n \equiv \boldsymbol{\beta}_o / \rho_n$:

$$\mathbb{E}[\mathbb{1}[\mathbf{G}_{st} = \mathbf{H}_{st}] | \mathbf{Z}_{st}^n, \boldsymbol{\beta}_o] = \Phi(\mathbf{H}_{st}^n | \mathcal{H}_{st}^n; \mathbf{Z}_{st}^n; \boldsymbol{\varphi}_o^n)$$

for $s = 1, \dots, S$ and $t = 1, \dots, T$. Note that since

$$|\Phi(\cdot | \cdot)| < \infty$$

and the categorical distribution belongs to the exponential family, the *unfeasible* subnetwork logit (USL) estimator based on $\mathcal{H}_{st}^n$ instead of $\mathcal{H}_{st}^{n*}$ and defined as:

$$\hat{\boldsymbol{\varphi}}_{USL}^n = \arg \max_{\boldsymbol{\varphi} \in \mathbb{R}^Q} \prod_{t=1}^T \prod_{s=1}^S \frac{\exp\left(\sum_{i \in \mathcal{I}_t^n} \sum_{j \in \mathcal{S}_{st}^n} g_{ijt} \mathbf{z}'_{ijt} \boldsymbol{\varphi}\right)}{\sum_{\mathbf{H}_{st} \in \mathcal{H}_{st}^n} \exp\left(\sum_{i \in \mathcal{I}_t^n} \sum_{j \in \mathcal{S}_{st}^n} h_{ijt} \mathbf{z}'_{ijt} \boldsymbol{\varphi}\right)},$$

satisfies the conditions of Theorem 5.4 and Corollary 5.5 by White (1994), which are based on Gourieroux, Monfort and Trognon (1984). Consequently,

$$\text{plim} \hat{\boldsymbol{\varphi}}_{USL}^n = \boldsymbol{\varphi}_o^n.$$

In addition, one can adapt the proof of Proposition 2 to the longitudinal dimension; this would show that

$$\text{plim} \hat{\boldsymbol{\varphi}}_{USL}^n = \text{plim} \hat{\boldsymbol{\varphi}}_{RSL}^n = \boldsymbol{\varphi}_o^n.$$

The second route leverages the standard theory of QMLE. Again by adapting the proof of Proposition 2, it is easy to show that:

$$\mathbb{E} [\nabla_{\boldsymbol{\varphi}} \log \Phi(\mathbf{G}_{st} | \mathcal{H}_{st}^{n*}; \mathbf{Z}_{st}^n; \boldsymbol{\varphi}_o^n)] = \mathbf{0}$$

for $s = 1, \dots, S$ and $t = 1, \dots, T$. Hence, by standard arguments the RSL estimator is consistent and asymptotically normal, with an asymptotic variance-covariance matrix which is a function of the dependence structure specific to the empirical setting.

This second approach is also useful to illustrate an alternative asymptotic regime for the RSL estimator. Let

$$N'_{st} = \sum_{i \in \mathcal{S}_{st}} \mathbb{1} [|\mathcal{K}_{it}^n| \geq 1]$$

be the total number of buyers of sector s active at time t and sourcing from type- n suppliers: even when S is fixed, the RSL estimator is consistent and asymptotically normal as $N'_{st} \rightarrow \infty$ for $s = 1, \dots, S$ and $t = 1, \dots, T$, and other regularity conditions hold. We elaborate next.

Write the gradient of the actual RSL log-likelihood function as:

$$(A.13) \quad \nabla_{\boldsymbol{\varphi}^n} \log \mathcal{L}^* \left(\boldsymbol{\varphi}^n \mid \{\mathbf{b}_t^n, \mathbf{d}_t^n, \mathbf{G}_t^n, \mathbf{Z}_t^n\}_{t=1}^T \right) = \sum_{t=1}^T \sum_{s=1}^S \sum_{i \in \mathcal{I}_t} \varsigma_{ist}(\mathbf{G}_{st}^n, \mathbf{Z}_{st}^n; \boldsymbol{\varphi}^n)$$

where $\varsigma_{ist}(\mathbf{G}_{st}^n, \mathbf{Z}_{st}^n; \boldsymbol{\varphi}^n) \equiv \sum_{j \in \mathcal{S}_{st}} (g_{ijt} - F_{ijt}(\mathbf{G}_{st}^n, \mathbf{Z}_{st}^n; \boldsymbol{\varphi}^n)) \mathbf{z}_{ijt}$, and:

$$F_{ijt}(\mathbf{G}_{st}^n, \mathbf{Z}_{st}^n; \boldsymbol{\varphi}^n) = \frac{\sum_{\mathbf{H}_{st} \in \mathcal{H}_{st}^{n*}} \exp \left(\sum_{i \in \mathcal{I}_t^d} \sum_{j \in \mathcal{S}_{st}^n} h_{ijt} \mathbf{z}'_{ijt} \boldsymbol{\varphi}^n \right) h_{ijt}}{\sum_{\mathbf{H}_{st} \in \mathcal{H}_{st}^{n*}} \exp \left(\sum_{i \in \mathcal{I}_t^d} \sum_{j \in \mathcal{S}_{st}^n} h_{ijt} \mathbf{z}'_{ijt} \boldsymbol{\varphi}^n \right)}.$$

The $\varsigma_{ist}(\mathbf{G}_{st}^n, \mathbf{Z}_{st}^n; \boldsymbol{\varphi}^n)$ terms express the deviation between the dyadic characteristics $\mathbf{z}_{ijt}$ of each buyer's suppliers in sector s at time t from the model's predictions under parameters $\boldsymbol{\varphi}^n$. Under the assumptions of the model, it is easy to verify that:

$$\mathbb{E} \left[\sum_{j \in \mathcal{S}_{st}} (g_{ijt} - F_{ijt}(\mathbf{G}_{st}^n, \mathbf{Z}_{st}^n; \boldsymbol{\varphi}^n)) \mathbf{z}_{ijt} \mid \mathcal{H}_{st}^{n*}, \mathbf{Z}_{st}^n \right] = \mathbf{0}$$

that is, the predictions are on average correct. Since $\mathcal{H}_{st}^{n*}$ is sampled uniformly

from $\mathcal{H}_{st}^n$, the Law of Iterated Expectations implies that $\mathbb{E}[\varsigma_{ist}(\mathbf{G}_{st}^n, \mathbf{Z}_{st}^n; \boldsymbol{\varphi}^n)] = 0$, which ensures consistency of the RSL estimator as all $N'_{st} \rightarrow \infty$ even if S and T are fixed.

We next establish sector-time specific Central Limit Theorem results of the kind:

$$(A.14) \quad \frac{1}{\sqrt{N'_{st}}} \sum_{i \in \mathcal{I}_t} \varsigma_{ist}(\mathbf{G}_{st}^n, \mathbf{Z}_{st}^n; \boldsymbol{\varphi}^n) \xrightarrow{d} \mathcal{N}(0, \mathcal{J}_{st})$$

for $s = 1, \dots, S$ and $t = 1, \dots, T$, where $\mathcal{J}_{st}$ is some information matrix specific to sector s and time t . Since this is conditional on the out-degree of sellers, the choices of different buyers are mutually dependent. In short, each seller can only be selected a finite and *fixed* number of times. By the Hájek Central Limit Theorem for sampling from finite populations without replacement, (A.14) holds if:

$$\max_{i \in \mathcal{I}_t} \|\varsigma_{ist}(\mathbf{G}_{st}^n, \mathbf{Z}_{st}^n; \boldsymbol{\varphi}^n)\| = o_p\left(\sqrt{N'_{st}}\right)$$

i.e. there are *no dominant buyers* that disproportionately drive the likelihood scores. Under this condition, (A.13) and, by extension, the RSL estimator are asymptotically normal. As discussed in the text, the true expression of the asymptotic variance may depend on sources of within-sector dependence (over time) or even cross-sector. The appropriate estimator of the RSL variance-covariance is thus context-specific.

B. Extended analysis of the model

This appendix elaborates on selected features of both the theoretical and econometric sides of the framework developed in Section I of the text, specifically: a. existence and uniqueness of equilibrium; b. distributional assumptions on the task-specific shocks; and c. extended assumptions that would enable structural dynamics.

Existence and uniqueness of equilibrium

This section rephrases results by Dhyne et al. (2023) in the context of our framework. Its objective is to formalize the concept of equilibrium in our setting, and to clarify under what assumptions equilibria exist and are unique. For simplicity, we restrict the discussion to the baseline framework from Section I in the text; absent structural dynamics, the results can be extended to the longitudinal model.

Demand side. We first introduce the consumers' preferences, which we neglected in the main text as they are not conducive to the development of the RSL estimator. As in Dhyne et al. (2023) and many other contributions, we assume consumers to have identical, homothetic CES preferences over the differentiated consumption goods sold by the firms of our economy. The utility U of the representative consumer is thus:

$$(B.1) \quad U = \left(\sum_{i \in \mathcal{I}} (\delta_i X_i^c)^{\frac{\psi-1}{\psi}} \right)^{\frac{\psi}{\psi-1}}$$

where X_i^c denotes the amount of the total output Y_i of any firm $i \in \mathcal{I}$ that ends up in the final consumer market (hence the superscript c), δ_i measures how salient firm i 's good is at determining consumer utility, while $\psi > 1$ is the elasticity of substitution. This leads to a standard expressions for consumers' final demand, for any $i \in \mathcal{I}$:

$$(B.2) \quad X_i^c = P_i^{-\psi} (\delta_i P)^{\psi-1} E$$

where P_i (with only one subscript) is the price charged to final consumers, E is the aggregate expenditure, whereas P (with no subscripts) is the domestic consumer price index that satisfies:

$$(B.3) \quad P^{1-\psi} = \sum_{i \in \mathcal{I}} \left(\frac{\delta_i}{P_i} \right)^{\psi-1}.$$

Furthermore, the representative consumer supplies one unit of labor inelastically.

Structure of firm sales. Given (1) and (2), the share of costs that any firm $i \in \mathcal{I}$ spends on a particular task $k \in \mathcal{K}_i$, given $\mathcal{J}_i$, is given by:

$$(B.4) \quad \frac{P_{ijk}X_{ijk}}{C_i Y_i} = \frac{(\alpha_k/P_{ijk})^{\sigma-1}}{\Theta_i}$$

where $\Theta_i = (C_i A_i)^{1-\sigma} = (\alpha_{0i}/W)^{\sigma-1} + \sum_{k \in \mathcal{K}_i} (\alpha_k/P_{ijk})^{\sigma-1}$ is, borrowing terminology by Antràs, Fort and Tintelnot (2017), the *sourcing capability* of firm i . Similarly:

$$(B.5) \quad \frac{WL_i}{C_i Y_i} = \frac{(\alpha_{0i}/W)^{\sigma-1}}{\Theta_i}$$

is the wage bill over a firm's cost. The *total sales* $\mathcal{Y}_j$ of a firm $j \in \mathcal{I}$ are given by the sum of sales to final consumers and sales to customer firms:

$$(B.6) \quad \begin{aligned} \mathcal{Y}_j &= P_j X_j^c + \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}_{ij}} P_{ijk} X_{ijk} \\ &= \left(\frac{\delta_j A_j P}{\mu_j^c} \right)^{\psi-1} \Theta_j^{\frac{\psi-1}{\sigma-1}} E + \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}_{ij}} \left(\frac{\alpha_{ijk} A_j}{\mu_{ijk}} \right)^{\sigma-1} \frac{\Theta_j}{\Theta_i} \frac{\mathcal{Y}_i}{\bar{\mu}_i} \end{aligned}$$

where $\mathcal{K}_{ij} \equiv \{k \in \mathcal{K}_i : j(k) = j\}$ is the subset of all tasks $\mathcal{K}_i$ of a firm $i \in \mathcal{I}$ that are supplied by j , given $\mathcal{J}_i$; $\alpha_{ijk} \equiv \alpha_k \tau_{ij}^{-1} \exp(\varepsilon_{ijk})$; μ_j^c is the markup charged by firm j on final consumers, while

$$(B.7) \quad \bar{\mu}_j \equiv \frac{1}{Y_j} \left(\mu_j^c X_j^c + \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}_{ij}} \mu_{ijk} X_{ijk} \right)$$

is a firm's "average" (quantity-weighted) markup. Notice that the dyadic frictions τ_{ij} appear in (B.6) through the α_{ijk} terms: despite their "iceberg" nature, they contribute to the monetary transfers from buyers to sellers. They also appear implicitly through the sourcing capabilities Θ_i , per (3), and through the markups μ_{ijk} , as shown next.

Determination of markups. In our model the transaction-task-specific markups μ_{ijk} are governed by Assumption 4 on limit pricing. Hence, writing j^* as shorthand for $j^*(i, k)$ and j^{**} as shorthand for $j^{**}(i, k)$, i.e. the *second best* eligible supplier for a given task k of any firm i :

$$(B.8) \quad j^{**}(i, k) = \arg \max_{j \in \{\mathcal{S}_{z(k)} \setminus j^*(i, k)\}} \gamma_j - \log \tau_{ij} + \varepsilon_{ijk},$$

we obtain:

$$(B.9) \quad \mu_{ij^*k} = \frac{C_{j^{**}}}{C_{j^*}} \frac{\tau_{ij^{**}}}{\tau_{ij^*}} \exp(\varepsilon_{ij^*k} - \varepsilon_{ij^{**}k}).$$

Final consumers are charged a monopolistically competitive markup:

$$(B.10) \quad \mu_i^c = \frac{\psi}{\psi - 1},$$

for all $i \in \mathcal{I}$. This is obtained as profit-maximizing firms treat $P^{\psi-1}E$ as given.

Aggregation. Firm profits are distributed to consumer-workers:

$$(B.11) \quad E = WL + \sum_{i \in \mathcal{I}} \left(1 - \frac{1}{\bar{\mu}_i}\right) \gamma_i$$

where L is the total size of the economy. Moreover, for the labor market to clear, total labor income must equal total labor costs:

$$(B.12) \quad WL = \sum_{i \in \mathcal{I}} \frac{\gamma_i (\alpha_{0i}/W)^{\sigma-1}}{\bar{\mu}_i \Theta_i}.$$

Equilibrium definitions. To expedite exposition, unlike Dhyne et al. (2023) here we define only one type of equilibrium in which the network is endogenous, and matching is initiated by buyers.

Definition 1. *In this economy, an equilibrium is the collection of: a. a network*

$$\mathcal{G}^* \equiv \bigcup_{i \in \mathcal{I}} \bigcup_{k \in \mathcal{K}_i} (i, j^*(i, k));$$

b. a sequence of firm marginal costs $\{C_i\}_{i \in \mathcal{I}}$; c. a sequence of firm labor demands $\{L_i\}_{i \in \mathcal{I}}$; d. a sequence of price-quantity combinations in the final consumer markets, $\{(P_i, X_i^c)\}_{i \in \mathcal{I}}$; e. a set of price-quantity combinations for each sourced input specified in the above definition of $\mathcal{G}^*$:

$$\left\{ \left\{ (P_{ij^*(i,k)k}, X_{ij^*(i,k)k}) \right\}_{k \in \mathcal{K}_i} \right\}_{i \in \mathcal{I}};$$

f. a consumer price index P ; and g. an aggregate expenditure E ; such that equations (2), (5) (B.2), (B.3), (B.4), (B.5), (B.6), (B.9), (B.10), (B.11), (B.12) are satisfied for some L , numeraire W , sequence $(A_1, \dots, A_N)$, matrix of friction measures τ_{ij} and set of task-specific shocks ε_{ijk} under the assumptions and parameters of the model.

Our analysis of equilibrium specializes to a limit case defined as follows.

Definition 2. *In a limit economy all sectors grow indefinitely in size: $|\mathcal{S}_s| \rightarrow \infty$ for $s = 1, \dots, S$.*

We consider a variation of the “acyclic potential (production) network” by Dhyne et al. (2023) as adapted to our setting.

Definition 3. *An acyclic potential network is one where the N firms can be ordered in such a way that buyers can only source their inputs from sellers that precede it in the sequence. Thus, without any loss of generality, if $i < j$ for a given pair $(i, j) \in \mathcal{I}^2$, the following restriction must hold: $s(j) \notin z(\mathcal{K}_i)$. This implies that the first firm in the sequence only employs labor: $\mathcal{K}_1 = \emptyset$.*

This definition is more general than the statement of Assumption 3 (which posits acyclicity *at the sectoral level*): for example, it allows buyers earlier in the sequence to buy from suppliers in their own sector that are listed later in the sorting order, but not vice versa. Our econometric results (Proposition 1 through 4) are still valid under this weaker assumption. We illustrate our framework under the stronger Assumption 3 for expositional reasons: in fact, it ties more intuitively with the empirical evidence from Fact 2 and the sectoral-level dimension of the RSL estimator.

Existence and uniqueness. We adapt the analysis by Dhyne et al. (2023) noting first that (2) can be recast as a system of N linear equations:

$$(B.13) \quad \mathbf{c} = \mathbf{A}(\mathbf{1} + \mathbf{\Gamma}\mathbf{c}),$$

where

$$\begin{aligned} \mathbf{c} &= (C_1^{1-\sigma}, \dots, C_N^{1-\sigma}), \\ \mathbf{A} &= \text{diag}(A_1^{\sigma-1}, \dots, A_N^{\sigma-1}), \\ \mathbf{1} &= W^{1-\sigma}(\alpha_{01}^{\sigma-1}, \dots, \alpha_{0N}^{\sigma-1}), \end{aligned}$$

while $\mathbf{\Gamma}$ is an $N \times N$ matrix whose Γ_{ij} elements, for $i, j = 1, \dots, N$, read as:

$$(B.14) \quad \Gamma_{ij} = \sum_{k \in \mathcal{K}_i} \mathbb{1}[j(k) = j] \left(\frac{\alpha_{ijk}}{\mu_{ijk}} \right)^{\sigma-1}$$

for some given $\mathcal{G}$. In what follows we write $\mathbf{\Gamma}^*$ as the version of $\mathbf{\Gamma}$ such that in (B.14), $\mu_{ijk} = 1$ for any $i \in \mathcal{I}$, $k \in \mathcal{K}_i$, and $j \in \mathcal{S}_{z(k)}$.

We next formulate two lemmata; the first one adapts a special case of the “fixed network” analysis by Dhyne et al. (2023).

Lemma 1. *If the production network $\mathcal{G}$ is given, the spectral radius of $\mathbf{A}\Gamma^*$ is less than one, and $\mu_{ij(k)k} = 1$ for any $k \in \mathcal{K}_i$ and $i \in \mathcal{I}$, it is possible to solve uniquely for $\mathbf{c}$, P , E and all input demands, prices and quantities as specified in the definition of equilibrium so that they meet the respective definitions or conditions.*

Proof. If the spectral radius of $\mathbf{A}\Gamma^*$ is less than one it is $\mathbf{c} = (\mathbf{I} - \mathbf{A}\Gamma^*)^{-1} \mathbf{A}\mathbf{l}$, where $\mathbf{I}$ is the identity matrix of dimension N . As all the markups are given, this determines all prices in the economy as well as the price index P . Because no profits are obtained from selling inputs to firms, combining the aggregation equation (B.11) together with (B.6) allows (with conventional algebraic manipulations) to recover the standard CES result that expenditures E are a constant multiple of labor income WL :

$$E = \frac{\psi}{\psi - 1} WL.$$

This allows to obtain the quantities sold to final consumers through (B.2). Lastly, all input demands are backed up via (B.9) and (B.10). $\square$

The second lemma is specific to our own setting: intuitively, it states that as the set of potential suppliers becomes increasingly larger, the markets for intermediate inputs approach perfect competition, in the sense that prices equal marginal costs.

Lemma 2. *If all firms choose their suppliers according to (5), while both A_i and τ_{ij} have bounded support, then in a limit economy equilibrium asymptotically it holds that $\mu_{ij^*(i,k)k} \rightarrow 1$ for all $i \in \mathcal{I}$ and $k \in \mathcal{K}_i$.*

Proof. Given (B.8) and (B.9), this lemma is equivalent to the following statement on convergence in probability, for any $\delta > 0$ and for all tasks and buyers in question:

$$\lim_{|\mathcal{S}_{z(k)}| \rightarrow \infty} \mathbb{P}(|(\varepsilon_{ij^*k} - \varepsilon_{ij^{**}k}) + (\theta_{ij^*} - \theta_{ij^{**}})| > \delta) = 0,$$

where $\theta_{ij} \equiv \gamma_j - \log \tau_{ij}$ for any $(i, j) \in \mathcal{I}^2$. This probability is the same as:

$$\mathbb{P} \left(\bigcap_{\substack{\ell \in \mathcal{S}_{z(k)} \\ \ell \neq j^*}} \{(\varepsilon_{ij^*k} - \varepsilon_{i\ell k}) > -(\theta_{ij^*} - \theta_{i\ell}) + \delta\} \right) = \frac{\exp(\theta_{ij^*} - \delta)}{\sum_{\ell \in \mathcal{S}_{z(k)}} \exp(\theta_{i\ell})},$$

which goes to zero at the limit as the number of addends in the denominator increase, so long as both constituent elements of θ_{ij} are drawn from a distribution with bounded support. In the case of τ_{ij} , this holds by assumption; as for γ_j , this follows from the analogous assumption on A_i and from the limit economy being

in equilibrium, which implies that γ_j is a deterministic function of the infinite sequence $(A_1, A_2, \dots)$. $\square$

While the “speed of convergence” here may depend on the specific distribution of both A_i and τ_{ij} , the intuition stands unaffected.

We can now state the main result by recombining the above lemmata. Specifically, similarly to Dhyne et al. (2023) we provide conditions for the existence and uniqueness of the equilibrium, along with an algorithm to derive that is implicit in the proof.

Theorem 1. *Under the hypotheses of Lemma 2, a limit economy with an acyclic potential production network whose all eligible $\mathbf{A}\mathbf{\Gamma}$ matrices have spectral radius less than one has a unique equilibrium.*

Proof. The proof is constructive and inductive. Start from firm $i = 1$ in the sequence implicit in Definition 3: because it only decides about how much labor to hire and what price and quantity to set in the final consumers market, its choices are uniquely determined. Firm $i = 2$ can potentially buy from firm 1 (which may or may not share their sector). Since the marginal costs of firm 1 are determined in the previous step, the choices of firm 2 are also uniquely determined. Firm $i = 3$ may source from both 1 and 2, and so forth. Iterate for $i = 1, \dots, N$ to obtain a candidate equilibrium *limit* network $\mathcal{G}^*$. By hypothesis, the spectral radius of the $\mathbf{A}\mathbf{\Gamma}^*$ matrix associated with $\mathcal{G}^*$ is less than one; hence, the conditions of Lemma 1 are satisfied and the economy’s equilibrium is fully determined *at the limit* where markups are all unitary. This is in turn ensured by Lemma 2. Consequently, the unique equilibrium of this particular limit economy is identified. $\square$

A limit economy is interesting from a theoretical viewpoint because one can easily prove that it is *constrained efficient*: if a planner cannot alter any of: (i) the matrix of τ_{ij} weights, (ii) the task-supplier-specific ε_{ijk} shocks, (iii) the final consumer markups μ_i^c , then, at the limit where the input markups are zero the market allocation coincides with the social optimum (policymakers might under some circumstances manipulate some of those elements: for example, the τ_{ij} “wedges”). Note that it is not necessary that the economy approaches the zero-markups limit for our quantitative framework to hold, as the RSL estimator accommodates nonzero firm-to-firm markups.

On the distribution of the errors

This section expands the discussion about Assumptions 5 and 5a of our framework, which specify distributional assumptions for the ε_{ijk} shocks of our model.

On other distributions. As is well known, the multinomial logit probabilities that are foundational to our model can be obtained otherwise. For example, if

instead of (3) prices were assumed as:

$$(B.15) \quad P_{ijk} = \frac{\mu_{ijk} C_j \tau_{ij}}{v_{ijk}},$$

where v_{ijk} is independent across buyers, suppliers and tasks and follows the Fréchet (type II generalized extreme value) distribution with location parameter equal to zero, arbitrary scale parameter, and unitary shape parameter, (6) would still hold exactly and our cross-sectional framework would stay unchanged; an analogous modification applies to the longitudinal case too. One can also derive (6) as an *approximation* under the assumption that v_{ijk} follows a type I Pareto distribution where both parameters are equal to one, see e.g. Panigrahi (2023). The statistical literature emphasizes more general assumptions leading to multinomial logit probabilities as approximations.

On scale parameters. The model allows unrestricted scale parameters for ε_{ijk} (or in the Fréchet case, shape parameters for v_{ijk}); setting them as one is a convenient normalization. Note, however, that the interpretation of the estimates is affected by the normalization. Suppose that the “true” ε_{ijk} has scale parameter equal to $\pi^{-1} > 0$ (or v_{ijk} has shape parameter equal to π); then (6) is replaced by:

$$(B.16) \quad \mathbb{P}\left(j^*(i, k) = j \mid \{(\gamma_\ell, \mathbf{z}_{i\ell})\}_{\ell \in \mathcal{S}_{z(k)}}\right) = \frac{\exp\left(\pi\left(\gamma_j + \mathbf{z}'_{ij}\boldsymbol{\beta}\right)\right)}{\sum_{\ell \in \mathcal{S}_{z(k)}} \exp\left(\pi\left(\gamma_\ell + \mathbf{z}'_{i\ell}\boldsymbol{\beta}\right)\right)}.$$

While the combined parameter set $\pi\boldsymbol{\beta}$ is still identified, the construction of the RSL estimator would proceed unchanged, and any interpretation of the estimates in terms of odd-ratios is still valid; those interpretations based on imputed or simulated values of γ_j must acknowledge that they are affected by the value of π in (B.16). Therefore, we suggest that researchers pursuing this route discuss how the calculated marginal effects change as a function of the unknown scale or shape parameter.

On microfoundations. The structural motivation for Assumption 5a summarized in Section II is as follows. Suppose that in every time period t , which for simplicity we equate to the time dimension of the data (though this is not necessary), all dyads, including those involving foreign sellers, receive a shock ω_{ijkt} for each task k of the buyer. This shock represents the value of a specific *production technique* that can be used to fulfill task k . If one such a technique is used, the effective price for the buyer is

$$P_{ijkt} \propto \frac{\mu_{ijkt} C_{jt} \tau_{ijt}}{\exp(\omega_{ijkt})}.$$

We assume that in each time period, these shocks are independent across buyers and tasks. However, it is:

$$(B.17) \quad \mathbb{Cov}(\omega_{ijkt}, \omega_{i\ell kt}) = \begin{cases} \varrho_d & \text{if } (j, \ell) \in (\mathcal{I}^d)^2, \\ \varrho_f & \text{if } (j, \ell) \in (\mathcal{I}^f)^2, \\ 0 & \text{otherwise;} \end{cases}$$

that is they present equal correlation across sellers of the same type, where $\varrho_n > 0$ for $n = d, f$. Techniques are never forgotten, but may be replaced as better alternatives become available. Thus, assuming that firms have started learning the techniques T_0 time periods before present:

$$(B.18) \quad \varepsilon_{ijkt} = \frac{\max \left\{ \omega_{ijkt}, \omega_{ijk(t-1)}, \dots, \omega_{ijk(t-T_0)} \right\} - b_{T_0}}{a_{T_0}},$$

where $a_{T_0} > 0$ and $b_{T_0} \in \mathbb{R}$ are two “standardizing” sequences. Multivariate extensions of the Fisher–Tippett–Gnedenko theorem show that under these hypotheses, $T_0 \rightarrow \infty$ implies convergence in distribution of ε_{ijkt} to the distribution specified by Assumption 5a for a vast set of underlying assumptions about the original distributions of the ω_{ijkt} shocks. Here, a_{T_0} and b_{T_0} are interpreted similarly as the location and scale parameters of ε_{ijk} discussed previously. A similar analysis applies to the “make” shocks ε_{0ikt} too. It is important to note that this particular microfoundation leads to ε_{ijkt} shocks that are strongly persistent in time for any given (i, j, k) triplet.

On extensions. More general assumptions on the random component of transaction values would lead naturally to extensions of our framework. Suppose for example that the covariance of the production techniques is constant for techniques coming from the same seller sector, and is zero otherwise. Then, Assumption 5a might be modified so as to accommodate the following joint distribution of the error term:

$$(B.19) \quad F_\varepsilon(\varepsilon_{ikt}) = \exp \left[-\exp(-\varepsilon_{0ikt}) - \sum_{n \in \{d, f\}} \left(\sum_{j \in \mathcal{S}_{z(k)t}^n} \exp \left(-\frac{\varepsilon_{ijkt}}{\rho_{nz(k)t}} \right) \right)^{\rho_{nz(k)t}} \right]$$

where the parameters $\rho_{nst} \in (0, 1]$ vary over $n = d, f$, $s = 1, \dots, S$ and $t = 1, \dots, T$. This is conducive to a random parameter treatment of our estimation framework as discussed in section II of the text.

Allowing for structural dynamics

As we discussed in Section I, allowing for structural dynamics would make our model misspecified by construction. To appreciate why, suppose that:

$$(B.20) \quad -\log \tau_{ijt} + \beta_0 = \mathbf{z}'_{ijt} \boldsymbol{\beta} + \phi \mathbb{1} [g_{ij(t-1)} > 0],$$

where parameter ϕ measures the extent to which past transactions reduce frictions (interpretations, e.g. in terms of set-up costs of a relationship, are easy to formulate). Because firms maximize their stream of future profits, $\phi > 0$ implies that a buyer i may source a task k from a supplier $j^\dagger \neq j^* = j^*(i, k, t)$ when $g_{ij^*(t-1)} > 0$, even if:

$$\gamma_{j^*t} + \mathbf{z}'_{ij^*t} \boldsymbol{\beta} + \phi + \varepsilon_{ij^*t} > \gamma_{j^\dagger t} + \mathbf{z}'_{ij^\dagger t} \boldsymbol{\beta} + \varepsilon_{ij^\dagger t} > \gamma_{j^*t} + \mathbf{z}'_{ij^*t} \boldsymbol{\beta} + \varepsilon_{ij^*t}$$

because in period $t + 1$, supplier $j^\dagger$ is expected to gain ϕ units of log-value. One may specify conditions about the expectation of future realizations of the state variables such that for firm i , switching to $j^\dagger$ at time t would lead to a higher present value of profits. A full-fledged characterization of the choice problem as dynamic and forward-looking would considerably increase the complexity of the model.

Nevertheless, it is possible to formulate assumptions and restrictions that allow the identification of structural dynamics as per (B.20) within the framework developed in the paper. Write the count of suppliers that have been fulfilling any task of a firm i for at least two years up to time t as:

$$\Delta_{it} = \sum_{j \in \mathcal{I}_t^d} \sum_{k \in \mathcal{K}_i} \mathbb{1} [j = j^*(i, k, t) \wedge j = j^*(i, k, t-1)]$$

and consider the following replacement for Assumption 6.

Assumption 1a. *The set of dyadic characteristics $\{\mathbf{z}_{ijt}\}_{j \in \mathcal{I}}$ are not Granger-caused by the sequence of past networks $\mathcal{G}_{(t-u)}^*$, $u \in \mathbb{N}_0$. However, the total factor productivity of every firm i evolves according to a stochastic process that satisfies:*

$$\mathbb{E} \left[A_{i(t+1)}^{\sigma-1} \middle| A_{it}, \Delta_{it}, \Delta_{i(t-1)} \right] \geq A_{it}^{\sigma-1} \exp \left(\tilde{\phi} \left(\Delta_{it} - \Delta_{i(t-1)} \right) \right)$$

where $\tilde{\phi} \geq \phi(\sigma - 1)$, for all past configurations of the network $\mathcal{G}_{t-u}$, $u \in \mathbb{N}$.

This assumption states that whenever a task k is sourced to a supplier j for at least two consecutive periods, j 's total factor productivity is then expected to increase by a percentage amount at least as high as ϕ . Holding (B.20), this offsets any loss ϕ due to “anticipated switching” as per the previous discussion. It follows that for any task $k \in \mathcal{K}_i$ of some buyer $i \in \mathcal{I}$, the optimal supplier at time t defined as

(14) is also *on expectation* the optimal one at every subsequent period $t+u$, $u \in \mathbb{N}$. Hence, a version of our model that allows for structural dynamics as per (B.20) is well-specified, since the “short-term” optimal choices at time t are consistent with forward-looking optimal plans based on the present value of all future profits. To formalize this intuition, it is useful to define $\mathcal{G}_{-(ik)t} = \mathcal{G}_t \setminus (i, j(k))$ as the network at time t minus the connection between firm i and the supplier (if there is any) that fulfills task $k \in \mathcal{K}_i$. In addition, let $j^*(i, k, t+u | \mathcal{G}_{-(ik)t+u})$ denote the *expected* optimal supplier for firm i ’s task k in any *future* time period, for $u \in \mathbb{N}$, conditional on $\mathcal{G}_{-(ik)t+u}$.

Theorem 2. *Let $\mathcal{G}_{-(ik)t+u} = \mathcal{G}_{-(ik)}$ be constant for all $u \in \mathbb{N}_0$. Under assumptions 1a, 2, 3, 4, 5a, 1a, and structural dynamics as specified in (B.20), for all $u \in \mathbb{N}$ it is:*

$$j^*(i, k, t) = j^*(i, k, t+u | \mathcal{G}_{-(ik)})$$

when $j^*(i, k, t) = j^*(i, k, t-1)$ for all $i \in \mathcal{I}^d$ and $k \in \mathcal{K}_i$: that is, suppliers that are “optimal” in the sense of (14) for a given buyer’s task for two consecutive periods are *on expectation* the ex ante optimal choice in every other future period, regardless of the buyer’s previous choices for that particular task.

Proof. Consider $u = 1$, and suppose that

$$j^* = j^*(i, k, t) \neq j^*(i, k, t+1 | \mathcal{G}_{-(ik)}) = j^\dagger.$$

As both z_{ijt} and ε_{ijt} are (conditionally) independent of firm i ’s choices, this implies that $j^\dagger$ would only be preferable to j^* at time $t+1$ if it were also chosen for the same task k at time t (anticipated switching), which occurs if:

$$\begin{aligned} \mathbb{E} \left[C_{j^\dagger(t+1)}^{1-\sigma} \middle| (i, j^\dagger(i, k, t)) \in \mathcal{G}_t^* \right] &= C_{j^\dagger t}^{1-\sigma} \exp(\Phi(1-\sigma)) \\ &\geq \mathbb{E} \left[C_{j^*(t+1)}^{1-\sigma} \middle| (i, j^*(i, k, t)) \in \mathcal{G}_t^* \right] \\ &\geq (C_{j^* t} A_{j^* t})^{1-\sigma} \cdot \mathbb{E} \left[A_{j^*(t+1)}^{\sigma-1} \middle| A_{j^* t}, (i, j^*(i, k, t)) \in \mathcal{G}_t^* \right] \\ &\geq C_{j^* t}^{1-\sigma} \exp(\tilde{\Phi}). \end{aligned}$$

The second inequality (third line) follows from (12) and from $\mathcal{G}_{-(ik)}$ being fixed, and the third inequality (fourth line), from Assumption 1a. This implies

$$j^\dagger = j^*(i, k, t),$$

which is a contradiction. The result for $u = 2$ obtains analogously, and by induction, it also extends to all other higher values of u . $\square$

While one could consider extensions for more elaborate versions of structural time-dependence on past realizations of the network, two considerations are in order. First, Assumption 1a has a meaningful economic interpretation: supplying inputs to other firms leads to increases, however small, of the supplier's productivity. Whereas this is reasonable in many settings (e.g. with learning-by-doing) this assumption must be defended and ideally, tested. Second, defending the assumption is increasingly more difficult the larger one expects ϕ to be. Our empirical application shows that the RSL estimates associated with the "past transactions dummy" are small in magnitude and not statistically significant: at least in this particular setting, this diminishes concerns about defensibility of the proposed approach.

C. Data and supporting stylized facts

To illustrate and support our quantitative framework, we use firm-to-firm transaction data from Costa Rica, which have already appeared in other contributions (among the others, Alfaro-Ureña, Manelici and Vasquez, 2022; Alfaro-Ureña et al., 2023). These data are actively maintained at the Central Bank of Costa Rica (BCCR) as part of the *Registro de Variables Económicas del BCCR* (“Revec”), and are elaborated from value added tax records. They collect all firm pairs that have been observed to transact with one another since 2006, and the total amount of the unilateral yearly transactions (i.e. the total transfers from a buyer to a seller), provided that in each year, such amount is higher than a threshold set at about \$4,800. Below that, firms are not obliged to report the identity of their transaction partners to tax authorities. Like similar firm-to-firm transaction datasets, it is impossible to appreciate any finer granularity of total transactions (e.g. if a buyer makes two separate orders from the same seller in the same year, one for \$6,000 and one for \$4,000, we only observe the total \$10,000). Thanks to unique firm identifiers, we can link balance sheet as well as additional firm information (such as location and four-digits sector code) to both sides of each transaction.

Table C.1—: Costa Rican production network: summary of key distributions

	Min	25p	50p	75p	Max
Buyers (total number: 68,385)					
Yearly no. of suppliers (“in-degree”)	1	1	3	7	402
Total no. of sourced sectors (2-digits)	1	2	3	6	67
Total no. of sourced sectors (4-digits)	1	2	4	8	182
Sellers (total number: 49,279)					
Yearly no. of buyers (“out-degree”)	1	1	2	6	3,861
Total no. of served sectors (2-digits)	1	1	2	6	73
Total no. of served sectors (4-digits)	1	1	3	7	259

Note: This table reports key percentiles (minimum, first quartile, median, third quartile, maximum) of some specific topological properties of the Costa Rican firm production network, for buyers and suppliers separately. A “yearly no. (number)” is conditional on a firm being observed to transact (as buyer or seller) in a given year. A “total no. (number)” is calculated over the entire 2008-2017 interval.

Source: *Revec*.

We use 2,283,102 total yearly transactions that occurred between 2008 and 2017. This number results from matching all firm-to-firm transactions to the firm balance sheets, and removing transactions whose total amount exceeds the revenue of buyers (these are arguably major investment projects rather than purchases of intermediate inputs). From the universe of 95,477 Costa Rican firms tracked in our panel of balance sheets, 68,385 are observed at least once as the

buyer of a transaction, whereas 49,279 are at least one-time sellers. Table C.1 summarizes key facts about the Costa Rican transaction data. In particular, the distributions of both the number of sellers (“in-degree”) and number of buyers (“out-degree”) that a firm has in a year are, as expected, highly skewed; the same applies to the distribution of the number of sectors that firms interact with over time, as either buyers or sellers. Notably, most firms interact with partners from a handful of sectors only, regardless of the sector classification one is adopting (two- versus four-digits). A full-fledged description of the dataset is outside the scope of this paper; interested readers can consult the cited contributions and the specialized paper by Alfaro-Ureña et al. (2018), which is dedicated to a comprehensive descriptive analysis of the Costa Rican production network from 2008 to 2015.

The rest of this appendix documents five stylized facts, each supported by a figure, that motivate some features of the quantitative framework and analysis of this paper.

Fact 1. *Buyer (seller) firms often transact with two or more seller (buyer) firms from the same industry in the same year: more than what random chance would predict.*

Figure C.1 shows that typically firms buy from, or sell to, only one partner

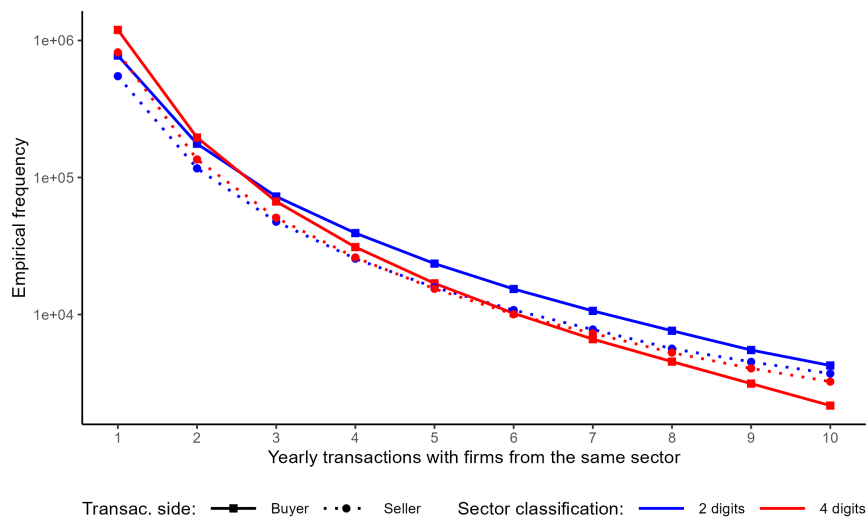


Figure C.1. : Frequency of transactions with multiple partners from the same sector

Note: This figure reports the empirical frequency of all those episodes, counted over the time interval spanned by the data (2008-2017), where in a given year a firm transacts with x “partners” belonging to the same sector, separately for the two sides of a transaction and for two classifications of sectors (two- versus four-digits). If the transaction side is the buyer’s the partners are sellers, and vice versa. The horizontal axis runs over x and is truncated at 10. The vertical axis is cast on a logarithmic scale.

Source: Revec.

from each industry (however the latter is classified). At the same time, firms usually report multiple partners from the same sector. Although the frequency of such occurrences is one order of magnitude smaller than instances of “only one partner per sector,” it is substantial and unlikely induced by chance. To elaborate, suppose that firms from the same industry source their inputs from non-overlapping subsets of sectors, as in the stylized, constructed example from Table C.2: this involves four buyer firms as well as four seller sectors that the former may source from. There, the pseudo-buyers display some variation in the allocation of their inputs across sectors, but a Fisher’s exact test of independence (which is more suited to contingency tables with low count values than a Pearson’s chi-squared test) cannot reject the null hypothesis that buyer “choices” are independent. In principle, Figure C.1 may be rationalized by a “balls and bins” model where buyers choose their suppliers from random sectors.

Table C.2—: Random sourcing from sectors: a constructed example

	Sector A	Sector B	Sector C	Sector D	Total
Buyer 1	1	2	1	1	5
Buyer 2	0	2	0	1	3
Buyer 3	1	1	2	0	4
Buyer 4	1	0	0	2	3
Total	3	5	3	4	15

Note: This contingency table outlines a stylized, constructed example about a supplier choice pattern consistent with the hypothesis that transactions are independent of the sellers’ sectors. A non-marginal cell reports the number of distinct transactions that any buyer firm (rows) makes with seller firms from a given sector (columns). A Fisher’s exact test of independence for this example yields a p -value of 0.679.

We test the “random sectors” hypothesis statistically. Specifically, in each selection of the data restricted to all buyer firms from a specific four-digits sector (pooling all years) we construct a contingency table akin to C.2, though real tables are typically larger. For each four-digits sector, we perform a Fisher exact test of independence like the one in the example of Table C.2, *after removing columns* relative to sectors that are *never observed* to have any buyer (which would skew the results against the null). Figure C.2 displays the p -values from all tests, arrayed in a histogram. The evidence overwhelmingly rejects the hypothesis that across industries, firms choose suppliers from random sectors, even if restricted to some industry-specific set. Instead, within each industry buyers purchase from very idiosyncratic combinations of sectors. This evidence supports modeling assumptions that restrict task allocation across sectors.

An alternative explanation to the high concentration of transactions as evidenced from Figure C.1 is that the sectors of suppliers are not correctly measured in the data. One reason for this is that multiproduct firms often operate in multiple industries, but are typically only associated to one industry code in administrative data (we thank an anonymous referee for this observation). To

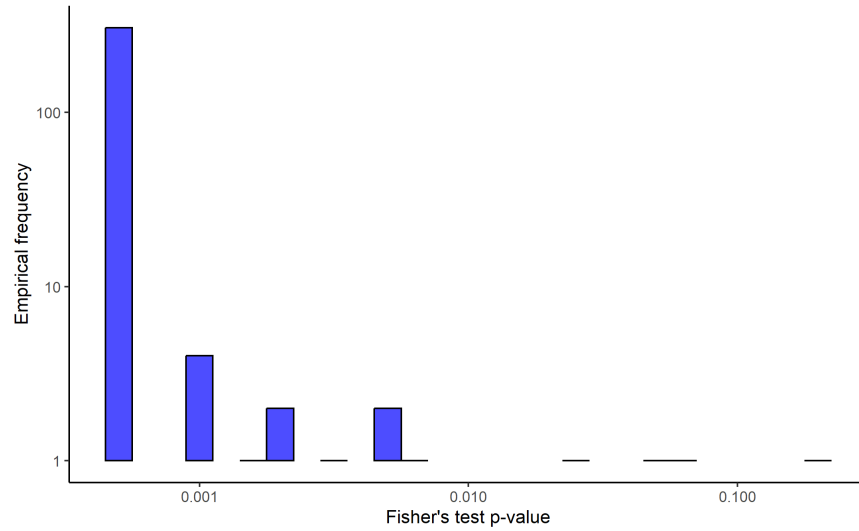


Figure C.2 : Distribution of Fisher test p -values across four-digits sectors

Note: This histogram displays the empirical distribution of the p -values from the Fisher exact tests, one per four-digit sector, as described in this appendix. Both the horizontal and the vertical axes are cast on a logarithmic scale.

Source: Revec.

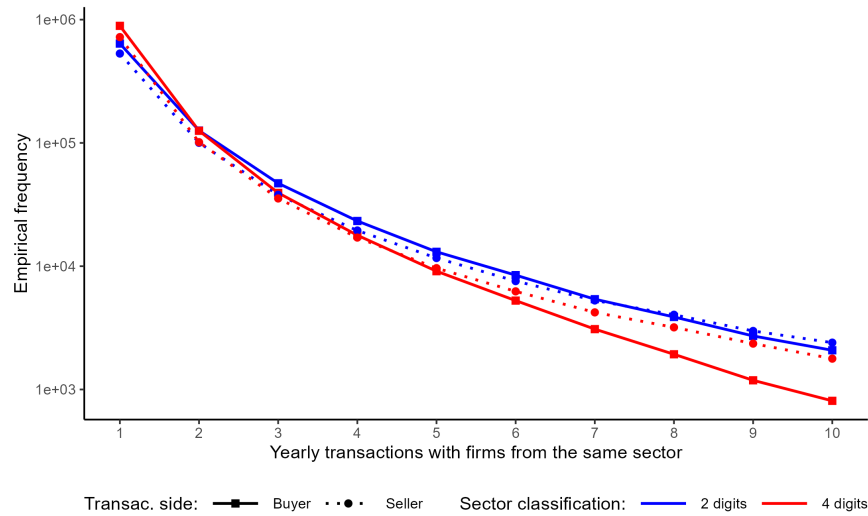


Figure C.3 : Frequency of transactions with small-medium partners from the same sector

Note: This figure reports the same information as Table C.1, but calculated by excluding transactions whose sellers have ever employed more than 50 workers between 2008 and 2017.

Source: Revec.

execute a first-order assessment of this possibility, we replicate Figure C.1 using a restricted sample that excludes suppliers ever observed with more than fifty employees: firms most likely to be multiproduct. The resulting pattern, reported in Figure C.3, is virtually identical to the baseline, suggesting that the observed concentration is not primarily driven by sectoral overlap but instead reflects features of the underlying production network formation process. A restricted version of figure C.2 similarly yields qualitatively identical implications.

Fact 2. *The input-output matrix is approximately acyclical.*

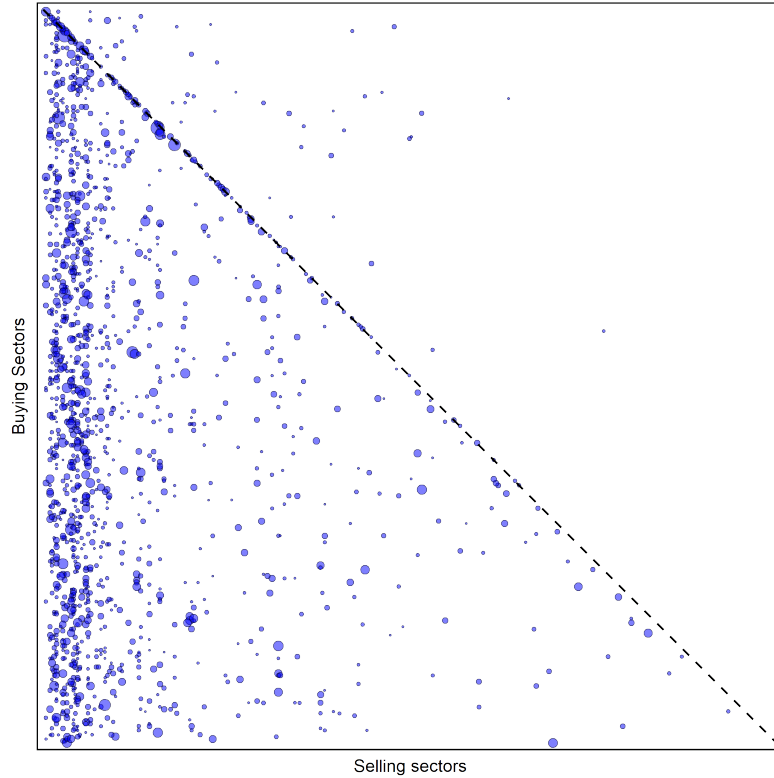


Figure C.4. : Pooled input-output matrix: relevant inter-sectoral connections

Note: The balls in this figure represent the intensity of input-output connections between four-digits sectors of the Costa Rican economy: each combination of sectors is a point in the figure. The size of each ball is proportional to share of purchases by all firms in a given buying sector that are made from all firms of another selling sector (pooling all years in the data). Directed sector pairs with calculated shares smaller than 0.05 are not represented in the figure via a ball. The position of each ball complies with a given ordering of sectors, which for buying sectors runs from top to bottom of the vertical axis, whereas for selling sectors, it runs from left to right of the horizontal axis. This ordering corresponds with the acyclical topological sorting of the input-output graph which is implied by the removal of the “feedback arc set” of edges that induce cycles in the graph, as per the algorithm by Eades, Lin and Smyth (1993). Small transaction shares are not dropped before the topological sorting is determined.

Source: Revec.

Figure C.4 represents the Costa Rican input-output network graphically. The balls it displays are sized proportionally to the share of all transactions directed, over time, from one sector to another, though small shares (amounting to less than 5 per cent) are not visualized. The remarkable sparsity of balls in the upper triangle of the matrix implies that the network is *approximately acyclical*: that is, there is a “pecking order” of industries such that sectors earlier in the sequence seldom buy from those in later positions. This observation echoes the analogous findings by Liu (2019) for China and South Korea. The total *value* of all transactions in the upper-right triangle (before normalizing entries of the input-output matrix to shares) is 6.2% of the total. This finding is instrumental towards motivating Assumption 3 of our model and the associated implications.

We find it useful to elaborate on the construction of Figure C.4: it is reproduced through the following steps. Here, s indicates a buyer’s sector, and z a seller’s sector.

- 1) Aggregate all (inflation-adjusted) transactions occurred over time involving the same pair of *sectors*. Calculate

$$Total_{sz} = \sum_{t=1}^T \sum_{i \in S_s} \sum_{j \in S_z} Transaction_{ijt}.$$

- 2) For each buyer sector, calculate the share, relative to every seller sector, of the grand total as

$$Share_{sz} = \frac{Total_{sz}}{\sum_{z'=1}^S Total_{sz'}}.$$

- 3) Remove the feedback arc set (the collection of sector pairs individually leading to violations of acyclicity) using the algorithm by Eades, Lin and Smyth (1993).
- 4) Obtain a *topological sort* of sectors: a cyclical ordering.
- 5) Reintroduce the feedback arc set removed two steps earlier.
- 6) Plot Figure C.4 without displaying pairs for which $Share_{sz} < 0.05$.

This procedure differs from the one adopted by Liu (2019) to show similar patterns for the Chinese and South Korean input-output matrices. Instead of following steps 3 to 5 just described, Liu (2019) sorts sectors according to the measure of distortion centrality that is central to his theory about industrial policy. Consistently with our distinct aims, our method identifies instead the “least cyclical” sorting order of sectors. Arguably, applying this method to the Chinese and South Korean data by Liu (2019) is likely to result in more apparent “approximately acyclical” patterns. Note that the temporary removal of the feedback arc

set (step 3) does not, *per se*, lead to a sparser upper triangle in Figure C.4, as all the involved connections are re-introduced at step 5. The topological sort from step 4, however, causes most of the non-truncated entries of the input-output matrix to appear on the left-hand side of Figure C.4.

The quasi-acyclical property of the Costa Rican input-output matrix cannot be appreciated without truncating lower-value entries and appropriately sorting sectors. Figure C.5 displays an unstructured version of the input output matrix whose entries are proportional to the logarithmic number of transactions observed over time (using levels instead of logs would seriously hinder the visualization of most small-sized cells). This figure is hardly suggestive of any specific pattern. Again, this observation echoes one by Liu (2019), who shared analogous considerations about the Chinese and South Korean input-output matrices. Accounting for small entries, the upper triangle of the input-output matrix accounts for about 29% of the total value of all transactions. It must be remarked that this is an overestimate, since small entries are not taken into account by the algorithm we adopted to derive the topological sort.

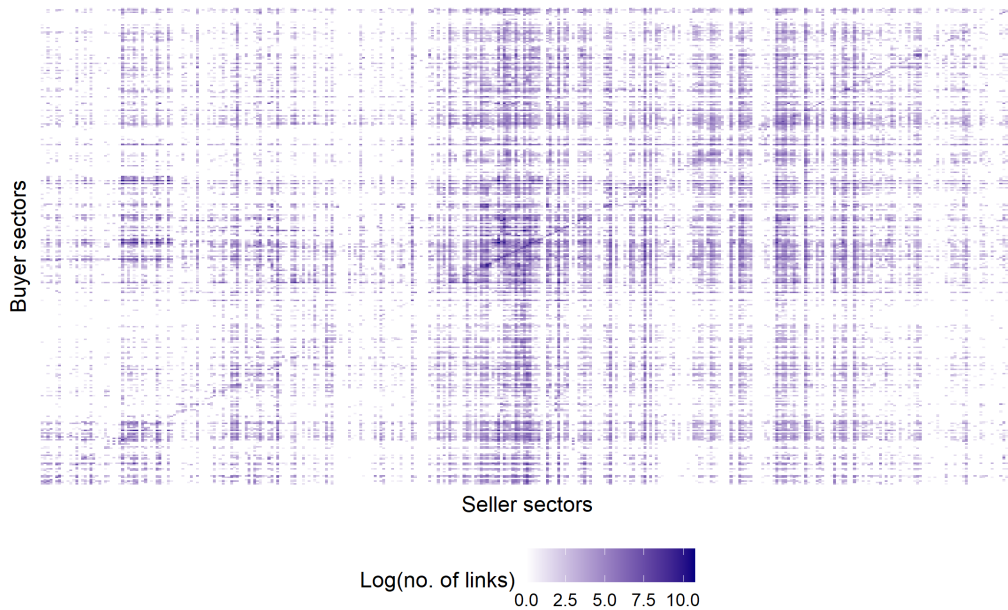


Figure C.5. : A raw, pooled Costa Rican input-output matrix

Note: This figure depicts the input-output matrix of the Costa Rican economy. Each cell represents a pair of sectors, one for buyers and one for sellers; the intensity of the color in a cell is proportional to the logarithm of yearly directed relationships observed in it over time (2008-2017). On both axes, sectors are ordered according to their numerical value in standard industry classification adopted by Costa Rica at the four-digits level (buyers: bottom to top; sellers: left to right).

Source: *Revec.*

Fact 3. *Buyer-supplier relationships are considerably persistent in time.*

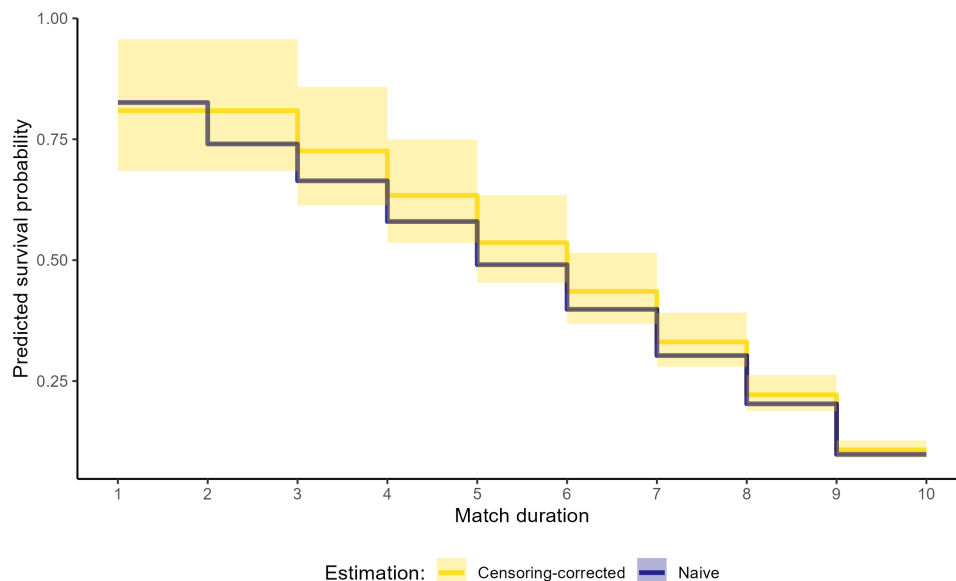


Figure C.6. : Estimates of survival probabilities for firm-to-firm transactions

Note: This figure reports estimates of survival probabilities of transactions, along with 95% confidence intervals, based on extended “transaction spells” that allow for interruptions (see the discussion in the text). The blue step function reports “naive” estimates based on the observed proportions of spells of different duration (the associated confidence intervals are too small to be visually appreciable). The yellow step function reports estimates that correct for data censoring at both ends of a spell based on the estimator by Turnbull (1974). Each step marks the estimated probability of survival for as many years as indicated at the right end of the corresponding interval. Because the Turnbull estimator does not yield an estimate for survival at year 2, its initial step extends over year 3.

Source: Revec.

Buyer-supplier relationships between firms are all but ephemeral. To quantify the persistence of matches, we define “transaction spells” as the time interval bounded by the first and the last observation of a transaction between any two distinct firms (this allows for occasional gaps of one-two years where no transactions are reported), and we estimate the survival probabilities of all observed spells. As Figure C.6 represents, a naive count suggests that about half of matches last at least five to six years, and about a quarter of them lasts nine years of more. However, these simple estimates do not account for censoring: matches observed in 2008 may have started earlier, while the termination of spells still ongoing during 2017 is unobserved. Censoring-corrected estimates based on the Turnbull (1974) estimator register, unsurprisingly, even larger magnitudes. We further replicated this exercise with a stricter definition of “spell,” where interruptions are not allowed (and where multiple short spells can be observed for the same pair of firms); the results, that we do not report for brevity, are virtually unchanged.

Corrections aside, even simple and naive estimates are substantially larger than the analogous survival probabilities estimated by Dhyne et al. (2021) on Belgian data. Motivated by this evidence, we introduce sources of path-dependence into the version of our framework adapted to longitudinal firm-to-firm transaction data.

Fact 4. *Across combinations of buyer and seller sectors, the empirical distribution of the share of transactions over the revenue of buyers appears at least bimodal.*

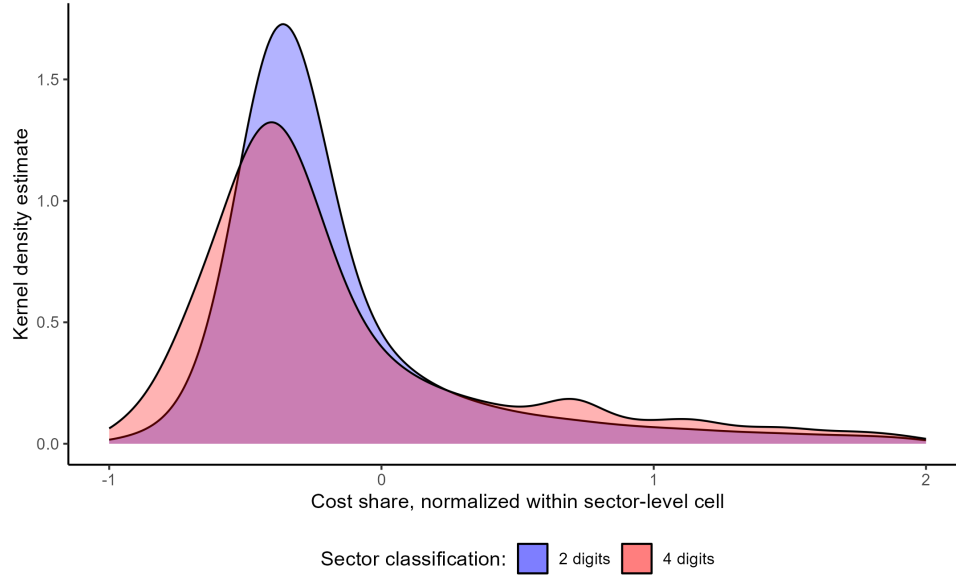


Figure C.7. : Distribution of normalized shares of transactions over buyer revenue

Note: This figure reports kernel density estimates (Gaussian kernel, 0.1 bandwidth size) of the share of transactions over total buyer revenues, standardized separately for each combination of buyer sector, seller sector and year for two classifications of sectors (two- versus four-digits).

Source: Revee.

Figure C.7 displays separate kernel density estimates of “normalized transaction revenue shares,” which are defined as the ratio between transactions and the revenue of the corresponding buyers, then standardized into z -scores within each combination of buyer sector, seller sector, and year. More formally, let

$$Share_{ijt} = \frac{Transaction_{ijt}}{Revenue_{it}}$$

for each transaction occurred between a buyer i and a seller j on year t ; the

normalized shares that yield the kernel density estimates are thus obtained as

$$\text{Normalized share}_{ijt} = \frac{(\text{Share}_{ijt} - \text{avg}(\text{Share}_{ijt}))}{\text{sd}(\text{Share}_{ijt})},$$

where $\text{avg}(\cdot)$ and $\text{sd}(\cdot)$ are the mean and standard deviations of the original shares calculated across all transactions where the buyer is in the same sector as i , the seller as in j , and the year is t . The standardization allows one to appreciate within a single figure features of the empirical distributions of shares that are shared across industry pairs, over time. The two kernel densities differ by the granularity of the sectors that inform the standardization. Both estimates display heavy right tails, as in the analysis by Dhyne, Kikkawa and Magerman (2022) of analogous (non-standardized) ratios. However, the estimate based on the four-digits classification appears bi- if not possibly multi-modal.

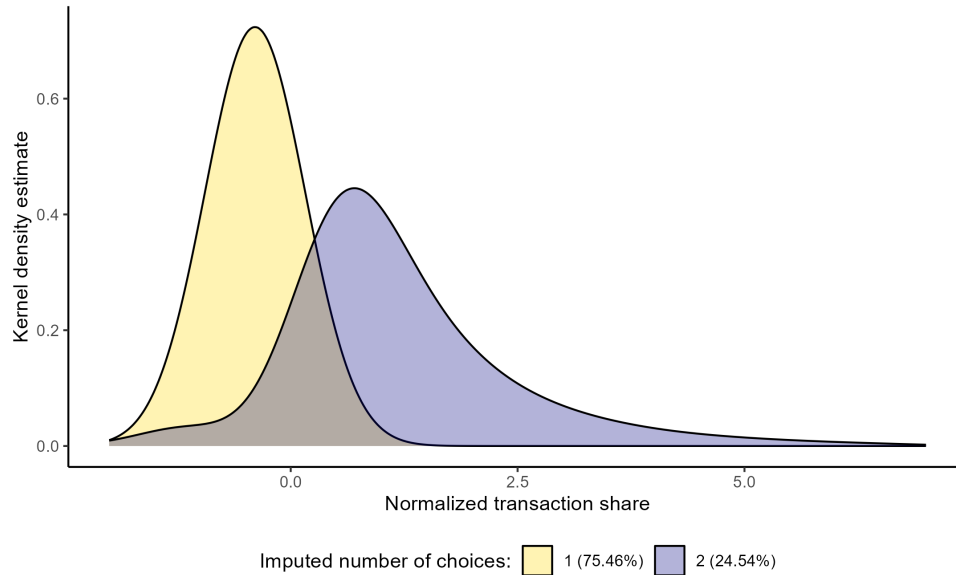


Figure C.8. : Imputation of two latent classes from the normalized shares of transactions

Note: This figure represents kernel density estimates (Gaussian kernel, 0.1 bandwidth size) of two latent classes of the normalized transaction revenue shares (at the four-digits level) imputed through a Gaussian mixture. The legend reports the sample share of each class (in percentage).

Source: Revee.

Normalized shares can be used to infer information from the intensive margin of transactions (typically observed in the data). As the inspection of (22) in the paper reveals, transaction shares map parameters and outcomes of the network formation process; they can be used, for example, to construct patterns of group match effects, as discussed in Section II. These would be inferred from estimating

Gaussian mixtures of regressions on the normalized shares (over the pooled panel) using z_{ijt} as predictors. A simpler use is the imputation of censored tasks, again via mixture models, to deliver outcomes as the one represented in Figure C.8. The latter is only intended for the sake of illustration: as the Monte Carlo results from Table 2 of Section III show, imputing tasks from transactions is an unnecessary, if not biased, approach.

Fact 5. *Connections are disproportionately concentrated at short spatial distances.*

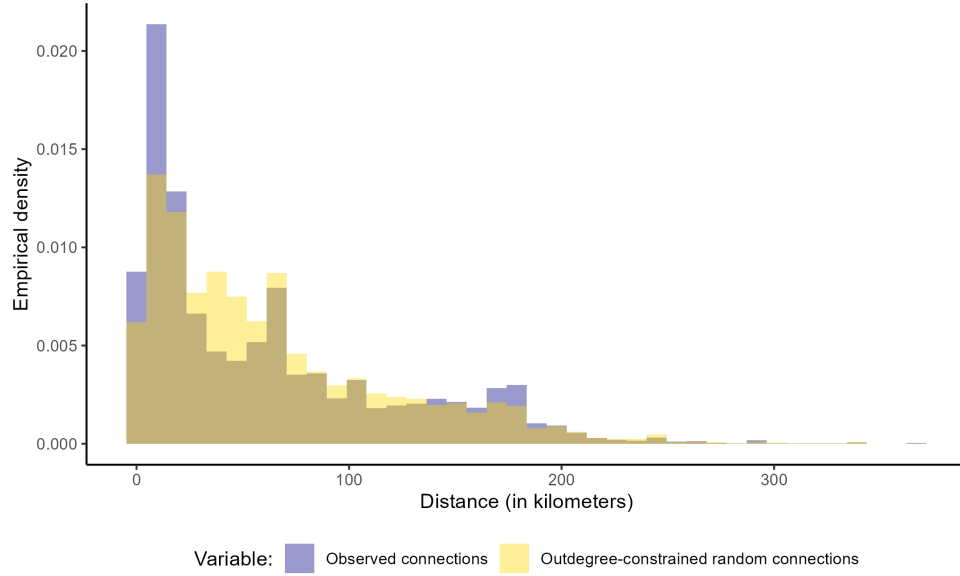


Figure C.9. : Observed and randomized connections as a function of spatial distance

Note: This figure reports histograms of the spatial distance (in kilometers) between the buyers and sellers of: the transaction observed in the data (blue histogram) and random transactions that preserve the outdegree of sellers across buyer sectors (yellow histogram).

Source: Revec.

Figure C.9 represents the histogram of distance calculated over two distinct lists of Costa Rican buyers and sellers *pairs*: all observed transactions in the data, and a list of randomized transactions that preserve the out-degree of all sellers across sectors in every year. As expected, the observed transactions are exceedingly concentrated at lower distances. This replicates the analogous Fact 3 by Bernard, Moxnes and Saito (2019), though to a quantitatively much less extreme extent. This contrast is easily explained through differences in geography, size, and longitudinal extension between Japan (the setting examined by Bernard, Moxnes and Saito, 2019) and Costa Rica.

D. Supplement to the empirical application

This appendix elaborates on some miscellaneous aspects of our empirical application discussed in Section IV of the text.

Reduced-form estimates. We begin by reporting the reduced-form estimates based on both definitions of binary treatment as discussed in the text. Figure D.1 displays the results of a conventional event study based on the “naive” treatment definition, where the unit of observation is a pair of Costa Rican cantons. As the latter is coarser than firm dyads (the unit of observations in the model and for the RSL estimator) we examine two aggregate outcome variables about firm-to-firm transactions: their total monetary volume and their total count between any two canton pairs. The estimates of Figure D.1 are vaguely suggestive of positive effects in later years, but are imprecise and typically not statistically significant.

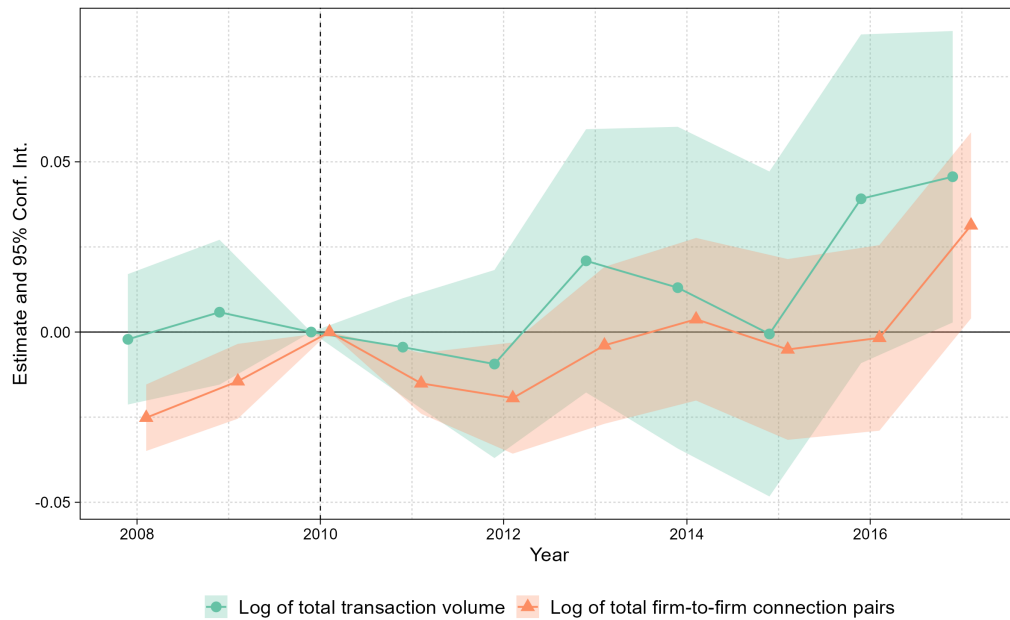


Figure D.1. : DiD estimation of the “naive” treatment effect at the municipality-pair level

Note: This figure displays selected estimated coefficients (along with the associated 95% confidence intervals, represented via shaded bands) from a Two-Way Fixed Effects regression model run on data aggregated at the municipality pair, i.e. buyer canton and seller canton, level. The dependent variables are the log of the total transaction value (green line) and the log of the total connections count (red line). The figure reports “event study” coefficients about the year-specific treatment effect, normalizing the effect at 2010 as zero. Two cantons are treated after 2011 if one belongs to the San José province, and the other belongs to either the Guanacaste or the Puntarenas province. Both regressions feature year fixed effects and municipality-pair fixed effects. Confidence intervals are calculated via two-way clustering (municipality pair and year).
Source: *Revec.*

This finding is easily explained by issues of treatment overlap and hence SUTVA violation, per the discussion in the text around Figure 3. We note that in previous versions of this paper, the RSL estimates of Table 5, Panel A were based on the naive treatment. In those estimates (now unreported), the treatment coefficient loses statistical significance upon adding covariates.

Figure D.2 reports analogous event study estimates, though based on the refined treatment definition described in the text (canton pairs deriving travel time savings from the *Ruta 27*). The estimates are uninformative about a possible treatment effect for two reasons. First, all the pre-treatment coefficients relative to 2010 are non-zero and of magnitude comparable to the post-treatment ones, suggesting that 2010 is an anomalous reference year for an event study (consistent with the *Ruta 27* being only partially operational during that year) and that the parallel trends assumption does not hold. Second, while post-treatment coefficients are positive, confidence intervals are exceedingly wide throughout. This arguably results from conflating heterogeneous travel-time advantages into a binary treatment: pairs receiving large gains are pooled with pairs receiving negligible ones, and the estimated coefficient corresponds to an opaque weighted average over the distribution of treatment intensities. The latter, in turn, is contaminated by network effects per the discussed issues of SUTVA violation.

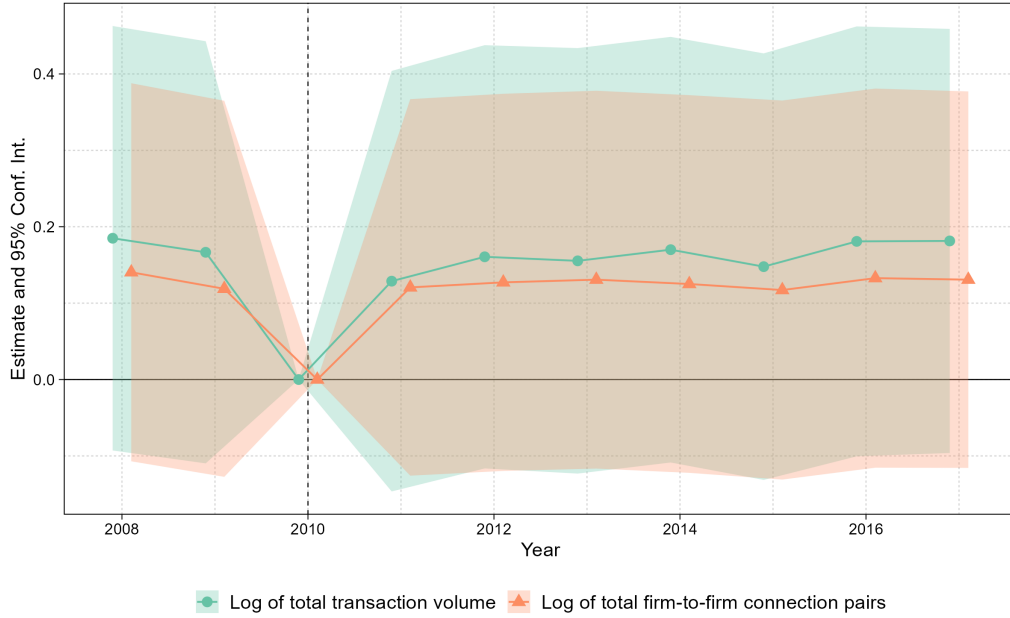


Figure D.2. : DiD estimation of the binary treatment effect at the municipality-pair level

Note: This figure displays event study estimates analogous to those of Figure D.1, but based on an improved definition of treatment: two cantons are treated if they enjoy travel time savings thanks to the *Ruta 27*. Refer to the notes to Figure D.1 for other details. *Source:* Revec.

Unreported estimates show that replacing the binary indicator with the continuous travel-time savings yields qualitatively identical patterns (only the scale of the vertical axis differs), confirming that even intensity-based identification cannot recover a clean treatment effect. We also implemented “triple differences” specifications as in Bernard, Moxnes and Saito (2019), in order to control for demand-side effects as summarized by the degree of input intensity across industries. Estimates of the productivity effect of post-2010 exposure to the *Ruta 27* on *all firms* located in Puntarenas and Guanacaste are not statistically significant, and are thus omitted.

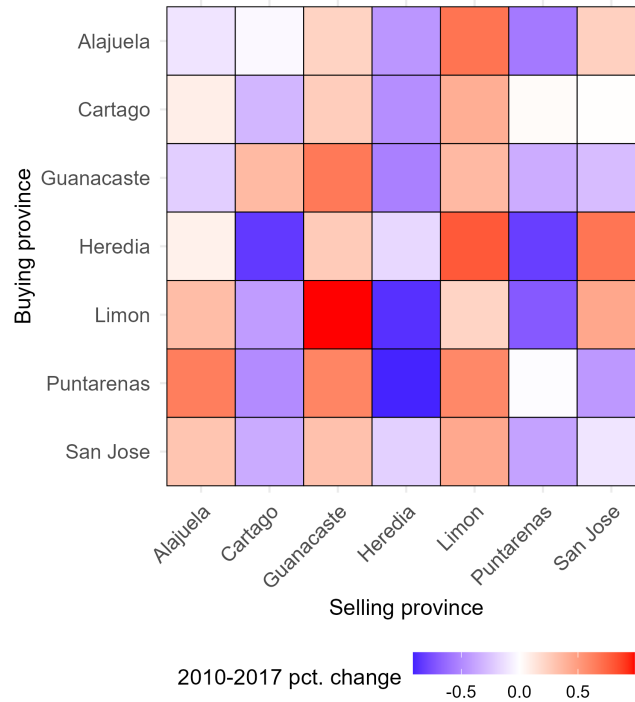


Figure D.3. : Changes in the geography of the Costa Rican production network, 2010-2017

Note: For every pair of Costa Rican provinces, this heatmap represents the differences between their share of total firm-to-firm connections in 2017 and that in 2010 (red: positive difference; blue: negative difference). Shares and their changes are expressed in percentage terms.

Source: *Revec.*

Geographical evolution of the production network. In figure D.3 we report some interesting statistics about the evolution of the Costa Rican production network in both time and space. For every pair of Costa Rican provinces p' and p'' , we calculate at $t \in \{2010, 2017\}$ the proportion of observed buyer-seller

connections:

$$P(\mathbf{G}_t, p', p'', t) = \frac{\sum_{i \in \mathcal{I}_t} \sum_{j \in \mathcal{I}_t} g_{ijt} \mathbb{1}[p(i) = p'] \mathbb{1}[p(j) = p'']}{\sum_{i \in \mathcal{I}_t} \sum_{j \in \mathcal{I}_t} g_{ijt}}$$

where $p(i)$ indicates the province where firm $i \in \mathcal{I}_t$ is located. For all province pairs, Figure D.3 reports the difference $P(p', p'', 2017) - P(p', p'', 2010)$.

The Figure shows that the Costa Rican production networks evolved over time in heterogeneous directions. Both the Heredia and Puntarenas provinces, for example, lost buyers for their sellers (in relative terms), despite their central position and their proximity to the *Ruta 27*. Conversely, the Guanacaste province (which is the farthest from the center, but that uniformly benefited from the *Ruta 27* in terms of reduced travel times by about thirty minutes) experienced a surge across all its relative shares on the supplier side. We interpret this as suggestive evidence of substitution patterns induced by the *Ruta 27*, which we discard in light of our ALDOR estimates.

Characteristics of the dropped subnetworks. As discussed in the text, before performing RSL estimation we drop subnetworks with less than 20 transactions, but this can cause issues of interpretation of the estimates if small subnetworks are selected on particular dimensions. Figure D.4 dispels this concern: it represents kernel density estimates for four subnetwork-level variables, each time separately for selected (cyan) and dropped (red) subnetworks. The figure shows that while the dropped subnetworks are smaller, which is so by construction, the share of treated dyads (pairs of observed transactions in a subnetwork that save on travel times thanks to the *Ruta 27*) share a common support between the two groups. Therefore, our sample cut rule does not appear to exceedingly select on the “treatment.”

Figure D.4 also shows that the retained subnetworks feature a lower dispersion of geographical distance between buyers and sellers (this is a mechanical effect: small subnetworks are often found within the metropolitan core) and about the same share of lagged transactions as dropped subnetworks, but with a much lower dispersion.

Multinomial logit estimates with GFEs. We compare how RSL estimators fare against a seemingly simpler model for buyer choice: the plain logit model with Group Fixed Effects (GFEs) as in the first Monte Carlo exercise from Section III in the text. Thus, we estimate a number of specifications, symmetric to those of our RSL estimates from Table 5, of a multinomial logit model where:

- the unit of observation is an observed transaction whose un-normalized revenue share does not exceed 1;
- the set of potential alternatives are those firms that share the same seller sector as the observed seller in a transaction;

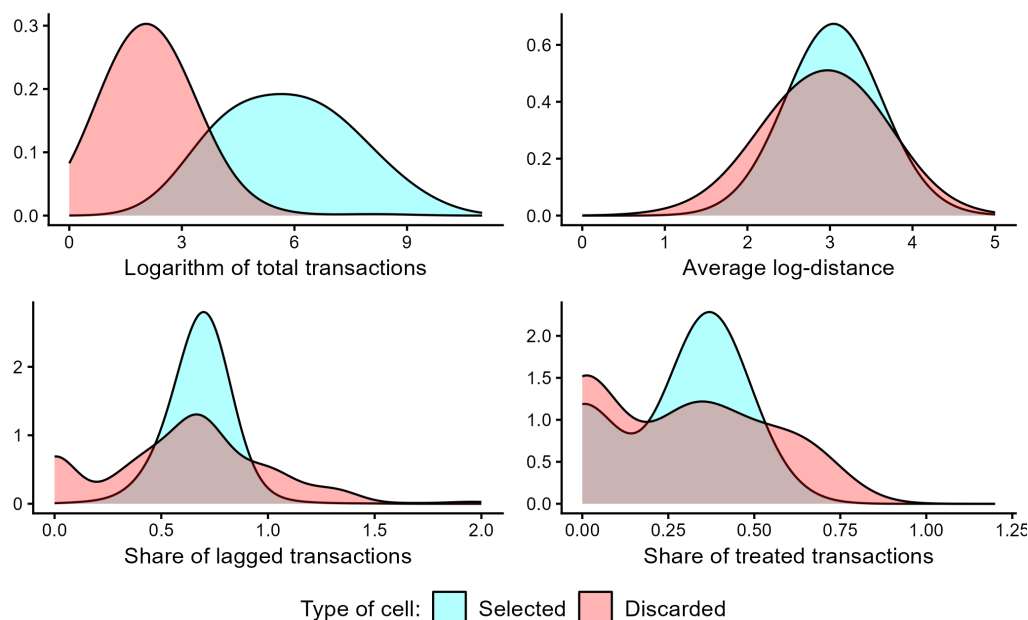


Figure D.4. : Analysis of subnetworks (un-)selected for RSL estimation

Note: This figure reports kernel density estimates (Gaussian kernel, variable bandwidth) about four variables, one per each subfigure, measured at the *subnetwork* or “cell” level, where subnetworks are as discussed in Section IV. Each subfigure reports two estimates: one for the subnetworks used to inform the RSL estimates of Table 5 (“selected,” cyan) and one for the unused subnetworks (“discarded,” red). In the bottom-right subfigure, “shares” are calculated using, in the denominator, the number of tasks imputed for each transaction as per the mixture model estimates reported in Figure C.8. All averages are then calculated at the cell level, hence the support running from 0 through 2.

Source: *Revec.*

- however, following McFadden (1978), the alternatives entering the denominator of the likelihood function probabilities are the observed seller and a number of uniformly sampled eligible firms (here: five);
- the estimates are weighted by the number of imputed tasks for each transaction, as per the clustering results of Table C.8.

To let the BCCR computers handle the estimation problem, we are compelled to use only a random sample of 5 per cent of all eligible transactions for each alternative. To construct GFEs, we perform an exercise in *k*-means clustering based on two firm-level variables: total customer count, weighted by the imputed number of tasks, and total value of sales, preliminarily normalized by the total yearly purchases of every buyer. As a result, we infer 12 clusters, and we associate a dummy variable to each.

The results from this estimation exercise are reported in Table D.1. Remarkably, the two variables of interest (dyadic treatment, travel times) register extremely small estimates, which are not statistically significant. Conversely, all

Table D.1—: Empirical application: multinomial logit estimates with GFEs

Panel A: discrete treatment specification				
	(1)	(2)	(3)	(4)
Treatment (“ <i>Ruta 27</i> ”)	0.002 (0.008)	0.001 (0.008)	0.001 (0.008)	−0.010 (0.012)
log(Seller size)		0.059*** (0.003)	0.054*** (0.003)	0.026*** (0.004)
log(Seller import intensity)			0.004*** (0.001)	0.007*** (0.001)
Transaction at $t - 1$				5.221*** (0.016)
Distance decile controls	YES	YES	YES	YES
Group Fixed Effects	YES	YES	YES	YES
Akaike information criterion	636,856.3	636,385.2	636,340.2	312,957.8
Number of choices	201,336	201,336	201,336	201,336
Panel B: continuous regressor specification				
	(5)	(6)	(7)	(8)
log(Travel time)	−0.002 (0.003)	−0.002 (0.003)	−0.002 (0.003)	0.003 (0.004)
log(Seller size)		0.059*** (0.003)	0.054*** (0.003)	0.026*** (0.004)
log(Seller import intensity)			0.004*** (0.001)	0.007*** (0.001)
Transaction at $t - 1$				5.221*** (0.016)
Group Fixed Effects	YES	YES	YES	YES
Akaike information criterion	636,840.1	636,370.0	636,325.3	312,950.6
Number of choices	201,336	201,336	201,336	201,336

Note: This table reports multinomial logit estimates of observed transactions using the same covariates as in Table 5. Each panel corresponds to one of two sets of specifications discussed in the text. All models control for unobserved heterogeneity through Group Fixed Effects, that are constructed via k -means estimates as described in this Appendix. Estimates are based on a random 10 percent sample of all transactions used to calculate the descriptive statistics in Table 4; the corresponding alternative sets are drawn following the sampling scheme detailed in that table’s notes. Each observation is weighted by its imputed number of tasks, derived from the clustering results displayed in Table C.8. Asterisks *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Source: *Revec.*

other point estimates are larger by one or more orders of magnitude than the corresponding RSL estimates from Table 5, and are statistically significant at the 1 per cent level throughout. Signs are usually preserved, except for the seller size predictor, as expected. This suggests that neglecting seller fixed effects can seriously bias one's estimates; interpretations vary by explanatory variable. For example, the binary treatment is likely negatively correlated with the seller effects because the *Ruta 27* connects San José with relatively less productive and developed areas of Costa Rica; conversely, travel time negatively correlates with the cost-effectiveness of sellers due to agglomeration economies. For control variables, the analysis proceeds analogously. In particular, the large bias that seems associated with the "lag connection" dummy must be attributed to persistence in the unobserved productivity of sellers.

Network counterfactuals: algorithm and analysis. Lastly, we elaborate on the ALDORs introduced in Section IV. We first describe the algorithm used to construct the counterfactual subnetworks $\mathbf{H}_{st}$. For $t = 2010, \dots, 2016$ and $s = 1, \dots, S$:

- 1) we select the adjacency submatrix $\mathbf{G}_{st'}$ in $t' = 2017$,
- 2) we draw $\tilde{R}$ candidate alternative networks via the algorithm by Patefield (1981),
- 3) we select the counterfactual submatrix $\mathbf{H}_{st}$ that minimizes some given measure of distance between $P(\mathbf{G}_{st'}, p', p'', t)$ and $P(\mathbf{H}_{st}, p', p'', t')$ across all (p', p'') pairs (the evaluation of both distributions is restricted to sellers from industry s).

We experiment with four related distance measures: Manhattan (L^1), Euclidean (L^2), Chebyshev (L^∞) and Hellinger. We set $\tilde{R} = 5$: as shown in Figure D.5, this is enough to select simulated subnetworks fairly close to the real ones across all four measures of distance. Subnetworks with less than 20 actually observed connections in one year are preliminarily dropped, consistently with the RSL estimates of Table 5.

The selected subnetworks are aggregated into a single counterfactual adjacency matrix $\mathbf{H}_t$ (with gaps where subnetworks are dropped), and ALDORs are computed by rescaling the denominator of (23) to account for dropped subsectors. In addition, each ALDOR is associated with an estimated variance, which is calculated as:

$$\begin{aligned} \widehat{\text{Var}} \left[\text{ALDOR}_t \left(\mathbf{H}_t, \mathbf{G}_t, \mathbf{Z}_t^0, \mathbf{Z}_t^1; \hat{\boldsymbol{\beta}}_{RSL} \right) \right] &= \\ &= \frac{1}{S^2} \sum_{i \in \mathcal{I}_t} \sum_{j \in \mathcal{I}_t} (h_{ijt} - g_{ijt})^2 \left(\mathbf{z}_{ijt}^1 - \mathbf{z}_{ijt}^0 \right)' \widehat{\text{Var}} \left[\hat{\boldsymbol{\beta}}_{RSL} \right] \left(\mathbf{z}_{ijt}^1 - \mathbf{z}_{ijt}^0 \right), \end{aligned}$$

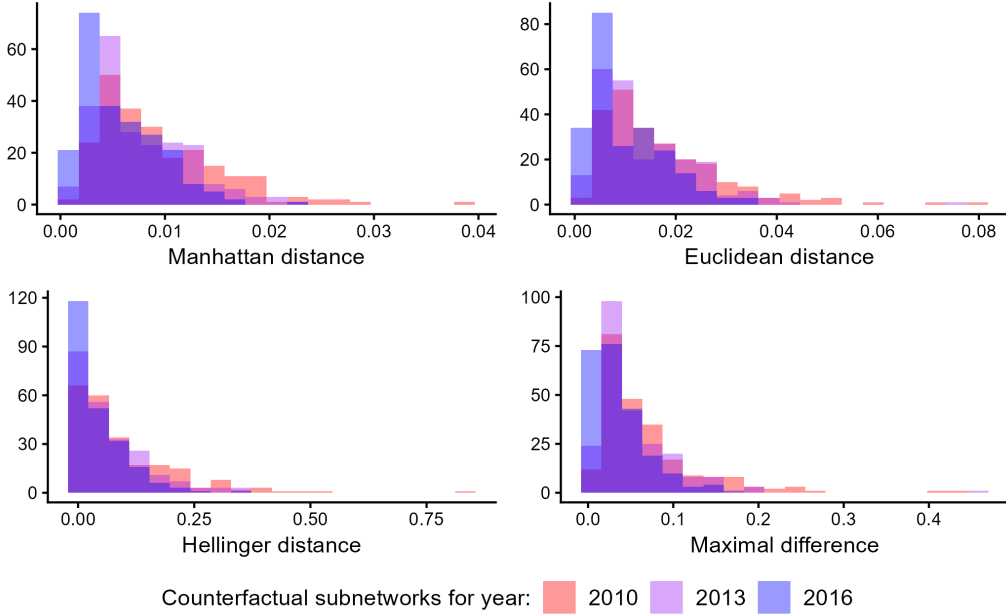


Figure D.5. : Benchmarking of the simulated subnetworks used for counterfactuals

Note: This figure reports histograms of distances between subnetworks $\mathbf{G}_{st'}$ (observed in 2017) and $\mathbf{H}_{st}$ (counterfactuals for ALDOR estimation) in terms of the empirical distribution of cross-province connections. Each histogram is specific to one of four distance measures used to select counterfactual subnetworks (Manhattan, Euclidean, Hellinger and Chebyshev). Different colors display distances for $t = 2010$, $t = 2013$ and $t = 2016$ (representative years), as indicated in the legend.

Source: Revee.

where, if necessary, S^2 in the initial denominator is adjusted to account for dropped subnetworks. The sandwich expressions that are summed over feature the estimated variance of the RSL estimates. This formula follows from simple probability theory under the assumption that $(\mathbf{H}_t, \mathbf{G}_t, \mathbf{Z}_t^0, \mathbf{Z}_t^1)$ are nonstochastic, and all the variability is due to the RSL estimates.

Figure D.6 reports the estimated ALDORs discussed in the text, along with their associated confidence intervals and a dashed line that displays the share of each year's ALDOR attributed to the top province pair contributor (as elaborated below). These estimates are based on counterfactual subnetworks $\mathbf{H}_{st}$ selected with the Manhattan distance; ALDORs based on other measures are virtually identical and thus omitted.

To evaluate the contribution of all different province pairs (p', p'') to the ALDOR, we calculate for all such pairs and for all $t = 1, \dots, T$ the following quantities:

$$LDOR_{p', p'', t}(\mathbf{H}_t, \mathbf{G}_t, \mathbf{Z}_t^0, \mathbf{Z}_t^1; \hat{\beta}_{RSL}) = \sum_{i \in \mathcal{P}_{p' t}} \sum_{j \in \mathcal{P}_{p'' t}} (h_{ijt} - g_{ijt}) (z_{ijt}^1 - z_{ijt}^0)' \hat{\beta}_{RSL},$$

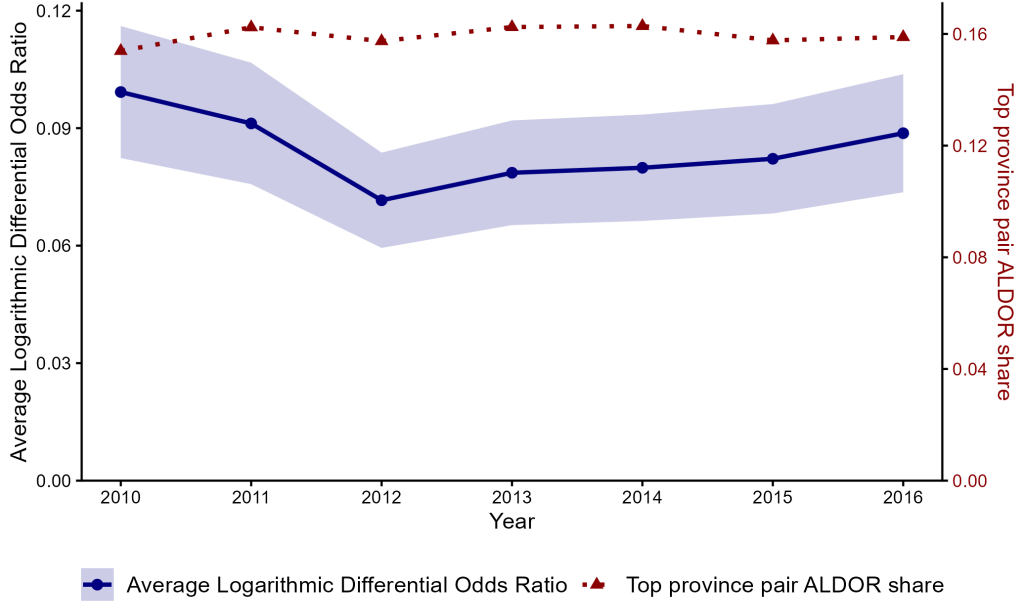


Figure D.6. : Average Logarithmic Differential Odds Ratios, over time

Note: The blue line reports the ALDOR as defined in the text for all years indicated on the horizontal axis; these are measured on the left vertical axis. Shades represent their confidence intervals at the 95 percent level, computed from the variance formula described above. All $\mathbf{H}_{st}$ subnetworks are selected with the Manhattan distance. The red dashed line indicates, for all years, the top contribution share to the ALDOR among pairs of Costa Rican provinces (consistently buyers from Alajuela, sellers from San José from 2010 to 2016); these are measured on the right vertical axis.

Source: Revee.

where $\mathcal{P}_{pt}$ is the set of all firms located in province p at time t . Because the collection of all $\mathcal{P}_{pt}$ sets is a partition of $\mathcal{I}_t$ for $t = 1, \dots, T$, the sum of all pair-specific LDORs in year t , divided by S , equals that year's ALDOR. However, the discussion in the text is informed by the *overall contribution shares* (OCS) over the relevant years:

$$OCS_{p',p''} = \frac{\sum_{t=2010}^{2016} LDOR_{p',p'',t}}{\sum_{t=2010}^{2016} \sum_{p'} \sum_{p''} LDOR_{p',p'',t}},$$

which we represent across pairs (p', p'') through the heatmap of Figure D.7 below.

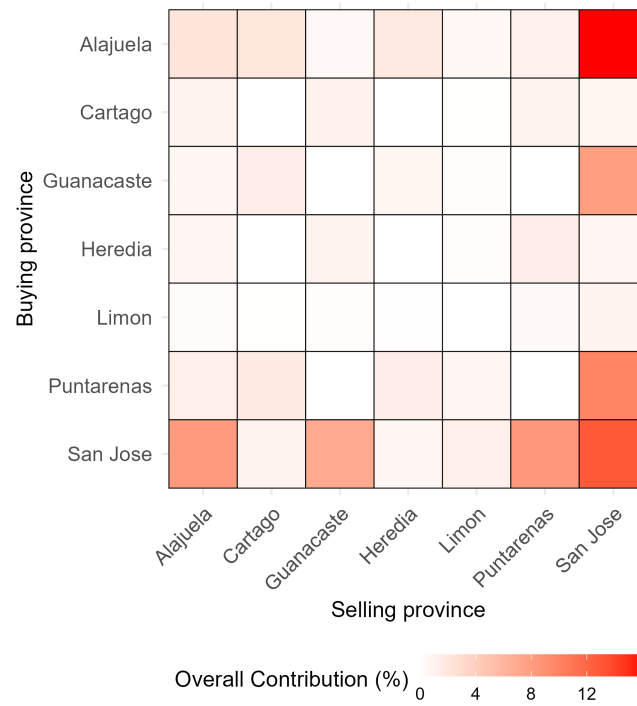


Figure D.7. : ALDOR contributions over time, by province pairs

Note: For every pair of Costa Rican provinces, this heatmap represents its contribution to the overall LDOR (OCS) between 2010 and 2016, in percentage terms. *Source:* Revec.

References

- Alfaro-Ureña, Alonso, Ben Faber, Cecile Gaubert, Isabela Manelici, and Jose Vasquez.** 2023. “Responsible Sourcing? Theory and Evidence from Costa Rica.” Unpublished working paper.
- Alfaro-Ureña, Alonso, Isabela Manelici, and Jose Vasquez.** 2022. “The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages.” *Quarterly Journal of Economics*, 137(3): 1495–1552.
- Alfaro-Ureña, Alonso, Mariany Fuentes-Fuentes, Isabela Manelici, and Jose Vasquez.** 2018. “Costa Rican Production Network: Stylized Facts.” Banco Central de Costa Rica: Documento de Investigación N. 002.
- Antràs, Pol, Teresa C. Fort, and Felix Tintelnot.** 2017. “The Margins of Global Sourcing: Theory and Evidence from U.S. Firms.” *American Economic Review*, 107(9): 2514–2564.
- Bernard, Andrew B., Andreas Moxnes, and Yukiko U. Saito.** 2019. “Production Networks, Geography, and Firm Performance.” *Journal of Political Economy*, 127(2): 639–688.
- D’Haultfœuille, Xavier, and Alessandro Iaria.** 2016. “A convenient method for the estimation of the multinomial logit model with fixed effects.” *Economics Letters*, 141: 77–79.
- Dhyne, Emmanuel, Ken Kikkawa, and Glenn Magerman.** 2022. “Imperfect Competition in Firm-to-Firm Trade.” *Journal of the European Economic Association*, 20(5): 1933–1970.
- Dhyne, Emmanuel, Ken Kikkawa, Magne Mogstad, and Felix Tintelnot.** 2021. “Trade and Domestic Production Networks.” *The Review of Economic Studies*, 88(2): 643–668.
- Dhyne, Emmanuel, Ken Kikkawa, Xianglong Kong, Magne Mogstad, and Felix Tintelnot.** 2023. “Endogenous Production Networks with Fixed Costs.” *Journal of International Economics*, 145. 103841.
- Eades, Peter, Xuemin Lin, and William F. Smyth.** 1993. “A fast and effective heuristic for the feedback arc set problem.” *Information Processing Letters*, 47(6): 319–323.
- Gourieroux, Christian, Alain Monfort, and Alain Trognon.** 1984. “Pseudo Maximum Likelihood Methods: Theory.” *Econometrica*, 52(3): 681–700.
- Liu, Ernest.** 2019. “Industrial Policies in Production Networks.” *Quarterly Journal of Economics*, 134(4): 1883–1948.

- McFadden, Daniel.** 1978. "Modeling the choice of residential location." *Transportation Research Record*, 637.
- Panigrahi, Piyush.** 2023. "Endogenous Spatial Production Networks: Quantitative Implications for Trade and Productivity." Working paper.
- Patefield, William M.** 1981. "Algorithm AS 159: An Efficient Method of Generating Random $R \times C$ Tables with Given Row and Column Totals." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 30(1): 91–97.
- Turnbull, Bruce W.** 1974. "Nonparametric Estimation of a Survivorship Function with Doubly Censored Data." *Journal of the American Statistical Association*, 69(345): 169–173.
- White, Halbert.** 1994. *Estimation, Inference and Specification Analysis. Econometric Society Monographs*, Cambridge University Press.