Matching to Suppliers in the Production Network: an Empirical Framework*

Alonso Alfaro-Ureña† and Paolo Zacchia‡

March 2024

Abstract

This paper develops a framework for the empirical analysis of the determinants of input supplier choice on the extensive margin using firm-to-firm transaction data. Building on a theoretical model of production network formation, we characterize the assumptions that enable a transformation of the multinomial logit likelihood function from which the seller fixed effects, which encode the seller marginal costs, vanish. This transformation conditions, for each subnetwork restricted to one supplier industry, on the out-degree of sellers (a sufficient statistic for the seller fixed effect) and the in-degree of buyers (which is pinned down by technology and by "make-or-buy" decisions). This approach delivers a consistent estimator for the effect of dyadic explanatory variables, which in our model are interpreted as matching frictions, on the supplier choice probability. The estimator is easy to implement and in Monte Carlo simulations it outperforms alternatives based on group fixed effects. In an empirical application about the effect of a major Costa Rican infrastructural project on firm-to-firm connections, our approach yields estimates typically much smaller in magnitude than those from naive multinomial logit.

JEL Classification Codes: C25, L11, R12, R15

Keywords: Production network, Supplier choice, Conditional logit, Infrastructures

---

*We are grateful to Pol Antràs, Bryan Graham, Jose Vasquez and all participants to the presentations held at the 2019 European Meeting of the Econometric Society and at the 2019 Northwestern junior workshop on the econometrics of networks for their insights and advice. We express gratitude for the central role of Isabela Manelici and especially Jose Vasquez in coordinating our collaboration on this project. We owe special thanks to Santiago Campos-Rodríguez as well as César Ulate for their invaluable research assistance. This project was enabled by financial support primarily from Charles University's PRIMUS/21/HUM/022 grant, and in addition, the University's UNCE project (UNCE/HUM/035). The processing of data required to produce the results shared in this document were conducted at the Banco Central de Costa Rica (BCCR) under a secure technological procedure designed to guarantee the integrity and confidentiality of the information. All results have been reviewed by the BCCR to ensure no confidential information is disclosed. The views expressed herein are solely those of the authors and are not those of the BCCR. All errors are our own.

†Banco Central de Costa Rica and Universidad de Costa Rica; e-mail: `alfaroua@bccr.fi.cr`.

‡Charles University and the Czech Academy of Sciences, e-mail: `Paolo.Zacchia@cerge-ei.cz`.

1 Introduction

Establishing a buyer-supplier relationship between a firm located downstream in the production network and one located more upstream is a consequential decision. For both buyers and sellers, the identity of their trading partner can affect performance (Alfaro-Ureña et al., 2022). For sellers, expanding the set of customer firms is a key driver of their growth (Bernard et al., 2022). For the economy at large, the structure of the production network that results from aggregating all such choices affects the propagation of macroeconomic shocks (Acemoglu et al., 2012; Carvalho et al., 2021). Understanding how buyer-supplier linkages are determined bears implications for the study of industrial policy,[1] agglomeration economies,[2] firms in developing economies,[3] and more. In these settings, factors as diverse as policies targeting selected segments of supply chains, spatial proximity, and personal connections across firms, are potential drivers of production network formation.

This paper develops an empirical framework to quantitatively assess the impact of factors like these on the extensive margin of firms' supplier choice, to be used on data about firm-to-firm transactions.[4] In particular, we study how to estimate the effect of explanatory variables which, as in the previous examples, display a *dyadic variation*: that is, they depend on the specific buyer-seller pair. Relative to conventional models of multinomial choice, this particular problem presents a central challenge: the prices

[1]Building upon insights originally by Hirschmann (1958), Liu (2019) has shown that industrial policies targeting upstream sectors of the production network can generate positive aggregate effects under market imperfections affecting the extensive margin of input demand. Lane (2023) has shown how policies of this sort were key for the historical industrial development of South Korea.

[2]Theories of agglomeration economies typically conjecture that the ability to access specialized suppliers that are close in space generates economies of scale (Duranton and Puga, 2004; Moretti, 2011). By calibrating a model of firm-to-firm matching in space on Japanese data, Miyauchi (2023) shows that this mechanism explains a substantial portion of the grand agglomeration force. Earlier, Bernard et al. (2019) showed that a decrease in search costs due to the introduction of high-speed trains makes firms more productive and more likely to source their inputs from farther locations.

[3]See e.g. Bartelme and Gorodnichenko (2015) and the discussion in Atkin and Khandelwal (2020). The latter in particular advocate the use of firm-to-firm transaction data to study distortions in trade and development. This resonates with both the main objective of this paper and with the theoretical framework, which features matching frictions, that leads to our econometric model.

[4]This kind of data, typically elaborated from administrative records on value added tax or from firm censuses, typically covers the (quasi-)universe of a country's domestic transactions in a given time period, e.g. a year. The availability of firm-to-firm transaction data for an increasing number of countries has benefited research on production networks. These countries include: Belgium (Bernard et al., 2022), Costa Rica (Alfaro-Ureña et al., 2022), Ecuador (Adão et al., 2022), Japan (Carvalho et al., 2021), Turkey (Demir et al., 2023), Uganda (Almunia et al., 2023) and others.

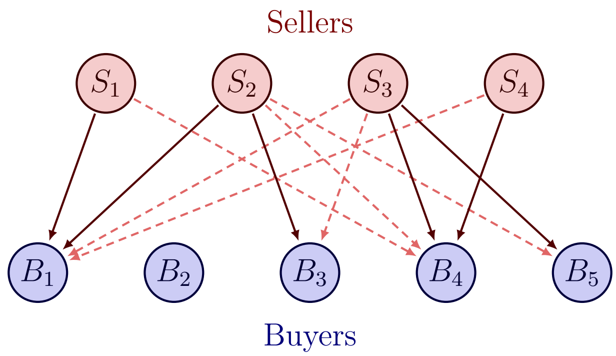
of the alternative choices available to a buyer are typically unobserved in transaction data. In addition, prices are also endogenous because they encode the marginal costs of direct suppliers, which are recursively dependent on the more upstream structure of the production network (better suppliers-of-suppliers make direct suppliers more productive). Since the explanatory variables are likely to co-vary with firm marginal costs, the construction of a consistent estimator for the effects of interest must take into account the implications of the network structure on firms' input demand.

Our framework tackles this problem via a sufficient statistic approach that extends Chamberlain (1980). This approach is supported by assumptions formulated within a model of production network formation where the explanatory variables of interest are treated as *matching frictions*, akin to iceberg trade costs in international trade models. More specifically, we assume that in every time period firms must perform a given number of *tasks* associated with a particular technology; as a result, the total number of distinct tasks fulfilled by a seller within a given time period, as approximated for example by the total number of its buyers in a year,[5] is a sufficient statistic for that seller's equilibrium marginal cost. The intuition is straightforward: a lower marginal cost enables sellers to undercut their competitors and increase their count of trading partners in equilibrium. Our assumption about technology is supported by a stylized fact that we document using Costa Rican data: firms typically purchase their inputs from suppliers that are restricted to a very idiosyncratic set of four-digits sectors.

This insight enables a transformation of the multinomial logit likelihood function that allows consistent estimation of the parameters of interest. This transformation, however, differs substantively from those typical of conditional logit models for panel data, which express the conditional probability for an observed sequence of outcomes *over time*. In fact, our transformation expresses the conditional probability that *in the same time period* (e.g., year) and in a specific section of the production network (the one restricted to sellers of a specific four-digits sector) the configuration of buyer-seller linkages is the one actually observed, conditional on all sellers supplying, and at the same time on all buyers sourcing, the respective observed number of tasks.

[5] A supplier may perform more than one task for the same buyer. To accommodate this possibility, our framework adopts a mixture-of-distributions approach to infer the total number of tasks from the total value of observed bilateral transactions, on the grounds that the share of input purchases for a single task over total firm revenue is pinned by technology. As we document with Costa Rican data, the empirical distribution of the standardized ratios of transaction values over buyers' revenue appears bimodal, which suggests that some buyers source multiple tasks via a single transaction.

In other words, in the transformation the observed subnetworks are pitted against alternative configurations of the same subnetworks that display the same out-degree and in-degree sequences,[6] respectively for buyers and for sellers, as the observed ones. To help build intuition, in Graph 1 we provide an illustration of a stylized subnetwork and two alternative configurations of linkages that share the same degree sequences. Think for example of the solid, darker edges as an observed subnetwork; while the dashed, lighter edges represent an alternative configuration.



Graph 1: Two subnetwork configurations with identical out- and in-degree

Notes. This graph represents two alternative configurations with identical out-degree and in-degree of a stylized bipartite network where "buyers" (blue nodes) and "sellers" (red nodes) are the two sides. The two configurations are represented, respectively, by the solid-dark and dashed-light directed edges.

The transformed likelihood function does not depend on firm marginal costs, but the contributions of each subnetwork-year to it feature a summation over alternative configurations in the denominator. As usual with networks, the enumeration and full specification of all subnetworks with the same degree sequences is a computationally expensive problem which scales non-linearly with (sub)network size. To overcome the resulting curse of dimensionality and operationalize our framework, we restrict the denominators to randomly sampled alternative subnetworks. As shown by McFadden (1978) for the conventional multinomial logit, and later by D'Haultfœuille and Iaria (2016) for its conditional (on fixed effects) version, this approach allows for consistent estimation at the cost of an efficiency loss. Our approach is easy to implement; still, it is useful to evaluate how it compares against alternatives that are arguably even easier

[6] In a directed network: one where connections between nodes are not symmetric, the *out-degree* of a node is the count of the linkages stemming from it (e.g. the number of tasks supplied by sellers), whereas the *in-degree* is the count of linkages directed to it (e.g. the tasks sourced by buyers).

while also more intuitive, such as a plain multinomial logit with fixed effects that are shared by appropriately defined groups of sellers. In Monte Carlo simulations, we show that such an alternative displays a substantial bias even under conditions that are favorable to it, with little to no gain in terms of variance to compensate.

We showcase our framework with an empirical application: we study how spatial distance measured as travel times between two different locations in Costa Rica affects the probability of a connection between a buyer and a seller from such places. Moreover, we examine the effect of a major infrastructural project, the *Ruta 27* (Highway 27) opened in 2011 to facilitate travel between the country's populous central valley and the developing Pacific coastline, with its seaports.[7] While this "treatment" displays a dyadic variation both in geographical space and in time, it is unlikely to be exogenous, since the regions that Highway 27 helps to connect are the most productive ones. In light of this, we find it unsurprising that our approach returns estimates that are much smaller in magnitude, though still statistically significant, with respect to those obtained via a naive multinomial logit that neglects seller fixed effects. The latter also yields predictions about choice probabilities that we find implausibly large, unlike those obtained via our proposed estimator.

Because of its aims, our paper connects with several diverse strands of literature. We consider our contribution primarily as an adaptation of multinomial logit models that condition on fixed effects (see e.g. Chamberlain, 1980; Honoré and Kyriazidou, 2000; D'Haultfœuille and Iaria, 2016; Crawford et al., 2021) to the particular setting of production networks. These models have a reputation for being difficult to implement; as a result, empirical applications are scant.[8] We show that the additional network dimension offered by the input-output structure, where choices are taken over time as well as across multiple tasks, if anything makes implementation easier, which we see as favorable towards applications with micro-level data in the expanding literature on production networks. We also examine a number of extensions of our framework, with features such as random parameters, richer specifications for the seller fixed effects,

[7]Our application thus speaks to the extant literature on the economic effects of infrastructures. Previous studies have typically focused on the effect upon regions (Faber, 2014) or firms (Holl, 2016). To the best of our knowledge, Bernard et al. (2019) is the only contribution that studies the effect on firm-to-firm connections; unlike our empirical application, it takes a reduced-form approach.

[8]Fixed effects are problematic for most non-linear models; for this reason, theoretical econometricians have recently investigated the properties of simpler-to-implement models with "group fixed effects" that assume an underlying discrete structure of unobserved heterogeneity (Bonhomme et al., 2022). These models inspire the estimators adversarial to ours in the Monte Carlo simulation.

and "structural dynamics" (dependence of current choices on past realizations of the network). Except for the extra assumptions that the latter demands, these extensions do not present particular challenges or issues about implementation.

This paper also relates to the literature on the econometrics of network formation, which we contribute to by developing the distinctive case of production networks. The problem of seller fixed effects that we address bears analogies with that of bilateral unobserved heterogeneity in dyadic models for binary, undirected networks. Similarly to Charbonneau (2017) and Graham (2017) for these models, we develop a conditional logit approach. The differences between our approach and theirs stem from the nature of the supplier choice problem: multinomial, constrained by technology, and resulting in a directed network. The econometrics of network formation also emphasizes issues due to multiple equilibria that in undirected networks arise from structural transitivity (Leung, 2015; Mele, 2017; de Paula et al., 2018; Sheng, 2020; Gualdani, 2021).[9] In our framework structural transitivity is absent because firm behavior is governed by cost functions, yet multiple equilibria are still possible as in other models of production network formation. Our sufficient statistic approach is robust to equilibrium selection, as the only variables that co-vary with the equilibria while affecting buyer choice are the seller fixed effects, which disappear in the transformation.

The conceptual framework of this paper is inspired by recent models of production network formation from macroeconomics and international trade, particularly those where matching is *buyer-initiated*,[10] such as Dhyne et al. (2023) (which builds on Antràs et al., 2017), and Panigrahi (2023).[11] Our model adapts that by Dhyne et al. (2023) by introducing some key differences: for example, our "task-based" production functions, our treatment of matching frictions, and our lower emphasis on relationship

[9]Structural transitivity is a primitive property of the preferences of the agents involved in network formation, specifically their taste for connections with agents that share connections with some third party agents ("being friends with friends of my friends"). In strategic models of network formation, structural transitivity yields multiple equilibria, which complicates econometric estimation. For an extended discussion, see the survey by de Paula (2020).

[10]In other models of production network formation, matching is either *seller-initiated* (Lim, 2018; Huneeus, 2020; Bernard et al., 2022) or based on a search-and-matching protocol (Arkolakis et al., 2022; Miyauchi, 2023). We believe that a buyer-initiated matching protocol is more appropriate for our aims, since it ties more naturally with econometric models of discrete choice.

[11]Oberfeld (2018) also developed an input-output model where equilibrium hinges on the choices of entrepreneurs-buyers, though limited to one input; our framework accommodates multiple input choices. In the model by Acemoglu and Azar (2020), buyers choose which and how many suppliers to source from, and the network becomes denser as technology improves or frictions are mitigated. Our framework is one for the short run, where network density is pinned down by technology.

fixed costs. We see the contribution by Panigrahi (2023) as especially close to ours, as it constitutes the attempt to develop an empirical framework for supplier choice to be used on firm-to-firm transaction data. Like ours, his model is based on a multinomial logit specification at the core; however, it differs in two key respects, which we believe warrant our distinct approach. First, Panigrahi (2023) assumes a Cobb-Douglas firm production function; we assume a constant elasticity of substitution (CES) technology instead. Second, he assumes tasks in the production function to be homogeneous and symmetric, potentially fulfilled by any firm regardless of its characteristics; while this allows one to elegantly solve for the seller fixed effects in closed form, we find this at odds with the highly sparse nature of the input-output network, which we believe is best captured by our assumptions on technology-determined tasks.

The rest of this paper proceeds as follows. Section 2 introduces the data and some motivating stylized facts. Section 3 develops the conceptual framework and derives from it the econometric estimator. Section 4 describes the Monte Carlo simulations. Section 5 illustrates the empirical application. Lastly, Section 6 concludes.

2 Data and stylized facts

To motivate and illustrate the proposed framework, we use firm-to-firm transaction data from Costa Rica, which have already appeared in other contributions (among the others, Alfaro-Ureña et al., 2022, 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 an amount is higher than a threshold set at about $4,800.[12] Importantly, like similar datasets cited in the introduction, 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.

[12]Below this threshold, firms are not obliged to report the identity of their transaction partners to the tax authorities. We treat unobserved transactions below this threshold as negligible.

Table 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

Notes. 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.[13] 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 1 summarizes some 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.[14]

[13] This interval corresponds with the second version of the dataset (the BCCR constantly updates the transaction and balance-sheet data with information from more recent years). The time window we utilize is appropriate for the empirical application discussed in section 5, which revolves around a policy intervention – the construction of a major highway – that was completed in 2011.

[14] This paper is based on the first version of the dataset, covering transactions from 2008 to 2015. The stylized facts that this paper describes hold across versions of the dataset.

Next, we document some stylized facts, each supported by a figure, that are useful to motivate certain features of the proposed empirical framework.

Fact 1. *Buyer (seller) firms occasionally transact with more than one seller (buyer) firm from the same narrowly defined industry in the same year.*

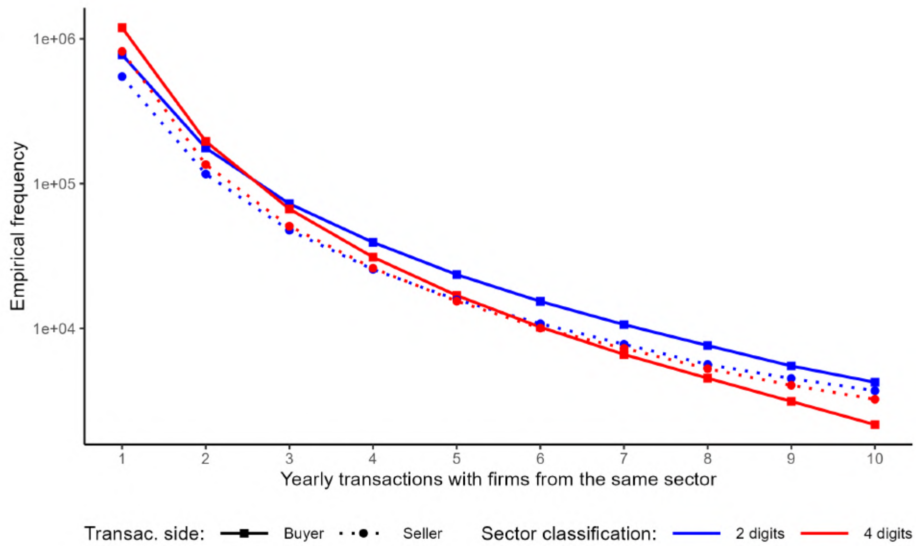


Figure 1: Frequency of transactions with multiple partners from the same sector

Notes. 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.

Figure 1 shows that typically firms buy from, or sell to, only another firm in the same industry (however the latter is defined), and yet, sometimes firms have multiple partners from the same sector. While 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 individual buyers (sellers) connect *at random* with sellers (buyers) from a given restricted subset of sectors. This "balls and bins" problem can, under careful parametrization, reproduce the patterns of Figure 1. However, we find overwhelming evidence against the hypothesis that firms link up with partners from random sectors: see the discussion in Appendix A.[15] This

[15]Furthermore, random association is unlikely to reproduce the observations for $x = 2$ and $x = 3$ in Figure 1, each of which displays similar values across sector classifications (and transaction sides).

suggests that buyers' choice of suppliers as well as sellers' ability to find customers are idiosyncratically determined by the technology (the empirical counterpart of which is the sector classification) of potential partners. In short, "sectors matter" in production network formation. We introduce this property explicitly in our empirical framework. Besides enabling our model to reproduce the empirical regularity from Figure 1, this property is key for econometric identification.

Fact 2. *Across combinations of buyer and seller sectors, the empirical distribution of the share of transactions over the revenue of buyers appears at least bimodal.*

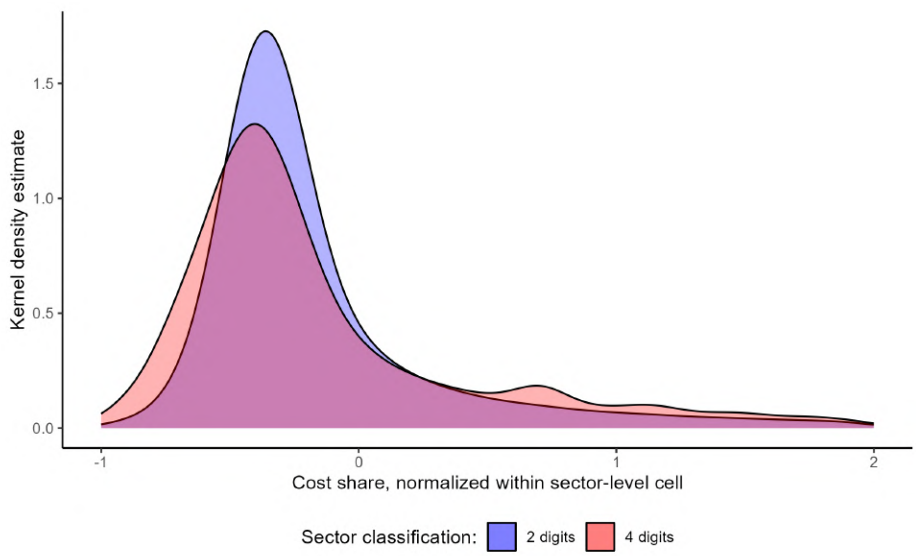


Figure 2: Distribution of normalized shares of transactions over buyer revenue

Notes. 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*: Revec.

Figure 2 displays two 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 for each combination of buyer sector, seller sector, and year.[16] The standardization allows one to appreciate

[16] More formally, let $Share_{ijt} = Transaction_{ijt}/Revenue_{it}$ 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 $Normalized\ share_{ijt} = (Share_{ijt} - avg\,(Share_{ijt}))\,/sd\,(Share_{ijt})$, where $avg\,(\cdot)$ and $sd\,(\cdot)$ 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.

via a single figure some features of the empirical distributions of raw shares shared across multiple combinations of industries. The only difference between the two kernel densities lies in the granularity of the sectors involved in the standardization. While both estimates display heavy right tails, the one based on four digits is clearly bi- if not multimodal. This suggests that at least in some industries, buyers occasionally make larger-than-average or multiple purchases. Along with a theoretical motivation, this finding informs our approach for inferring a discrete number of "input choices" from the observed value of transactions, aimed at addressing a censoring problem typical of firm-to-firm transaction data: we can only observe the total value of transactions between two firms, and not separate orders for possibly distinct intermediate inputs.

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

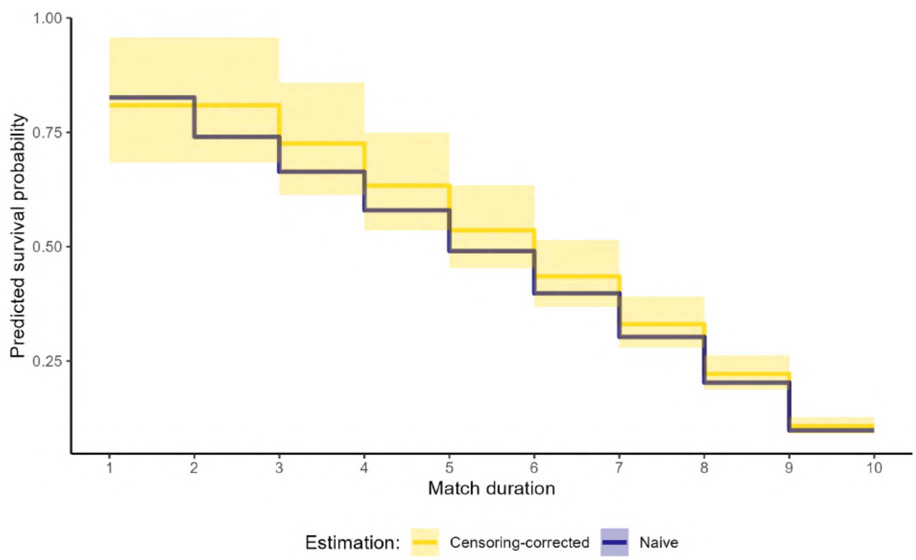


Figure 3: Estimates of survival probabilities for firm-to-firm transactions

Notes. 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 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 on the right of an interval. The Turnbull estimator does not yield an estimate for survival at year 2, hence the initial step extending up to 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 observed), and we estimate the survival probabilities of such spells. As Figure 3 shows, 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.[17] However, such simple counts do not account for censoring (matches observed in 2008 may have started earlier, and we cannot observe the termination of spells still occurring in 2017). Censoring-corrected estimates based on the Turnbull (1974) estimator are unsurprisingly even larger.[18] This evidence calls for an econometric model that accommodates sources of path-dependence if estimated on longitudinal firm-to-firm transaction data.

These stylized facts are largely novel in the literature on production networks.[19] As shown in the cited descriptive paper by Alfaro-Ureña et al. (2018), stylized facts about production networks reported in other contributions (e.g. Bernard et al., 2019, 2022) are reproduced on the Costa Rican data; this suggests that Facts 1, 2 and 3 are likely, while tentatively, reproducible on transaction data from other countries too. Verifying this is outside the scope of this paper; however, we find the evidence of this section as encouraging for the idea that some key features of the framework outlined in the next section bode well with the empirical facts.

3 The empirical framework

This section develops our econometric framework. For the sake of exposition, we first outline the basic version of the model cast in a cross-sectional environment. Later, we illustrate a more extended version that is adapted for longitudinal data and that introduces additional elements, such as foreign sellers and "make or buy" decisions. Lastly, we provide a miscellaneous discussion about additional extensions and issues of implementation in practice. These three parts are organized as distinct subsections. The proofs of our key analytical results or "propositions" are developed in Appendix B, while Appendix C provides further discussion of selected features of the model.

[17] We calculate confidence intervals associated with the naive estimates; however, due to the sheer size of the data, they are too small to be appreciable in the figure.

[18] We also conducted this exercise under 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 are virtually unchanged.

[19] The high persistence of transactions is a well-known fact but, to the best of our knowledge, there are no extant estimates of match survival probability that correct for censoring.

3.1 Cross-sectional framework

An economy is populated by a discrete number N of firms, indexed as $i, j = 1, \ldots, N$; we denote the set of all firms by $\mathcal{I}$. Each firm belongs to a *unique* sector (industry), which identifies the technological type of output that it produces. Two firms belonging to the same sector may still produce output that is perceived as differentiated by final consumers, by other firms, or both. There are in total S sectors in the economy, which are indexed as $s, z = 1, \ldots, S$. We let $S \ll N$: typically, many firms share the same sector. With some convenient abuse of notation, we denote by $s(i)$ the function that associates a firm i to its sector, and we write the set of all firms that belong to some sector s by $\mathcal{S}_s = \{i \in \mathcal{I} : s(i) = s\}$. The collection $\{\mathcal{S}_s\}_{s=1}^{S}$ is a partition of $\mathcal{I}$.

To produce its own output Y_i, a firm i must perform an idiosyncratic set of "tasks," which are denoted by $\mathcal{K}_i$, that represent the organization and/or assembly of multiple goods and services. Hence, a task k requires firm i to use of some intermediate inputs X_{ik} (interpreted as quantity of the intermediate good or service), for any $k \in \mathcal{K}_i$. We introduce two assumptions that regulate the association between tasks, sectors and suppliers for all firms in the economy.

Assumption 1. *To accomplish a task $k \in \mathcal{K}_i$, a firm $i \in \mathcal{I}$ must use inputs $X_{ijk} = X_{ik}$ supplied from only* one *firm $j \in \mathcal{I}$, $j \neq i$. We denote such a firm-supplier by $j(k)$.*

This assumption implies no loss of generality, as tasks can always be re-defined so as to meet it. We uninspiringly call function $j(k)$, which associates an individual supplier to any given task k, the "supplier function," and $\mathcal{J}_i \equiv \bigcup_{k \in \mathcal{K}_i} j(k)$ the firm's "supplier set." All firms choose their supplier sets endogenously, under the restrictions and assumptions that are discussed next. Importantly, a firm $j \in \mathcal{I}$ can appear multiple times in another firm's supplier set $\mathcal{J}_i$, as it may provide inputs for distinct tasks. A collection of *transactions* is denoted by $\mathcal{G} \equiv \bigcup_{i \in \mathcal{I}} \bigcup_{k \in \mathcal{K}_i} (i, j(k))$, and the ordered pair $(\mathcal{I}, \mathcal{G})$ fully characterizes the topological structure of the economy's production network, which is directed and (in the case of multiple-task suppliers) weighted.

Assumption 2. *For any task $k \in \bigcup_{i \in \mathcal{I}} \mathcal{K}_i$, inputs can only be provided by suppliers of a specific sector, denoted as $z(k)$. Hence, $j(k)$ must always satisfy $(s \circ j)(k) = z(k)$.*

This assumption imposes a fundamental technological restriction: only firms from the appropriate industry can provide inputs to fulfill a certain task. We view this as a realistic hypothesis: one would typically not source car tires from fruit companies;

similarly, tire producers are not the appropriate providers of pineapples. Importantly, a sector can potentially fulfill multiple tasks for the same producer i. Hence, for any two tasks $k, k' \in \mathcal{K}_i$ it is either $\mathcal{S}_{z(k)} = \mathcal{S}_{z(k')}$ or $\mathcal{S}_{z(k)} \cap \mathcal{S}_{z(k')} = \varnothing$. This remark helps clarify why our framework adopts a distinction between tasks and sectors, rather than a more parsimonious approach. Thanks to the distinction, the framework can in fact replicate the evidence from Figure 1, where buyers are observed to source more than once from the same sector in the same year.

All firms are characterized by a general Constant Elasticity of Substitution (CES) production function with constant returns to scale and elasticity of substitution $\sigma > 1$, which combines labor L_i with the intermediate inputs X_{ijk}. For a given $\mathcal{J}_i$, it reads:

$$Y_i = A_i \left[(\alpha_{0i} L_i)^{\frac{\sigma-1}{\sigma}} + \sum_{k \in \mathcal{K}_i} \left(\alpha_k X_{ij(k)k} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

where A_i is firm i's exogenous total factor productivity, while α_{0i} and α_k (for $k \in \mathcal{K}_i$) are firm-specific parameters that weigh the saliency of, respectively, the labor and the task-specific inputs.[20] From (1) one derives the firm's marginal cost C_i as:

$$C_i = \frac{1}{A_i} \left[\left(\frac{\alpha_{0i}}{W} \right)^{\sigma-1} + \sum_{k \in \mathcal{K}_i} \left(\frac{\alpha_k}{P_{ij(k)k}} \right)^{\sigma-1} \right]^{\frac{1}{1-\sigma}}, \tag{2}$$

where W is the exogenous real wage while P_{ijk} is the *effective* unit real price charged by an eligible supplier $j \in \mathcal{S}_{z(k)}$ for the inputs it provides to fulfill some task $k \in \mathcal{K}_i$. We allow suppliers to price-discriminate and offer buyer-specific prices.

The endogenous choice of firms' supplier sets $\mathcal{J}_i$ is dictated by cost minimization: per (2), firms select for any task k the eligible supplier j that offers the lowest effective price P_{ijk}. The latter can be decomposed, for all $j \in \mathcal{S}_{z(k)}$, as:

$$P_{ijk} = \frac{\mu_{ijk} C_j \tau_{ij}}{\exp\left(\varepsilon_{ijk}\right)}, \tag{3}$$

where μ_{ijk} is a task-specific mark-up on j's marginal cost C_j; ε_{ijk} is a match-specific random shock that measures how convenient it is, in relative terms, for firm i to source *specifically* from firm j for task k; while $\tau_{ij} \geq 1$ is a measure of *total matching frictions*

[20] The saliency parameters as well as A_i play only a small role in our framework. The i subscript in α_{0i} indicates that we allow firms to be heterogeneous in their intensity of labor utilization.

akin to iceberg costs in models of international trade and economic geography, which effectively inflate the cost borne by a firm for its inputs, because only a fraction τ_{ij}^{-1} of the inputs that are paid for can be actually used by firm i.

The τ_{ij} measure is central in our framework, as it encodes those dyadic observable variables whose effect on linkage formation we aim to estimate. More specifically:

$$\tau_{ij} = \tau_{ij}\left(\boldsymbol{z}_{ij}\right) = \exp\left(\beta_0 - \boldsymbol{z}'_{ij}\boldsymbol{\beta}\right) \tag{4}$$

where $\boldsymbol{z}_{ij} = (Z_{ij1}, \ldots, Z_{ijQ})$ is a vector of some Q dyadic observable characteristics, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_Q)$ is a vector of associated parameters, whereas $\beta_0 \geq 0$ is a constant, which vanishes in estimation, that can be arbitrarily set to ensure that $\beta_0 - \boldsymbol{z}'_{ij}\boldsymbol{\beta} \geq 0$ (and hence, $\tau_{ij} \geq 1$) always holds. The most straightforward example of a dyadic Z_{ij} variable is a direct measure of travel costs, such as spatial distance or travel times. Other examples may be setting- or data-specific: to make a few of them, a Z_{ij} variable may encode some exogenous policy treatment (say, a governmental policy that favors linkages between firms with selected characteristics) or more micro-level information about firms (like the existence of personal connections between their workers, possibly due to those workers' prior careers, provided that these can be observed). Our model clearly accommodates also variables that only vary at the level of the seller j.

A major motivation for this framework is that prices are typically not observed separately from quantities in real-world transactions, but even if they were, we would never be able to observe the prices for *unrealized* transactions (which did not occur). Both considerations make the estimation of $\boldsymbol{\beta}$ via some variation of (3) impractical, if not altogether unfeasible. We thus propose an indirect approach based on a multinomial logit of supplier choice derived from (3). To construct it, we need two additional assumptions: a distributional one about the shock ε_{ijk}, and one about the formation of mark-ups in equilibrium. A rigorous treatment of the latter would require closing the model and fully characterizing its equilibria, but to the detriment of exposition. Thus, here we only outline a limited assumption about supplier pricing behavior, and we defer a more extended discussion of the model to Appendix C.

Assumption 3. *Potential suppliers compete* à la *Bertrand and enact limit pricing.*

This assumption states that in equilibrium, all suppliers offer marginal cost pricing for a given task, except the one that can provide the *ex ante* (before mark-up) most favorable conditions. This supplier is able to charge a mark-up that makes its effective price (almost) equal to the second best's; we leave a formal characterization to the appendix. Thus, taking the marginal costs of all eligible suppliers as given, the *optimal* choice for buyer's i task k, denoted as $j^*(i,k)$, follows from (3) and (4) as:

$$j^*(i,k) = \arg\max_{j \in \mathcal{S}_{z(k)}} \gamma_j + \boldsymbol{z}_{ij}' \boldsymbol{\beta} + \varepsilon_{ijk} \tag{5}$$

where $\gamma_j \equiv -\log C_j$. In what follows, we name the collection $\{\gamma_j\}_{j \in \mathcal{I}}$ the *seller fixed effects*. While intuitive, this denomination can be misleading: these quantities, which are inversely related to marginal costs, are not really "fixed," but rather endogenous and recursively dependent, as an inspection of (2), (3) and (5) would show. Intuitively, sourcing tasks from cheaper upstream suppliers lowers a firm's marginal cost, and hence also that of its downstream buyers. However, the expression "fixed effects" is useful to convey the idea that these quantities shift the probability to select a given supplier across tasks, and that they pose a challenge for our econometric approach.

Assumption 4. *The ε_{ijk} shocks are known by firms under perfect information; they are exogenous and mutually independent across buyers, suppliers and tasks; and they all follow a standard Gumbel (type I extreme value) distribution.*

This is a typical assumption that conveniently yields a multinomial logit expression for the probability of choosing a given supplier for a task. However, it is not the only hypothesis that would deliver it, exactly or approximately: alternatives are discussed in Appendix C. Following a standard derivation, *conditional on the fixed effects of the eligible suppliers $\ell \in \mathcal{S}_{z(k)}$*: γ_ℓ, and conditional also on the dyadic "friction" variables $\boldsymbol{z}_{i\ell}$, the probability that supplier j is chosen for task k obtains as:

$$\mathbb{P}\left(j^*(i,k) = j \,\middle|\, \{(\gamma_\ell, \boldsymbol{z}_{i\ell})\}_{\ell \in \mathcal{S}_{z(k)}}\right) = \frac{\exp\left(\gamma_j + \boldsymbol{z}_{ij}'\boldsymbol{\beta}\right)}{\sum_{\ell \in \mathcal{S}_{z(k)}} \exp\left(\gamma_\ell + \boldsymbol{z}_{i\ell}'\boldsymbol{\beta}\right)}. \tag{6}$$

This familiar expression is not very amenable to estimation in this setting due to the challenge posed by the fixed effects γ_j, which is threefold. First, estimating tens of thousands of fixed effects can be computationally very expensive, if at all feasible. Second, many buyers and sellers only appear in a handful of transactions, which can cause or amplify the incidental parameter problem. Third, the seller fixed effects are themselves endogenous, as they co-vary with $\mathcal{G}$. A possible solution is to condition the likelihood function on the full $\{\gamma_j\}_{j \in \mathcal{I}}$ sequence: this would require to either observe, compute, or estimate firms' marginal costs via external information, which might be data-demanding and subject to assumptions and statistical uncertainty; alternatively, one could establish assumptions that allow to solve for the fixed effects in closed form, which is the route pursued by Panigrahi (2023). On the other hand, omitting the seller fixed effects $\boldsymbol{\gamma}_j$ from estimation would yield inconsistent estimates if firms' marginal costs correlate with the friction variables $\boldsymbol{z}_{ij}$ of their potential matches; for example if, because of agglomeration economies, those potential suppliers that are closer to a buyer in geographical space tend to be also more productive, on average.

We solve the problem via a sufficient statistic approach *à la* Chamberlain (1980), which conditions in particular on the observed sequence of each seller's total number of customers (the "out-degree" of the production network). To illustrate our key result it is useful to establish some additional notation. Let $h_{ij} = |\{\ell \in \mathcal{J}_i : \ell = j\}|$ be the number of times that a firm j is selected as the supplier for *any* of firm i's tasks, for *some* (possibly non-optimal) $\mathcal{J}_i$ and for any pair $(i, j) \in \mathcal{I}^2$. Let g_{ij} be the value of h_{ij} that occurs in the data.[21] By construction, $h_{ij} = g_{ij} = 0$ if $i = j$. Let $\boldsymbol{H}$ and $\boldsymbol{G}$ be the *adjacency matrices* of size $N \times N$ that array the h_{ij} and g_{ij} values, respectively. Let $\boldsymbol{H}_s$ be the subset of columns of $\boldsymbol{H}$ such that $j \in \mathcal{S}_s$, and arrayed into an $N \times |\mathcal{S}_s|$ matrix; define $\boldsymbol{G}_s$ analogously. Let $\boldsymbol{d} = (D_1, \dots, D_N)$ be the *out-degree sequence* that collects the observed counts of all tasks sourced to each seller, with $D_j = \sum_{i \in \mathcal{I}} g_{ij}$ for $j = 1, \dots, N$. Lastly, let $\boldsymbol{Z} = \{\boldsymbol{z}_{1j}, \dots, \boldsymbol{z}_{Nj}\}_{j \in \mathcal{I}}$. Our main result follows.

Proposition 1. *By conditioning on the out-degree sequence* $\boldsymbol{d}$*, one can formulate the following likelihood function for* $\boldsymbol{\beta}$*:*

$$\mathscr{L}\left(\boldsymbol{\beta} \middle| \boldsymbol{d}, \boldsymbol{G}, \boldsymbol{Z}\right) = \prod_{s=1}^{S} \frac{\exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} g_{ij} \boldsymbol{z}_{ij}' \boldsymbol{\beta}\right)}{\sum_{\boldsymbol{H}_s \in \mathcal{H}_s} \exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} h_{ij} \boldsymbol{z}_{ij}' \boldsymbol{\beta}\right)} \tag{7}$$

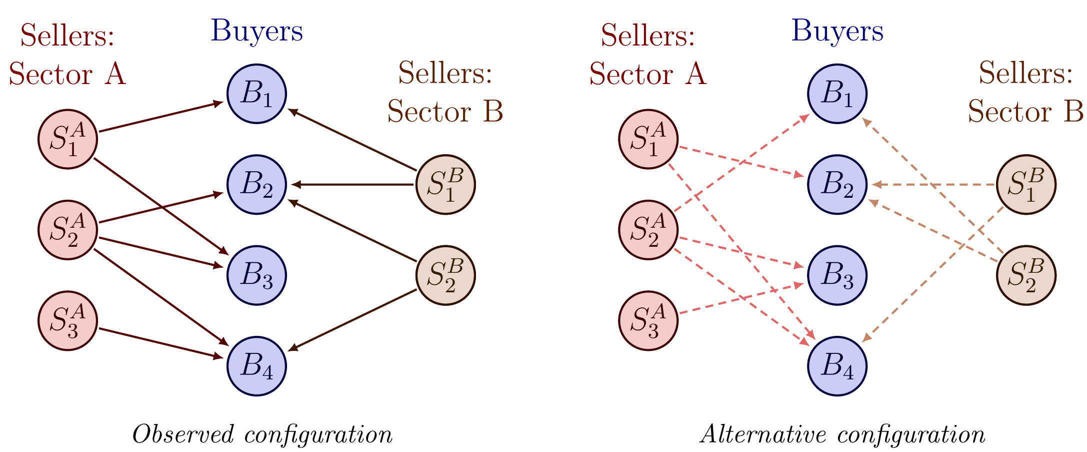
where:

$$\mathcal{H}_s \equiv \left\{ \boldsymbol{H}_s : \sum_{i \in \mathcal{I}} (h_{ij} - g_{ij}) = 0 \; \forall j \in \mathcal{S}_s, \sum_{j \in \mathcal{S}_s} (h_{ij} - g_{ij}) = 0 \; \forall i \in \mathcal{I} \right\} \tag{8}$$

is defined as the collection of all potential $\boldsymbol{H}_s$ *matrices that share the same "margins" (sums along rows and sums along columns) as* $\boldsymbol{G}_s$*, for* $s = 1, \dots, S$.

[21] According to our model, this is the value that emerges from buyers' optimal choices. Formally, $g_{ij} = |\{\ell \in \mathcal{J}_i^* : \ell = j\}|$ for $\mathcal{J}_i^* \equiv \bigcup_{k \in \mathcal{K}_i} j^*(i, k)$, and for any $(i, j) \in \mathcal{I}^2$.

The likelihood function (7) features no seller fixed effects, similarly to applications of the sufficient statistic approach to (binary or multinomial) logit models for panel data. With respect to the latter, our approach presents two distinctive features. First, rather than conditioning on "success counts" (here, the number of tasks-inputs that a firm supplies) that occur over time, our approach exploits the spatial dimension given by the network: intuitively, the number of buyers a firm displays at any point in time conveys information about how intrinsically convenient that particular firm-seller is for all its potential buyers. Second, the sets of alternatives in the denominator, which is defined in (8), also constrains the sectoral in-degree of buyers (the number of tasks-inputs that a firm sources from a given sector). This follows from Assumption 2 and bears two implications: as argued, it disciplines the ability of the model to reproduce empirical facts, and it facilitates the implementation of our approach for reasons to be discussed shortly. By segregating task choices across sectors, we construct a likelihood function based on the probability that some *collection of subnetworks*, each restricted to sellers from a particular sector, is the one being actually observed. This probability is defined over a sample space restricted to collections that share both in-degree and out-degree sequences across all sectors. Graph 2 below illustrates this idea, extending the representation from Graph 1 to allow for two different seller sectors ("A" and "B"). This way, the graph captures a key dimension of the likelihood function (7).



Graph 2: Two network configurations with two distinct seller sectors

Notes. This graph represents two alternative network configurations (similarly as in Figure 1, although displayed on two distinct panels: left and right), with identical out-degree and in-degree for two distinct seller sectors ("A" and "B"). Using our notation, they correspond to two distinct $(\boldsymbol{H}_A, \boldsymbol{H}_B)$ pairs. The edges need not be configuration-specific: a typical linkage usually appears in multiple configurations.

To obtain a maximum likelihood estimator from (7), one should in principle specify all sets $\mathcal{H}_s$ for $s = 1, \ldots, S$. If feasible at all, this is a very computationally demanding task for sectors featuring even a moderate number of sellers, and their corresponding buyers. Thus, we propose an approach based on the random sampling of the elements of $\mathcal{H}_s$ across sectors, which builds on McFadden (1978) and D'Haultfœuille and Iaria (2016), and that allows for a consistent, feasible estimator at the cost of an efficiency loss. This is expressed as our next result.

Proposition 2. *We define the Random Subnetwork Logit (RSL) estimator of* $\boldsymbol{\beta}$ *as:*

$$\widehat{\boldsymbol{\beta}}_{RSL} = \underset{\boldsymbol{\beta} \in \mathbb{R}^Q}{\arg\max} \prod_{s=1}^{S} \frac{\exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} g_{ij} \boldsymbol{z}_{ij}' \boldsymbol{\beta}\right)}{\sum_{\boldsymbol{H}_s \in \mathcal{H}_s^*} \exp\left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} h_{ij} \boldsymbol{z}_{ij}' \boldsymbol{\beta}\right)} \tag{9}$$

where, for $s = 1, \ldots, S$, $\mathcal{H}_s^*$ *is a random collection of possibly repeated elements of* $\mathcal{H}_s$ *which satisfies the* uniform conditioning property, *i.e. for any two* $\boldsymbol{H}_s' \in \mathcal{H}_s^*$ *and* $\boldsymbol{H}_s'' \in \mathcal{H}_s^*$, *and for* $\boldsymbol{Z}_s = \{\boldsymbol{z}_{1j}, \ldots, \boldsymbol{z}_{Nj}\}_{j \in \mathcal{S}_s}$*:*

$$\mathbb{P}\left(\mathcal{H}_s^* \middle| \boldsymbol{G}_s = \boldsymbol{H}_s'; \boldsymbol{Z}_s\right) = \mathbb{P}\left(\mathcal{H}_s^* \middle| \boldsymbol{G}_s = \boldsymbol{H}_s''; \boldsymbol{Z}_s\right). \tag{10}$$

The RSL estimator thus defined is consistent and asymptotically normal, but is less efficient than an unfeasible estimator based on the maximization of (7)*.*

To implement the RSL estimator, it is necessary to draw the $\mathcal{H}_s^*$ collections via a sampling scheme that is consistent with (10). We propose the following approach: to construct each $\mathcal{H}_s^*$ collection as the union of $\boldsymbol{G}_s$ together with R_s more elements of $\mathcal{H}_s$ that are uniformly sampled *with replacement.*[22] In this setting, uniform sampling from $\mathcal{H}_s$ corresponds with the well-studied problem of uniformly sampling matrices with given row and column sums. The algorithm by Patefield (1981) is an established, fast solution to this problem, and it is easy to implement;[23] we adopt it in both our Monte Carlo simulations as well as in our empirical application. Assumption 2 is key for this approach, as it imposes constant row-sums across the elements of $\mathcal{H}_s$. Because

[22]This means making R_s identical random draws where each element of $\mathcal{H}_s$, including the observed subnetwork $\boldsymbol{G}_s$ itself, can occur with probability equal to the inverse of the dimension of the $\mathcal{H}_s$ set. It is straightforward to see that this complies with the uniform conditioning property. This scheme is independent of $\boldsymbol{Z}$; note that (10) also conditions on $\boldsymbol{Z}_s$ for notational consistency with the proof of the proposition (see Appendix B).

[23]See e.g. function `r2dtable()` from the `stats` package for the R computing language.

random sampling with replacement can possibly return multiple realizations of the same subnetwork $\boldsymbol{H}_s$ (including $\boldsymbol{G}_s$ itself), the statement of the proposition explicitly allows for repeated elements in $\mathcal{H}_s^*$. This is novel in the practice of multinomial logit estimation; thus, we show in the proof of the proposition that this does not affect the asymptotics of the RSL estimator.

A typical issue with sufficient statistic approaches to fixed effects in logit models is that the interpretation of the estimated parameters is hampered by the inability to observe the fixed effects and thus evaluate marginal effects. However, the seller fixed effects in our framework have a structural interpretation in terms of marginal costs. Hence, marginal effects can still be calculated via (6) for a hypothetical distribution of sectoral fixed costs which is empirically plausible, possibly informed via a separate exercise in cost function estimation. This does not invalidate one of the motivations for our approach: to enable consistent estimation of $\boldsymbol{\beta}$ even if firms' marginal costs are unknown. In fact, one can draw meaningful interpretations of the RSL estimates in terms of marginal effects even via uncertain estimates or hypothetical guesses of the marginal costs. We provide more concrete examples in our discussion of the empirical application.

3.2 Longitudinal framework

Typical firm-to-firm transaction data possess a time dimension, which is ideally fully exploited for the sake of estimating $\boldsymbol{\beta}$; this is even more important if some Z_{ij} variables of interest feature variation over time, as in our empirical application. In what follows, we extend the cross-sectional framework developed thus far to allow for a longitudinal dimension of the data. In the rest of this section, we index time as $t = 1, \dots, T$, where t is either a subscript or an argument of the variables and functions introduced earlier, and T is the number of time periods observed in the data. In case of mild ambiguity, we clarify the use of the index more explicitly. We allow the set of firms observed at time t in the economy, which we denote as $\mathcal{I}_t^d$, to change over time; in addition, we write $\mathcal{I}^d = \bigcup_{t=1}^T \mathcal{I}_t^d$. This implies variation in the composition of sectors over time, though the *number* of sectors is fixed at S. We write the set of local firms from sector s that are observed at time t as $\mathcal{S}_{st}^d$. The upperscript d, for reasons that are to be clarified immediately, stands throughout for "domestic."

Longitudinal data pose two main issues. First, firms are observed to change over time the set of sectors which they source from; this makes Assumptions 1 and 2 too restrictive in their present formulation. Since we adopt the view that the requirements of production encoded by the task sets $\mathcal{K}_i$ (note the lack of the t subscript) are fixed in the short run, buyers must be substituting domestic suppliers either with foreign ones, or with in-house execution of tasks. Import relationships with foreign suppliers are not stable either,[24] suggesting that the model should accommodate the choice of individual foreign supplier firms. Second, transactions are highly persistent in time, as per our discussion of Fact 3. To incorporate this feature into our framework, one can make the error terms ε_{ijkt} dependent over time, or introduce structural dependence in $\boldsymbol{z}_{ijt}$ (for example, via a dummy variable that identifies past transactions) or both. We mainly pursue the former approach; we also discuss what additional complications or assumptions would be entailed by the latter.

In what follows, we outline a version of our framework that allows us to extend our RSL estimator to longitudinal transaction data. To this end, we replace Assumptions 1 and 4 with "augmented" versions of them, enumerated as 1a and 4a. Furthermore, we allow for *foreign* suppliers to provide inputs to domestic buyers. Specifically, we let there be $\mathcal{I}_t^f$ foreign suppliers in each time period t, and we write $\mathcal{I}^f = \bigcup_{t=1}^{T} \mathcal{I}_t^f$. The latter set is partitioned between the same S sectors of domestic firms; we denote the set of foreign suppliers that belong to sector s and that are available at time t by $\mathcal{S}_{st}^f$. The upperscript f, as opposed to d, means "foreign." We keep slightly abusing notation and use the j index to also denote individual foreign suppliers in $\mathcal{I}^f$. Hence, function $s(j)$ can also associate a foreign supplier $j \in \mathcal{I}^f$ to a particular sector s. We thus enable more options for domestic firms to fulfill their tasks.

Assumption 1a. *To fulfill a task $k \in \mathcal{K}_i$ at time t, a firm $i \in \mathcal{I}_t^d$ alternatively can:* a. *use inputs supplied from only* one *firm $j \in \mathcal{I}_t^d$, $j \neq i$;* b. *import them from a single foreign firm $j' \in \mathcal{I}_t^f$; or* c. *produce ("make") them in-house.*

This assumption introduces more realism into our model as it allows additional options for firms to secure their production inputs, consistently with the observation that buyers' supplying sectors can change over time. In what follows, we partition $\mathcal{K}_i$, for every firm $i \in \mathcal{I}_t^d$ and for every time period t, between the three subsets $\mathcal{K}_{it}^d$, $\mathcal{K}_{it}^f$, and $\mathcal{K}_{it}^m$, which, respectively, collect tasks sourced to domestic firm, tasks sourced to foreign firms, and tasks fulfilled internally ("make"). For any firm i, this partition can

[24] See Fitzgerald et al. (2023) for a recent study of the dynamics of international trade, although more focused on exporters than on importers.

change over time. Like in the static framework, restricting tasks in either $\mathcal{K}_{it}^{d}$ or $\mathcal{K}_{it}^{f}$ to one supplier bears no loss of generality. The supplier function $j(k)$ is interpreted here according to the origin of suppliers (domestic or foreign); we express it formally as $j : \bigcup_{t=1}^{T} \bigcup_{i \in \mathcal{I}_t^d} \mathcal{K}_{it}^{d} \rightarrow \mathcal{I}^d$ and $j : \bigcup_{t=1}^{T} \bigcup_{i \in \mathcal{I}_t^d} \mathcal{K}_{it}^{f} \rightarrow \mathcal{I}^f$. The supplier set at time t now collects both types of suppliers: $\mathcal{J}_{it} \equiv \bigcup_{n \in \{d,f\}} \bigcup_{k \in \mathcal{K}_{it}^{n}} j(k)$.[25] With this notation, we can keep Assumption 2 in its original wording; hence, Assumption 2 now also applies to foreign suppliers, with the same interpretation as for the domestic ones.

If a firm chooses to execute a task k internally instead of relying on other sellers, it establishes an internal division that employs some workers L_{0ikt} hired specifically to produce some substitute inputs X_{ikt}. These inputs are obtained according to a simple linear production function: $X_{ikt} = L_{ikt} M_{ikt}$, where M_{ikt} gives both the marginal and the average product of labor. Furthermore, we decompose M_{ikt} as a function of two elements, as follows:

$$M_{ikt} = \exp\left(m_{is(k)t} + \varepsilon_{0ikt}\right). \tag{11}$$

Here, $m_{is(k)t}$ is a deterministic factor that is specific to firm i, sector $s(k)$ and time t; while ε_{0ikt} is a task-specific random shock. One could think of $m_{is(k)t}$ as a function of measurable firm-level variables such as the technological proximity of the firm's own output $s(i)$ with that of sector $z(k)$; however, such a function would not be identified in our framework. Because the firm lacks the knowledge that is necessary to produce the X_{ikt} inputs, it must delegate the management of the division to an external agent, who demands a compensation equal to $(\mu_{0ikt} - 1) W_t$, $\mu_{0ikt} \geq 1$, per managed worker. Ultimately, the firm spends $P_{0ikt} \equiv \mu_{0ikt} W_t$ for each worker hired in the division. The agent-manager acts strategically and makes a "take-it-or-leave-it" request for μ_{0ikt} at the beginning of each period; if the division is not set up the agent's payoff is zero.

In the time-varying version of the CES production function (1), the total factor productivity A_{it} is understood as stochastic, whereas the saliency parameters α_{0i} and α_k, as well as the elasticity of substitution σ, are treated for simplicity's sake as fixed. In this setup, the marginal cost of firm i at time t, given the supplier set $\mathcal{J}_{it}$, is:

$$C_{it} = \frac{1}{A_{it}} \left[\left(\frac{\alpha_{0i}}{W_t} \right)^{\sigma - 1} + \sum_{n \in \{d,f\}} \sum_{k \in \mathcal{K}_{it}^{n}} \left(\frac{\alpha_k}{P_{ij(k)kt}} \right)^{\sigma - 1} + \sum_{k \in \mathcal{K}_{it}^{m}} \left(\frac{\alpha_k M_{ikt}}{P_{0ikt}} \right)^{\sigma - 1} \right]^{\frac{1}{1-\sigma}}, \tag{12}$$

[25] The *domestic* production network at time t is given by $\left(\mathcal{I}_t^d, \mathcal{G}_t\right)$, with $\mathcal{G}_t \equiv \bigcup_{i \in \mathcal{I}_t^d} \bigcup_{k \in \mathcal{K}_{it}^{d}} (i, j(k))$. The *bipartite* network of domestic buyers and foreign suppliers is obtained analogously.

where $P_{ijkt} = \mu_{ijkt} C_{jt} \tau_{ijt} \exp(-\varepsilon_{ijkt})$ is the longitudinal extension of (3). Note that here, the marginal cost C_{jt} of foreign suppliers (for $j \in \mathcal{I}^f$) is exogenous to the model. As in the static case, it is $\log \tau_{ijt} = \beta_0 - \boldsymbol{z}'_{ijt}\boldsymbol{\beta}$; without loss of generality, $\boldsymbol{z}_{ijt}$ is the same random vector of length Q whether j denotes a domestic or a foreign supplier. Foreign suppliers set markups μ_{ijkt} strategically as specified by Assumption 3, thereby enacting limit pricing.

Every domestic firm $i \in \mathcal{I}^d$ chooses an optimal partition of $\mathcal{K}_i$ between $\mathcal{K}^d_{it}$, $\mathcal{K}^f_{it}$ and $\mathcal{K}^m_{it}$ in every time period t. To characterize buyers' optimal choices, it is necessary to make additional assumptions about the stochastic components of (12). The following "augmented" version of Assumption 4 addresses the model's random shocks.

Assumption 4a. *For each firm i, task k and time t, the joint cumulative distribution of the random vector $\boldsymbol{\varepsilon}_{ikt} = \left(\varepsilon_{0ikt}, \{\varepsilon_{ijkt}\}_{j \in \mathcal{S}^d_{z(k)t}}, \{\varepsilon_{ijkt}\}_{j \in \mathcal{S}^f_{z(k)t}}\right)$ is given by:*

$$F_\varepsilon(\boldsymbol{\varepsilon}_{ikt}) = \exp\left[-\exp(-\varepsilon_{0ikt}) - \sum_{n \in \{d,f\}} \left(\sum_{j \in \mathcal{S}^n_{z(k)t}} \exp\left(-\frac{\varepsilon_{ijkt}}{\rho_n}\right)\right)^{\rho_n}\right]$$

i.e. a Generalized Extreme Value distribution in the sense of McFadden (1978), with scale parameters $(\rho_d, \rho_f) \in (0,1]^2$. In addition, the random vectors $\boldsymbol{\varepsilon}_{ikt}$ are exogenous and mutually independent across buyers and tasks.

This assumption is clearly conducive to a nested logit formulation of buyers' choice probabilities: this is appropriate in this setting, since importing inputs and especially "making" them are distinctive options for the buyers. A nested logit model introduces a correlation structure regulated by the two scale parameters ρ_d and ρ_f, such that any comparison between the choice probabilities for two options of different kind (say, a domestic supplier versus a foreign one) is not independent of irrelevant alternatives.[26] Note that the statement of the assumption implicitly allows errors that are dependent in time for the same buyer and task, which would lead to persistence in buyers' choices. In Appendix C we illustrate a simple model of random technological learning which, under minimal distributional assumptions, leads to a time-dependent structure of the shocks which complies with Assumption 4a. In short, firms learn in every period some "techniques" that they can use to fulfill their buyers' tasks, without forgetting them.

[26]To substantiate with a practical example, consider the problem of sourcing electric appliances. Foreign suppliers may adopt different country standards for power outlets and plugs, which leads to correlation between the ε_{ijkt} shocks associated with them.

Thus, suppliers always offer the techniques that deliver the best value ε_{ijkt}. Results from multivariate extreme value theory ensure that the joint distribution of the errors is asymptotically as given in the assumption, provided that the dependence between the original techniques shares the same nesting structure.[27]

We complete the description of the model by introducing a fifth assumption that is specific to the longitudinal setting.

Assumption 5. *For each firm i and time period t, the total factor productivity A_{it} and the set of dyadic characteristics $\{\boldsymbol{z}_{ijt}\}_{j\in\mathcal{I}^n}$ (for $n = d, f$) are* not *Granger-caused by the sequence of* all *past supplier sets expressed as $\left\{\mathcal{J}_{\ell(t-u)}\right\}_{\ell\in\mathcal{I}^d_{(t-u)}}$, for $u \in \mathbb{N}$.*

This assumption states that for every firm i, any variables of the A_{it} and $\boldsymbol{z}_{ijt}$ kind are statistically independent, conditional on any other information possibly available at time t (such as for example their own past realizations), of all past realizations of the networks, both domestic and international. This rules out any structural dependence in $\boldsymbol{z}_{ijt}$; other examples where Assumption 5 is violated are those where A_{it} or some Z_{ijt} variable (such as a measure of personal connections across employees of firms i and j) are endogenous to past transactions. Observe that this assumption implicitly allows time-dependence in either A_{it} or $\boldsymbol{z}_{ijt}$ (or both), which would make transactions more persistent over time. Assumption 5 simplifies the analysis considerably, as it removes the intertemporal dimension of the problem. Current networks cannot predict future profits: thus, buyers simply pick for each of their tasks in $\mathcal{K}_i$ the option that appears most cost-effective in the current period. It is possible to relax Assumption 5, but to the detriment of exposition: our framework would still be well-specified if future state variables were endogenous to current choices, but in a way that the most convenient suppliers of today are expected to remain so in the future. We elaborate on this idea in Appendix C, where we formally outline an assumption alternative to 5.

We can thus characterize buyers' optimal choices in the longitudinal framework. By Assumption 5, for every firm i and time t, a task $k \in \mathcal{K}_i$ is assigned to $\mathcal{K}^m_{it}$ if and only if the set:

$$\mathcal{B}_{ikt} \equiv \left\{ j \in \left(\mathcal{S}^d_{z(k)t} \cup \mathcal{S}^f_{z(k)t}\right) : \gamma_{jt} + \boldsymbol{z}'_{ijt}\boldsymbol{\beta} + \varepsilon_{ijkt} \geq \beta_0 - \log W_t + m_{is(k)t} + \varepsilon_{0ikt} \right\} \tag{13}$$

[27] The only restriction on the probability distribution that generates the original techniques is that it falls within the "basins of attraction," of the Gumbel or Fréchet cases; this includes a wide range of distributions. In the multivariate case, additional assumptions on the strength of cross-dependence may be necessary, see e.g. Genest and Rivest (1989).

is empty, i.e. there is no convenient eligible supplier. In particular, (13) follows since the agents who would manage the "make" divisions of a firm are strategic, and compete against actual firm-suppliers; in equilibrium, they undercut their request for μ_{0ikt} to the level that makes buyers indifferent between "make" and the best eligible suppliers. Conversely, task k is assigned to a supplier identified by the following function:

$$j^*(i,k,t) = \underset{j \in \mathcal{B}_{ikt}}{\arg\max}\, \boldsymbol{\gamma}_{jt} + \boldsymbol{z}'_{ijt}\boldsymbol{\beta} + \varepsilon_{ijkt}, \tag{14}$$

if $\mathcal{B}_{ikt}$ is nonempty. Hence, the "make" choice can also be expressed as $j^*(i,k,t) = \varnothing$. It follows from Assumption 4a that the all choice probabilities for domestic suppliers, foreign suppliers and the "make" option take the familiar nested logit structure; these are functions of the seller effects $\boldsymbol{\gamma}_{jt} = -\log C_{jt}$, similarly to the cross-sectional case.

We can thus reformulate Proposition 1 by restricting the attention to a likelihood function that conditions on the observations only from a single time period t. In what follows, we maintain the use of the $n = d, f$ upperscripts to represent information that is specific to dyads and/or (sub)networks that involve domestic and foreign suppliers, respectively. Thus, for example $\boldsymbol{d}_t^d$ is the out-degree sequence of domestic firms at time t, while $\boldsymbol{G}_t^f$ is the adjacency matrix that collects information about all tasks sourced from foreign firms at t. We also introduce the random vector $\boldsymbol{b}_{st}^n = (B_{1st}^n, \ldots, B_{N_t st}^n)$, where $B_{ist}^n = \sum_{j \in \mathcal{S}_{st}^n} g_{ijt}$ for $s = 1, \ldots, S$; g_{ijt} is the (i,j) entry of $\boldsymbol{G}_t^d$ or $\boldsymbol{G}_t^f$ depending on context, and N_t is the size of $\mathcal{I}_t^d$; $\boldsymbol{b}_{st}^n$ is thus the collection of *in-degree sequences* for each sector of the economy, that is the total number of tasks that each domestic firm sources at time t from firms of type n in sector s. Lastly, we gather all in-degree sequences of a given type across sectors via the vector $\boldsymbol{b}_t^n = (\boldsymbol{b}_{1t}^n, \ldots, \boldsymbol{b}_{St}^n)$. Our revised proposition follows.

Proposition 3. *Under the revised assumptions, and for $n = d, f$, by conditioning in particular on both the in-degree and out-degree sequences $\boldsymbol{b}_t^n$ and $\boldsymbol{d}_t^n$ observed at time t, one can formulate the following likelihood function for the combined parameter $\boldsymbol{\varphi}^n = \boldsymbol{\beta}/\rho_n$, for $t = 1, \ldots, T$:*

$$\mathscr{L}\left(\boldsymbol{\varphi}^n \middle| \boldsymbol{b}_t^n, \boldsymbol{d}_t^n, \boldsymbol{G}_t^n, \boldsymbol{Z}_t^n\right) = \prod_{s=1}^{S} \frac{\exp\left(\sum_{i \in \mathcal{I}_t^d} \sum_{j \in \mathcal{S}_{st}^n} g_{ijt} \boldsymbol{z}'_{ijt} \boldsymbol{\varphi}^n\right)}{\sum_{\boldsymbol{H}_{st} \in \mathcal{H}_{st}^n} \exp\left(\sum_{i \in \mathcal{I}_t^d} \sum_{j \in \mathcal{S}_{st}^n} h_{ijt} \boldsymbol{z}'_{ijt} \boldsymbol{\varphi}^n\right)}, \tag{15}$$

where $\mathcal{H}_{st}^n$ is defined as in (8) given $\mathcal{S}_{st}^n$ and the subnetwork $\boldsymbol{G}_{st}^n$ observed at time t.

There are three key differences with Proposition 1. First, our result is expressed in terms of the combined parameter sets $\boldsymbol{\varphi}^d$ and $\boldsymbol{\varphi}^f$; short of making more assumptions or gathering external information that enables the calculation of the seller fixed effects $\gamma_{jt} = -\log C_{jt}$, $\boldsymbol{\beta}$ is not identified under the current formulation of Assumption 4a. Second, each of the two parameter sets is associated with a specific likelihood function, which conditions on distinct observations. This arises from the structure of the nested logit model which here formulates probabilities to select a given supplier conditional on it being a domestic or a foreign one. Third, (15), unlike (7), explicitly conditions on the observed in-degree sequences because these are now endogenous and incorporate the outcome of the "make-or-buy" decision for each task; this follows from the nested logit structure of the choice probabilities, as we show in the proof. Note that the $\boldsymbol{H}_{st}$ matrices hold all in-degree and out-degree sequences fixed. This leads us to formulate an RSL estimator adapted to the longitudinal dimension of the data.

Proposition 4. *Under the revised assumptions, and for $n = d, f$, the RSL estimator of $\boldsymbol{\varphi}^n$ is a quasi-maximum likelihood estimator (QMLE) defined as:*

$$\widehat{\boldsymbol{\varphi}}^n_{RSL} = \underset{\boldsymbol{\varphi}\in\mathbb{R}^Q}{\arg\max} \prod_{t=1}^{T}\prod_{s=1}^{S} \frac{\exp\left(\sum_{i\in\mathcal{I}_t^d}\sum_{j\in\mathcal{S}_{st}^n} g_{ij}\boldsymbol{z}'_{ij}\boldsymbol{\varphi}\right)}{\sum_{\boldsymbol{H}_{st}\in\mathcal{H}_{st}^{n*}}\exp\left(\sum_{i\in\mathcal{I}_t^d}\sum_{j\in\mathcal{S}_{st}^n} h_{ij}\boldsymbol{z}'_{ij}\boldsymbol{\varphi}\right)} \tag{16}$$

where, for $s = 1, \ldots, S$ and for $t = 1, \ldots, T$, $\mathcal{H}_{st}^{n*}$ is a random collection of possibly repeated elements of $\mathcal{H}_{st}^n$ which satisfies the uniform conditioning property *as in (10). This RSL estimator is consistent and asymptotically normal; its asymptotic variance-covariance matrix depends on the assumed time-dependence structure of the model.*

The extended RSL estimator for the longitudinal framework is a QMLE, as stated by the Proposition, because we allow for arbitrary time-dependence in the ε_{ijkt} shocks. Even if the structure of this dependence was relatively simple, the aggregation of all dyad-level observations across subnetworks would make it difficult to formulate the "true" likelihood function. This makes a QMLE approach suitable to this problem; in the proof we show that this RSL estimator converges in probability to the true value of $\boldsymbol{\varphi}^n$ by applying established results by White (1994) that extend Gourieroux et al. (1984). To conduct statistical inference, we advocate clustered covariance estimation of the standard errors, where a cluster is a collection of subnetworks for a particular industry s observed at different points in time. In terms of practical implementation, the $\mathcal{H}_{st}^{n*}$ collections shall be drawn with replacement via Patefield's algorithm for each combination of industry s and year t.

The RSL estimator specified in Proposition 4 can be independently calculated on both the domestic production network ($n = d$) and the bipartite network based on domestic buyers and foreign exporters ($n = f$). The latter requires observing detailed cross-border transactions, which is typically obtained via customs data. Although this can be useful in empirical studies on international trade, our main intention is to show that our framework can accommodate substitution between domestic and foreign suppliers, and yet $\boldsymbol{\varphi}^d$ can be estimated only using firm-to-firm domestic transaction data. How to interpret such estimates? First, even if $\boldsymbol{\beta}$ is not identified separately of ρ^d (or ρ^f), the estimates of $\widehat{\boldsymbol{\varphi}}_{RSL}^n$ can still be used to perform odds-ratio comparisons between pairs of alternative counterfactual subnetworks, like in all logit models that condition on fixed effects. Second, our framework does not prevent practictioners to make more assumptions to facilitate the interpretation of the estimates. For example, if one believes that foreign exporters are a more selected group than domestic firms, it is reasonable to assume $\rho^d = 1$ while allowing for $\rho^f \leq 1$. Hence, (16) would return a consistent estimator of $\boldsymbol{\beta}$ (for $n = d$) which allows one to calculate marginal effects on the probability that a single supplier is chosen, conditional on a task being sourced domestically, under some assumptions on the distribution of sellers' marginal costs.

3.3 Implementation and extensions

In this final part of the section, we discuss some miscellaneous issues about practical implementation of the RSL estimator, as well as additional extensions of it.

Inferring the true (censored) task count. Typical firm-to-firm transaction data do not enumerate the number of distinct *purchases* ("tasks") of a buyer, as they merely provide the total yearly amount of bilateral transactions; as discussed in Section 2, this is also the case for our illustrative Costa Rican data. Hence, if according to our framework's lenses a firm sources more than one task from a distinct seller ($g_{ijt} > 1$), this information is effectively censored. This can lead to misspecifying the likelihood functions (7) or (15), possibly introducing biases in the estimates. To mitigate this problem in practical applications, in what follows we suggest a data pre-processing approach to *infer* the task count from actually observed transaction *values*, and which builds upon our previous discussion of Fact 2.

In our framework, the intensive margin of firm-to-firm transactions is determined by input demand functions. Given (1), the cost share for one task k over total input costs is:

$$\frac{P_{ijk}X_{ijk}}{C_iY_i} = \frac{(\alpha_k/P_{ijk})^{\sigma-1}}{(\alpha_{0i}/W)^{\sigma-1} + \sum_{k\prime\in\mathcal{K}_i}(\alpha_{k'}/P_{ijk'})^{\sigma-1}}. \tag{17}$$

Suppose that "similar" firms (e.g. firms in the same four-digits industry s), have equal technologies, hence equal $\mathcal{K}_i$ sets and saliency parameters α_k. Then any variation in (17) across transactions that involve any buyer $i \in \mathcal{S}_s$ and any seller $j \in \mathcal{S}_{s(k)}$ should depend on variation in the determinants of prices as per (3). It is not immediate to see how this can lead to the type of multimodal distribution discussed in Fact 2. Our favorite explanation for this is that the transaction values we observe are the sum of multiple task-specific purchases obtained from a finite mixture of distributions, each component of which corresponds to a given number of tasks in $\mathcal{K}_i$. This suggests using clustering algorithms designed for statistical mixture models to infer the number of distinct transactions; we pursue this approach in our empirical application.[28]

Structural dynamics. A researcher may want to study how supplier choice depends on a function of the past realizations of the network, say a dummy variable equal to $Z_{ijt} = \mathbb{1}\left[g_{ij(t-1)} > 0\right]$. The resulting model would feature "structural dynamics:" the frictions associated with bilateral transactions decrease after the first match, offering an additional explanation of transaction persistence (Fact 3).[29] However, structural dynamics violates Assumption 5: our model would be misspecified as forward-looking buyers must take into account the consequences that, in expectation, current choices have on future profits. Introducing intertemporal buyer choice would entail analytical complications that deserve separate research.

In Appendix C, we discuss alternatives to Assumptions 5 that allow for structural dynamics. To build intuition, suppose that choosing a supplier j for two consecutive

[28]A related issue is the *truncation* problem due to administrative data not reporting transactions below a certain threshold. Our framework can accommodate small transactions as a fourth decision allowed by Assumption 4a: "buying inputs from the retail market" as final consumers do, implying non-strategic markups. Thanks to our sufficient statistics approach, the RSL estimator would not be affected by the addition of such a fourth option. However, per (17), even some formal transactions may be truncated if both the buyer and the saliency parameter α_k associated with a task k are too small. This is an empirical issue that must be assessed by the researcher given the data at hand.

[29]This explanation can be related to the idea of fixed costs of transactions, which feature in a number of (static) theoretical models of endogenous production networks (e.g. Bernard et al., 2022; Dhyne et al., 2023) as the initial transaction in a sequence would imply a larger "set-up" cost.

periods positively affects its future productivity in such a way that increases in the seller fixed costs $\gamma_{j(t+1)}$ offset the reduced frictions. So long as the "effect" of lagged connections is small enough, in expectation the optimal supplier of today is also anticipated to be the best option always in the future; hence the choice probabilities are well-specified. Under such assumptions, the effect of past realizations of the network is identified. Intuitively, this follows because unlike in the standard multinomial logit with individual-alternative fixed effects and lagged dependent variables (Honoré and Kyriazidou, 2000), our likelihood transformation is based on the network, rather than the temporal dimension of the data. Motivated by this insight, in our empirical application we experiment with a "transaction lag" dummy variable.

Sellers' sectoral effects. Suppose that the match-specific shocks are actually given by $\varepsilon_{ijkt} = \eta_{s(i)jt} + \widetilde{\varepsilon}_{ijkt}$, where $\eta_{s(i)jt}$ is a constant effect of seller j that is specific for buyers from sector $s(i)$: a *seller's sectoral effect*; while $\widetilde{\varepsilon}_{ijkt}$ complies with Assumption 4a. This decomposition conveys the idea that firms are heterogeneous in their intrinsic capacity to supply different sectors, which is a realistic hypothesis. In such a setting, one could obtain an augmented version of the RSL estimator in (16) as:

$$\widehat{\boldsymbol{\varphi}}^n_{RSL} = \arg\max_{\boldsymbol{\varphi}\in\mathbb{R}^Q} \prod_{t=1}^{T}\prod_{s=1}^{S}\prod_{z=1}^{S} \frac{\exp\left(\sum_{i\in\mathcal{S}^d_{zt}}\sum_{j\in\mathcal{S}^n_{st}} g_{ij}\boldsymbol{z}'_{ij}\boldsymbol{\varphi}\right)}{\sum_{\boldsymbol{H}_{szt}\in\mathcal{H}^{n*}_{szt}}\exp\left(\sum_{i\in\mathcal{S}^d_{zt}}\sum_{j\in\mathcal{S}^n_{st}} h_{ij}\boldsymbol{z}'_{ij}\boldsymbol{\varphi}\right)}, \tag{18}$$

where $\mathcal{H}^{n*}_{szt}$ is obtained via uniform draws from:

$$\mathcal{H}^n_{szt} \equiv \left\{\boldsymbol{H}_{szt} : \sum_{i\in\mathcal{S}^d_{zt}}(h_{ijt}-g_{ijt}) = 0\ \forall j\in\mathcal{S}^n_{st}, \sum_{j\in\mathcal{S}^n_{st}}(h_{ijt}-g_{ijt}) = 0\ \forall i\in\mathcal{S}^d_{zt}\right\}, \tag{19}$$

while $\boldsymbol{H}_{szt}$ denotes a block of some adjacency matrix $\boldsymbol{H}^n_t$ which is restricted to sellers from sector s and buyers from sector z. Intuitively, $\eta_{s(i)jt}$ adds up to the seller fixed effects γ_{jt} resulting in seller-and-buyer-sector effects; our likelihood transformation would still remove such total effects if applied to every combination of buyer sector, seller sector, and year; all the derivations are analogous. The implementation of this extended RSL estimator does not pose special challenges, and the choice between (16) and (18) shall be taken by the researcher.[30]

[30] We plan to study differences in statistical efficiency between the two estimators in future work. We expect these to depend on the details of the clustering scheme.

Random parameters. Our discussion so far has assumed that $\boldsymbol{\varphi}^n$ is constant across buyers and sellers of different kinds, and over time. However, this implicit assumption is easily relaxed when one allows for random parameters. Because the RSL estimator is a multinomial logit, the situation where $\boldsymbol{\varphi}_{st}^n$ is a random parameter that varies by the sector of sellers s and by year t is well understood: a leading case is that where one assumes a more elaborate dependence structure than implied by Assumption 4a, one where the scale parameters ρ_{nst} vary along s and t, then $\boldsymbol{\varphi}_{st}^n = \boldsymbol{\beta}/\rho_{nst}$ and ρ_{nst} may be treated as a random draw from an appropriate distribution (e.g. the Beta). One can design even more elaborate random parameter schemes; suppose for example that in the cross-sectional framework, $\log \tau_{ij} = \beta_0 - \boldsymbol{z}_{ij}' \boldsymbol{\beta}_i$, where $\boldsymbol{\beta}_i$ is a buyer-specific parameter set. Thus, (7) becomes:

$$\mathscr{L}\left(\boldsymbol{\beta} | \boldsymbol{d}, \boldsymbol{G}, \boldsymbol{Z}\right) = \prod_{s=1}^{S} \int_{\mathbb{R}^N} \frac{\exp\left(\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{S}_s} g_{ij} \boldsymbol{z}_{ij}' \boldsymbol{\beta}_{ij}\right)}{\sum_{\boldsymbol{H}_s \in \mathcal{H}_s} \exp\left(\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{S}_s} h_{ij} \boldsymbol{z}_{ij}' \boldsymbol{\beta}_{ij}\right)} dF_{\beta}\left(\boldsymbol{\beta}_i\right) \qquad (20)$$

where $F_{\beta}\left(\boldsymbol{\beta}_i\right)$ is the joint distribution from which the random parameters are drawn. A simulation-based RSL estimator would be obtained accordingly; in practice though, it may be too computationally expensive because it would feature as many random parameters as there are buyers in the data. Thus, the appropriate random parameters scheme shall be evaluated in light of one's particular data and setting.

4 Monte Carlo

Our RSL estimator is a variation of a multinomial logit: hence, if the transformation of the likelihood function that motivates it is grounded on appropriate assumptions, it is expected to have asymptotic properties that are acceptable in practice. However, the mild complications involved in the construction of the $\mathcal{H}_s^*$ sets beg two questions: to what extent do the RSL asymptotics depend upon the details of implementation, and to what extent do they improve, if at all, upon alternative estimators that may be mildly inconsistent, but at the same time simpler and more intuitive to implement? While definite answers to these questions are not immediate to obtain, we conducted a number of Monte Carlo experiments based on a streamlined version of our framework. These experiments allow us to provide some tentative answers to our questions, which are favorable to the RSL estimator and to a simple implementation of it.

In each or our experiments we simulate 1,000 times a stylized economy observed over "time," setting all key parameters as constant. In each repetition we construct S "sectors" each populated by N_s firms; these firms must choose "suppliers" from pre-determined sectors. We simulate the task sets $\mathcal{K}_i$ for every firm $i \in \mathcal{I}$ via random draws from a Poisson distribution with parameter λ; in particular, every firm i receives S such draws, each corresponding with the number of tasks that must be sourced from a sector s, including $s(i)$. This determines the "technology" specific to a repetition, which is kept constant over T "years." In each year t, a different production network is obtained by aggregating simulated "choices" given by:

$$j^*(i,k,t) = \underset{j \in \mathcal{S}_{s(k)}}{\arg\max}\, \gamma_{jt} + \beta Z_{ijt} + \varepsilon_{ijkt},$$

where γ_{jt} is randomly drawn from one of two distributions as discussed below; ε_{ijkt} is a random draw from the standard Gumbel distribution that is independent across buyers, sellers, tasks as well as years; and, for any two firms $(i,j) \in \mathcal{I}^2$ and year t:

$$Z_{ijt} = \xi \phi_{ijt} + \zeta \left|\gamma_{it} - \gamma_{jt}\right|,$$

where ϕ_{ijt} is a random draw from the standard normal distribution, whereas ξ and ζ are two real parameters. If $\zeta \neq 0$ the simulation features endogeneity, since dyadic characteristics are a function of the distance between two firms' fixed effects. Hence, depending on the sign of β and ζ firms are relatively more attracted to suppliers with "fixed effects" γ_{jt} that are either closer to, or farther from, their own.[31]

This is a much sanitized version of our longitudinal framework, where the "make-or-buy" problem, foreign suppliers, as well as time-dependence in the random shocks are all removed. The latter choice in particular strikes a balance between simplicity of the simulation and its adherence to a key feature of the model: the time-invariant, exogenous task sets $\mathcal{K}_i$. Network formation is also simplified in the simulations, as the seller fixed effects γ_{jt} are exogenous random draws instead of being endogenous to the choices themselves. Since we aim to compare the RSL estimator against an alternative which is not designed to address the recursive structure of the fixed effects, we find

[31] If one thinks about Z_{ijt} for example as a measure of spatial distance, this particular specification would be consistent, if $\zeta > 0$, with the idea that Z_{ijt} correlates with productivity differentials at the level of firm dyads. Conversely, if Z_{ijt} represents cross-firm connections (e.g. between their workers or managers) we may expect $\zeta < 0$ due to the operation of knowledge spillovers.

it appropriate to remove such a feature; hence, our comparison must be interpreted as a conservative one, conducted under conditions that are expected to be relatively unfavorable to the RSL.[32] We treat the fixed effects γ_{jt} as unobservable; hence, both our competing estimators are computed using only the simulated values of the dyadic variables Z_{ijt} and the simulated choices $j^*(i,k,t)$ at each repetition.

We compare RSL estimates of β against multinomial logit estimates where in lieu of the seller fixed effects, we introduce κ dummy variables inspired by the "group fixed effects" (GFE) approach by Bonhomme et al. (2022). Specifically, in every repetition we assign each firm to one of κ groups following a k-means partition of the simulated out-degree distribution; the intuition being, as in our sufficient statistics approach, that higher γ_{jt} leads to higher D_{jt}. Bonhomme et al. (2022) show that if the support of unobserved heterogeneity is discrete and low-dimensional, their approach leads to a consistent estimator of the parameters of interest; we conjecture that this may apply to our setting, and perhaps deliver an estimator that is even more performant (in terms of mean squared error) than RSL. Furthermore, we conjecture that even if the support of γ_{jt} was continuous (which is arguably more realistic), such a GFE-based approach may be a viable option if it traded off a small bias for improved precision, in addition to being likely easier to implement in practice. To evaluate both conjectures we conduct two groups of experiments: one where $\exp(\gamma_{jt})$ has a continuous support, being drawn at every repetition for each firm-year from a Type I Pareto distribution with unit scale parameter and tail index equal to two, and one with discrete support, where draws are taken from a geometric distribution with parameter 0.5, and are then increased by one unit. All draws are independent over pseudo-time for each firm j.

Table 2 provides a summary of our Monte Carlo results for fifty combinations of estimators and experiments. We examine five estimators: two variations of the RSL, which differ by the number $R \in \{5, 20\}$ of random subnetworks $\boldsymbol{H}_{st}$ that, along with the "observed" $\boldsymbol{G}_{st}$ ones, enter the construction of each $\mathcal{H}^*_{st}$ set; and three variations of the GFE-augmented multinomial logit, for $\kappa \in \{3, 6, 12\}$. Each group of experiments (continuous versus discrete fixed effects) features five variations, symmetrical across groups. In our baseline we set $(S,T) = (10, 10)$ as well as $(\lambda, \xi, \zeta) = (0.10, 0.25, 0.25)$; in addition, sectors are *ex ante* symmetric, each being made up of fifty pseudo-firms.

[32] To introduce recursive fixed effects in the simulation, one must model firm production functions explicitly and introduce an algorithm which iteratively searches for one combination of firms' choices and fixed effects such that the former are optimal, and the latter are determined by the production functions evaluated at the given choices.

Variations of this baseline are obtained either by perturbing one of the parameters or by changing the size of sectors while keeping the size of the pseudo-economy constant. Across all experiments the parameter of interest is $\beta = 1$.

Table 2: Monte Carlo simulations

Estimator	Base	$\lambda = 0.2$	$\xi = 0.0$	$\zeta = 0.1$	HSS
			Continuous fixed effects		
RSL, $R = 5$	1.004	1.019	1.017	1.009	1.010
	(0.122)	(0.174)	(0.199)	(0.119)	(0.120)
RSL, $R = 20$	1.006	1.007	1.010	1.003	1.002
	(0.087)	(0.103)	(0.179)	(0.087)	(0.089)
GFE, $\kappa = 3$	1.827	1.828	4.374	1.360	1.810
	(0.180)	(0.165)	(0.551)	(0.086)	(0.173)
GFE, $\kappa = 6$	1.839	1.886	4.269	1.356	1.835
	(0.138)	(0.129)	(0.397)	(0.078)	(0.131)
GFE, $\kappa = 12$	1.742	1.893	3.921	1.320	1.749
	(0.124)	(0.106)	(0.320)	(0.076)	(0.119)
			Discrete fixed effects		
RSL, $R = 5$	1.008	1.014	1.003	1.006	1.003
	(0.126)	(0.193)	(0.163)	(0.110)	(0.122)
RSL, $R = 20$	1.006	1.001	0.996	1.002	1.004
	(0.086)	(0.103)	(0.140)	(0.083)	(0.085)
GFE, $\kappa = 3$	1.378	1.387	2.837	1.176	1.373
	(0.090)	(0.080)	(0.368)	(0.065)	(0.084)
GFE, $\kappa = 6$	1.332	1.397	2.563	1.153	1.342
	(0.073)	(0.054)	(0.257)	(0.060)	(0.068)
GFE, $\kappa = 12$	1.271	1.359	2.260	1.125	1.290
	(0.064)	(0.049)	(0.201)	(0.059)	(0.060)

Notes. This table reports the results of the Monte Carlo simulations described in the text. For each experiment, we report the results obtained across five estimators: the RSL estimator for $R \in \{5, 20\}$, and the GFE estimator based upon k-means partitions of the out-degree sequences for $\kappa \in \{3, 6, 12\}$ as described in the text. For each combination of experiment and estimator, the table reports the median of the estimated β coefficient across 1,000 repetitions and, in parentheses, the corresponding standard deviation. Baseline ("Base") experiments are all based on $(S, T) = (10, 10)$, $(\lambda, \xi, \zeta) = (0.10, 0.25, 0.25)$, and "sectors" of homogeneous size $N_s = 50$. Other experiments follow either from manipulating one parameter in (λ, ξ, ζ) as indicated in the table's header, or through heterogeneous sector sizes ("HSS") with four sectors of size $N_s = 30$, two of size $N_s = 40$, two of size $N_s = 50$, and two of size $N_s = 100$. All experiments are conducted with either continuous or discrete fixed effects, as described in the text; each group is reported in one of the two panels in the table: respectively, top and bottom.

The results show that across all experiments, the median RSL estimate is always virtually equal to one; the associated standard deviations are one order of magnitude smaller. The latter decrease as expected when we increase R: the number of sampled alternative subnetworks; yet the difference does not appear pronounced, except for the experiment variation where we set $\lambda = 0.2$, implying that the simulated networks are denser: in this case, the standard deviation falls by almost 50 per cent. By contrast, our adapted GFE logit appears markedly biased across all experiments, with median estimates up to about 4.4 times the true parameter. As expected, the bias is larger the higher the proportion of Z_{ijt}'s variance that is explained by the fixed effects (higher ξ and/or lower ζ) and under continuous, rather than discrete fixed effects. Network density, as set by λ, does not seem to play a role. Furthermore, increasing the number of dummies κ typically appears inconsequential, with one exception: the experiment most favorable to GFE, the one with milder endogeneity ($\zeta = 0.1$) and discrete fixed effects. In this case, $\kappa = 12$ yields a median bias of about 12.5 per cent the true value, but despite the precision gains, GFE is outcompeted by RSL even in that experiment. Sector size does not appear to impact the estimates, neither for RSL nor for GFE.

In summary, our original concerns about the RSL estimator appear reassuringly unfounded: the estimator is precise even for low values of R (which are obtained faster in actual implementations) and a fairly small "effective" sample size ST. In addition, the alternative estimator that we entertained, based on group fixed effects, does not seem to be viable, even under conditions that are more favorable to it. In future work, we plan to conduct additional simulations to examine the relative performance of the RSL estimator under different setups, for example by allowing for time dependence or by developing some of the extensions outlined in Section 3.3.

5 Empirical application

We employ the RSL estimator to study how transportation infrastructures and travel distance affect buyers' choice of suppliers and hence production network formation. In particular, we focus on a major Costa Rican highway, the *Ruta Nacional Primaria 27* (more simply, *Ruta 27*) that officially opened in 2010. Our expectation is that *Ruta 27* has facilitated the formation of buyer-seller relationships across pairs of locations that are directly or indirectly connected via the highway. Echoing the discussion by Bernard et al. (2019) in their study of the *Shinkansen* high-speed train in Japan, such a result would identify a particular channel by which infrastructures benefit economic activity: lowered input costs due to the ability to connect with better suppliers that are located farther away. The RSL estimator allows one to evaluate such a mechanism within a structural model of buyer choice based on cost minimization.

The *Ruta 27* is a strategic piece of infrastructure: it connects Costa Rica's Greater Metropolitan Area, encompassing the capital San José and home to about 60 per cent of the country's population, with Caldera, the main seaport along the Pacific coast. Originally designed in 1978, the *Ruta 27* was plagued by construction delays caused by financial and political issues. When it officially opened in early 2010, segments of the highway were still incomplete; the *Ruta 27* only became fully functional in 2011. At present, passage through the highway is typically optimal when planning a land route between San José and some location in one of the two Costa Rican "provinces" adjacent to the Pacific Ocean, and vice versa; this is due to the particular orography of the country, as mountain ranges hinder more direct connections between the center of Costa Rica and both its westernmost and southernmost tips (see Figure 4). This fact informs one of the two empirical specifications that we examine.

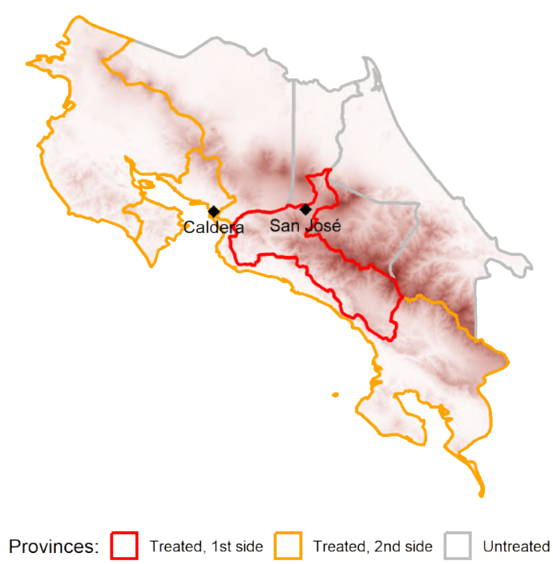


Figure 4: Costa Rican geography and the *Ruta 27*

Notes. This map of Costa Rica represents: (a) the altitude of the country, displayed via shaded reliefs; (b) the two endpoints of the *Ruta 27*, namely the capital city of San José and the seaport of Caldera; (c) the borders of the seven provinces (first-level administrative division) of Costa Rica. The latter are colored according to the role played by a province in defining our dyadic binary treatment: first side (San José, red); second side (Puntarenas or Guanacaste, orange); untreated (others, grey).

To evaluate the effect of *Ruta 27* on buyer choice, we estimate two specifications of the linear combination $\boldsymbol{z}'_{ijt}\boldsymbol{\beta}$. The first is based on a *binary dyadic treatment* that equals one if one of firms i and j is located in the province of San José (delimited by red borders in Figure 4), the other is located in either the Guanacaste or the Puntarenas province (orange borders), and t is 2011 or later; the dyadic treatement is otherwise zero. Dyads based on any other province pair are never treated. In this specification we nonparametrically control for the overall distance between two firms by introducing ten dummy variables in $\boldsymbol{z}_{ijt}$, one for each decile of the empirical distribution of dyadic *commuting distances*; these are calculated as the lengths of the shortest land routes connecting the centroids of the *cantons* (second-level administrative division[33]) where the two firms in a dyad are established.[34] One can interpret this specification, and the conditional choice probabilities it implies, under the familiar differences-in-differences framework. Our second specification is based on a simple *continuous regressor*, defined as the logarithm of the shortest travel time between any pair of cantons (calculated via a Google Maps API); however, for years antecedent to 2011, we do not allow the *Ruta 27* to be covered by the calculated paths. In this case, the effect of the *Ruta 27* is evaluated in terms of the marginal reduction in travel times. In both specifications we treat firm locations, which are fixed in our data, as predetermined.

We include additional dyadic controls in both our specifications. All our estimates include dummy variables that equal one if two firms are located in the same province, and are zero otherwise. In selected estimates we include three other dyadic controls, namely: the logarithmic "size ratio" (seller's employees divided by buyer's employees) capturing the role of seller size in linkage formation, inspired by Bernard et al. (2022); the logarithmic "trade relative exposure" (seller's imports *plus one* divided by buyer's exports *plus one*, adding ones accounts for observed zeroes), a measure we conjecture correlates with a seller's attractiveness, as access to foreign inputs could make sellers more valuable; lastly, in some specifications we introduce structural time-depedence via a dummy variable that equals one if the two firms were transacting in the previous year, and that is zero otherwise. The terms in the denominators of the two controls defined as log-ratios are normalization factors that help interpret the measures; they

[33]Costa Rica is partitioned into seven *provinces* (first-level administrative division) and eighty-four *cantons* (second-level). Each canton is entirely contained within the borders of a province.

[34]Calculations do not consider land routes passing through the *Ruta 27*. However, this is irrelevant: other routes between San José and Caldera have length comparable to that of the *Ruta 27*, but due to characteristics of the roads, they require about twice the travel time to be traversed.

do not affect neither the logit probabilities nor the estimates. For sure, the compliance of these measures with Assumption 5 is at best debatable, and the lagged transaction variable mechanically violates it; while our framework could still accommodate these variables under weaker assumptions, our primary objective is to assess the sensitivity of the parameters related to *Ruta 27* to the inclusion of additional controls.

Table 3: Empirical application: descriptive statistics

Dyadic variable	Actual transactions		Sampled alternatives	
Treatment (*"Ruta 27"*)	0.050	(0.218)	0.075	(0.263)
log(Commuting distance)	3.019	(1.484)	3.701	(1.356)
log(Travel time)	3.203	(1.769)	3.969	(1.443)
log(Size ratio)	0.622	(3.073)	-1.602	(2.570)
log(Trade relative exposure)	6.426	(9.322)	0.706	(8.160)
Same province	0.465	(0.499)	0.304	(0.460)
Transaction at $t-1$	0.578	(0.494)	0.007	(0.082)

Notes. This table reports the mean and, in parentheses, the standard deviation of all dyadic variables listed in the left column, separately for firm pairs that are observed to transact in a year ("actual transactions") and for alternative pairs obtained by randomly sampling, without replacement, sellers sharing the same four-digits sector with the observed one ("sampled alternatives"). Calculations are based upon all 2,192,003 transactions from 2009 to 2011. For each observed transaction, up to five alternatives are sampled (if available), yielding on average 5.993 dyads (0.118 standard deviation) associated with each observed transaction. The "Treatment," "Same province," and "Transaction at $t-1$" variables are coded as dichotomous dummies. The "commuting distance" variable is expressed as the number of kilometers between two cantons' centroids *plus one*, while the "travel time" is measured as the duration of the commute estimated by Google maps in terms of hours *plus one*. In both cases, adding one approximates the effective "door-to-door" journey. *Source*: Revec.

Table 3 displays descriptive statistics about the variables included in our empirical analysis, separately for dyads that are actually observed to transact and for a sample of "alternative dyads" such that in each of them, the seller shares its four-digits sector with at least one of the buyer's actual sellers. The table shows that sellers involved in actual transactions are on average larger, make a much more intensive use of imported inputs, and are typically closer in space, as it takes shorter distances and less time to reach them; in addition, they are more often located in the same province. However, a relatively lower share of observed transactions is exposed to *Ruta 27* according to our definition of binary treatment, since many firms located in the province of San José transact with provinces never exposed to the treatment (colored grey in Figure 4). Unsurprisingly in light of Fact 3, more than half of actual transactions were connected in the previous year, while only a very few of the alternative dyads were.

Before performing RSL estimation we take some preparatory steps. Following the discussion in Section 3.3, we first impute the number of "tasks" associated with each transaction. Specifically, we fit a gaussian mixture model with two components on the transaction cost shares normalized across sector pairs (pooled over years); we assign transactions to one of two groups according to their highest posterior mixture weight. As a result, about three quarters of all transactions are classified as "one task," the others as "two tasks." Second, we draw the random sets $\mathcal{H}^*_{st}$ via Patefield's algorithm for all subnetwork-year (s,t) combinations, setting $R = 5$ throughout. To avoid using combinations where the observed subnetworks $\boldsymbol{G}_{st}$ are likely to be drawn repeatedly (which would make the estimates harder to interpret) we drop subnetwork-years with less than twenty observed transactions (a conservative threshold): about 19 per cent of the total. Because our data feature 314 four-digit sectors and ten years, this results in 2,425 effective observations (2,192 if all observations from 2008 are also dropped). In Appendix D we provide further details about both preparatory steps; in particular, we show that the empirical distribution of the share of transacting dyads exposed to the *Ruta 27* treatment is analogous across retained and dropped subnetworks, which diminishes concerns about our data selection choices.

Table 4 displays our RSL estimates. Specifically, Panel A reports estimates based on the binary treatment specification; Panel B reports those based on a continuous "travel time" regressor. Column (1) of Panel A shows the parameter estimate for the binary treatment in case no control variables are added to the specification: it equals 0.02, indicating that *Ruta 27* makes transactions more likely across provinces, and it is statistically significant at the 10 per cent level. Columns (2) through (4) show results obtained by incrementally adding control variables: the point estimates for the *Ruta 27* treatment become larger in magnitude, though also noisier. An analogous pattern is observed in Panel B. The simple specification of column (5) yields a point estimate equal to –0.008 for the logarithm of travel time, negative as expected; it is statistically significant at the 1 per cent level. As shown in columns (6) through (8), adding more controls generally yields estimates that are larger in magnitude, but still statistically significant at the 1 per cent level. In both panels, the point estimates for the control variables indicate that larger suppliers are less likely to be selected, while those that import more of their inputs are more likely. Interestingly, the point estimate for the "lagged transaction" dummy is only sizable and statistically significant in column (8) from Panel B, and not in the corresponding Panel A specification. All estimates are weighted by the number of transactions observed in each subnetwork, magnifying the importance of larger sectors. This would yield estimates that are more representative of the whole economy in the likely case that the true parameters are heterogeneous.

Table 4: Empirical application: RSL estimates

Panel A: binary treatment specification

	(1)	(2)	(3)	(4)
Treatment (*"Ruta 27"*)	0.020*	0.025**	0.026	0.032
	(0.011)	(0.012)	(0.016)	(0.021)
log(Size ratio)		−0.004**	−0.002	−0.008***
		(0.002)	(0.002)	(0.003)
log(Trade relative exposure)			0.004***	0.004***
			(0.001)	(0.001)
Transaction at $t-1$				0.007
				(0.007)
Distance decile controls	YES	YES	YES	YES
Same province controls	YES	YES	YES	YES
Akaike information criterion	175.71	164.31	112.94	89.21
Number of subnetworks	2,425	2,425	2,425	2,192

Panel B: continuous regressor specification

	(5)	(6)	(7)	(8)
log(Travel time)	−0.008***	−0.007***	−0.010***	−0.018***
	(0.0002)	(0.0002)	(0.0004)	(0.004)
log(Size ratio)		−0.004***	0.001	−0.011**
		(0.0004)	(0.001)	(0.005)
log(Trade relative exposure)			0.001***	0.008***
			(0.0003)	(0.002)
Transaction at $t-1$				0.096***
				(0.011)
Same province controls	YES	YES	YES	YES
Akaike information criterion	512.02	536.40	455.64	34.84
Number of subnetworks	2,425	2,425	2,425	2,192

Notes. This table reports RSL estimates for the empirical application described in the text, for both specifications (each associated to one of the two panels in the table). Both panels display four sets of estimates, with varying sets of explanatory variables; at the bottom, they also report on the additional controls included, the computed Akaike information criteria, as well as the total number of subnetworks ST associated with each set of estimates. All $\mathcal{H}_{st}^{*}$ sets are constructed by sampling $R = 5$ alternatives (with replacement) using Patefield's algorithm. Estimates are weighted by subnetwork size, defined as the total number of both buyers and sellers associated with a subnetwork in a year. Estimate sets (4) and (8) obtain after removing the first year of the data, i.e. 2008. Asterisk sequences *, ** and *** denote significance at the 10, 5, and 1 per cent level, respectively. *Source*: Revec.

The interpretation of the estimates depends on context, and on the kind of thought experiment one is willing to entertain. Consider for example a single buyer from the San José area whose twenty potential sellers are evenly split between two locations: one that is served by *Ruta 27*, and one that is not. Suppose that the sellers only differ by their marginal costs C_{jt}, which do not depend on the actual choice of the buyer and are randomly drawn from a log-normal distribution with standard parameters.[35] It is easy to calculate via simulations that an estimate of 0.025 for our binary treatment yields an increase in the expected probability to pick a seller served by *Ruta 27* by 1.24 percentage points. Noting that *Ruta 27* decreases commuting times by about 80 minutes, an estimate of –0.01 for logarithmic travel time in the continuous regressor specification implies instead an increase of the above probability by 0.66 percentage points in case it originally took exactly three hours to reach either location. Let there now be one hundred buyers, all located in San José. The same estimates return an odds ratio between the probability that sixty of the selected suppliers are from the location served by *Ruta 27*, and the probability that the choices are evenly split across the two places, that is respectively equal to 1.284 (binary treatment case) and 1.061 (continuous regressor). These are all economically significant magnitudes.

In Appendix D we also report estimates from a "naive" multinomial logit model where the contributions of the likelihood function are individual choice probabilities similar to (6), where the alternatives in the denominator are randomly sampled from the four-digits sector of the actually observed supplier and where, importantly, seller fixed effects are utterly omitted. The point estimates are typically larger in magnitude than those from Table 4; any exercise at interpreting them akin to those discussed above is likely to return implausible effects and odds ratios. This suggests not only that, as already shown in Section 4, failing to account for the fixed effects can seriously bias one's estimates, but also that the variables employed in this empirical application, such as for example travel times, correlate with the unobserved productivity and cost-effectiveness of sellers. See the Appendix for further discussion.

6 Conclusion

Economic research has shown that knowledge of a production network's structure is key to assessing how industries and economies react to shocks and policy interventions.

[35] Observe that this exercise is invariant to multiplicative transformations of C_{jt}.

However, our knowledge of how production networks emerge from the aggregation of many firms' choices is still scant. This paper develops an econometric framework that we envision being useful for empirical researchers aiming at uncovering the empirical determinants of production network formation using firm-to-firm transaction data. The framework we propose is quite flexible: we examined only a few of all the possible extensions to the basic setup. It is, however, grounded on two premises that are likely restrictive in some real-world settings. First, the extensive margin of transactions does not imply any trade-off for sellers because these simply seek to maximize their buyer sets. This might be inaccurate, say, under capacity constraints. Second, buyers and sellers alike act under perfect information. This is likely unrealistic when firms make choices in sectors or "subnetworks" populated by many agents (buyers and sellers). We look forward to overcoming both limitations in future work.

References

Acemoglu, Daron and Pablo D. Azar (2020) "Endogenous Production Networks," *Econometrica*, 88 (1), 33–82.

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi (2012) "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80 (5), 1977–2016.

Adão, Rodrigo, Paul Carillo, Arnaud Costinot, Dave Donaldson, and Dina Pomeranz (2022) "Imports, Exports, and Earnings Inequality: Measures of Exposure and Estimates of Incidence," *Quarterly Journal of Economics*, 137 (3), 1553–1614.

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, 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.

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.

Almunia, Miguel, Jonas Hjort, Justine Knebelmann, and Lin Tian (2023) "Strategic or Confused Firms? Evidence from "Missing" Transactions in Uganda," *Review of Economics and Statistics*, Forthcoming.

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.

Arkolakis, Costas, Federico Huneeus, and Yuhei Miyauchi (2022) "Spatial Production Networks," Working paper.

Atkin, David and Amit K. Khandelwal (2020) "How Distortions Alter the Impacts of International Trade in Developing Countries," *Annual Review of Economics*, 12, 213–238.

Bartelme, Dominick and Yuriy Gorodnichenko (2015) "Linkages and Economic Development," NBER Working Paper no. 21251.

Bernard, Andrew B., Emmanuel Dhyne, Glenn Magerman, Kalina Manova, and Andreas Moxnes (2022) "The Origins of Firm Heterogeneity: A Production Network Approach," *Journal of Political Economy*, 130 (7), 1765–1804.

Bernard, Andrew B., Andreas Moxnes, and Yukiko U. Saito (2019) "Production Networks, Geography, and Firm Performance," *Journal of Political Economy*, 127 (2), 639–688.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2022) "Discretizing Unobserved Heterogeneity," *Econometrica*, 90 (2), 625–643.

Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito, and Alireza Tahbaz-Salehi (2021) "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake," *The Quarterly Journal of Economics*, 136 (2), 1255–1321.

Chamberlain, Gary (1980) "Analysis of Covariance with Qualitative Data," *The Review of Economic Studies*, 47 (1), 225–238.

Charbonneau, Karyne B. (2017) "Multiple fixed effects in binary response panel data models," *The Econometrics Journal*, 20 (3), S1–S13.

Crawford, Gregory S., Rachel Griffith, and Alessandro Iaria (2021) "A survey of preference estimation with unobserved choice set heterogeneity," *Journal of Econometrics*, 222 (1), 4–43.

de Paula, Áureo (2020) "Econometric Models of Network Formation," *Annual Review of Economics*, 12, 775–799.

de Paula, Áureo, Seth Richards-Shubik, and Elie Tamer (2018) "Identifying Preferences inNetworks with Bounded Degrees," *Econometrica*, 86 (1), 263–288.

Demir, Banu, Ana Cecília Filier, Daniel Y. Xu, and Kelly K. Yiang (2023) "O-Ring Production Networks," *Journal of Political Economy*, Forthcoming.

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, Xianglong Kong, Magne Mogstad, and Felix Tintelnot (2023) "Endogenous Production Networks with Fixed Costs," NBER Working Paper no. 30993.

Duranton, Gilles and Diego Puga (2004) "Micro-foundations of Urban Agglomeration Economies," in Duranton, Henderson, and Thisse eds. *Handbook of Regional and Urban Economics*, 4, 2063–2117: Amsterdam: North-Holland.

Faber, Benjamin (2014) "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *The Review of Economic Studies*, 81 (3), 1046–1070.

Fitzgerald, Doireann, Stefanie Haller, and Yaniv Yedid-Levi (2023) "How Exporters Grow," *Review of Economic Studies*, Forthcoming.

Genest, Christian and Louis-Paul Rivest (1989) "A characterization of Gumbel's family of extreme value distributions," *Statistics & Probability letters*, 8 (3), 207–211.

Gourieroux, Christian, Alain Monfort, and Alain Trognon (1984) "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52 (3), 681–700.

Graham, Bryan (2017) "An Econometric Model of Network Formation with Degree Heterogeneity," *Econometrica*, 85 (4), 1033–1063.

Gualdani, Cristina (2021) "An Econometric Model of Network Formation with an Application to Board Interlocks between Firms," *Journal of Econometrics*, 224 (2), 345–370.

Hirschmann, Albert O. (1958) *The Strategy of Economic Development*: Yale University Press.

Holl, Adelheid (2016) "Highways and productivity in manufacturing firms," *Journal of Urban Economics*, 93, 131–151.

Honoré, Bo E. and Ekaterini Kyriazidou (2000) "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68 (4), 839–874.

Huneeus, Federico (2020) "Production Network Dynamics andthe Propagation of Shocks," Working paper.

Lane, Nathaniel (2023) "Manufacturing Revolutions: Industrial Policy and Networks in South Korea," *Quarterly Journal of Economics*, Forthcoming.

Leung, Michael P. (2015) "Two-step estimation of network-formation models with incomplete information," *Journal of Econometrics*, 188 (1), 182–195.

Lim, Kevin (2018) "Endogenous Production Networks and the Business Cycle," Working paper.

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.

Mele, Angelo (2017) "A structural model of dense network formation.," *Econometrica*, 85 (3), 825–850.

Miyauchi, Yuhei (2023) "Matching and Agglomeration: Theory and Evidence from Japanese Firm-to-Firm Trade," Working paper.

Moretti, Enrico (2011) "Local Labor Markets," in Ashenfelter, O. C. and D. Card eds. *Handbook of Labor Economics*, 4, 1237–1313: Elsevier.

Oberfeld, Ezra (2018) "A Theory of Input–Output Architecture," *Econometrica*, 86 (2), 559–589.

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 × C Tables with Given Row and Column Totals," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 30 (1), 91–97.

Sheng, Shuyang (2020) "A structural econometric analysis of network formation games through subnetworks," *Econometrica*, 88 (5), 1829–1858.

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.

A The sectoral structure of the network: discussion

This appendix provides more evidence and discussion about the sectoral dimension of firm-to-firm matching in the production network; this is aimed at providing additional motivation to Assumption 2, which underpins our empirical framework. We start by presenting the input-output matrix of the Costa Rican economy in graphical form via Figure A.1. This shows that the Costa Rican production network is, like that of other countries (and as one would expect), highly sparse, suggesting that technology plays a fundamental role at shaping it.

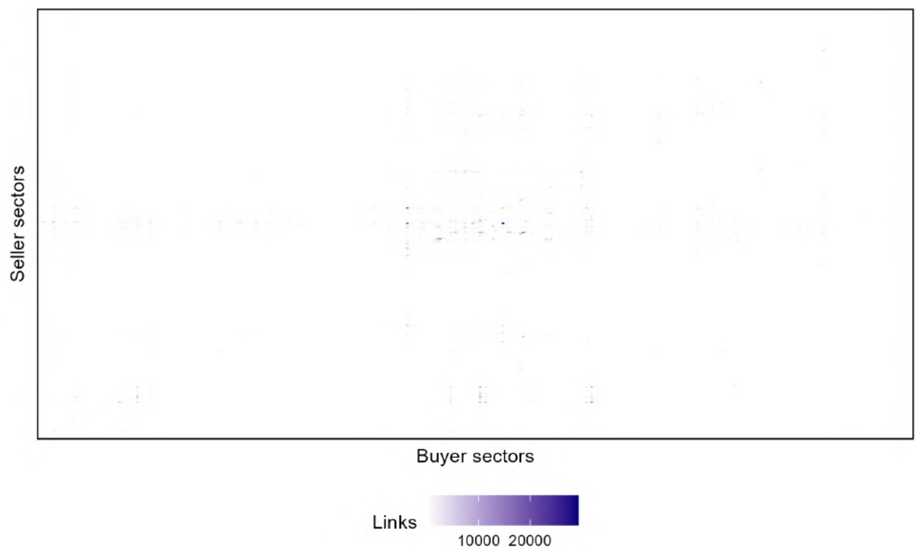


Figure A.1: The Costa Rican input-output matrix

Notes. 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 total number of unidirectional transactions observed in it over time (2008-2017). *Source*: Revec.

Sparsity of the input-output matrix is not, however, *per se* enough to motivate the hypothesis that the choices of *individual buyer firms* are technologically constrained, as stated by Assumption 2; instead, it could be the outcome of a random process. To elaborate, suppose that similar firms can *potentially* source their inputs from a subset of sectors, but not necessarily always from all of these, as in the stylized, constructed example in Table A.1, involving four buyer firms and four seller sectors that the former can source from. There, buyers display some variation in their choice of sectors, but a Fisher's exact test of indepedence (better suited to contingency tables with low count values than a Pearson's chi-squared test) cannot reject the null hypothesis that their "choices" are independent. Both sparsity of the input-output matrix and Figure 1 can be rationalized by a "balls and bins" model where buyers choose their suppliers from random sectors, provided that the average number of choices is sufficiently low.

Table A.1: 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

Notes. 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.

A scenario where buyers chose suppliers from random sectors would invalidate our empirical framework, as the latter would impose undue constraints on buyers' choices. Thus, we test this hypothesis statistically. Specifically, we conduct a number of Fisher exact tests, each restricted to all buyer firms from a given four-digits sector, pooling all years in the data. For every such selection of buyer-year observations, we construct a contingency table akin to A.1, though typically much larger. Figure A.2 displays the p-values from the resulting Fisher exact tests via a histogram. Clearly, the evidence overwhelmingly supports the favorable case where the choice of supplying sectors is idiosyncratic for buyers, a central assumption of our empirical framework.

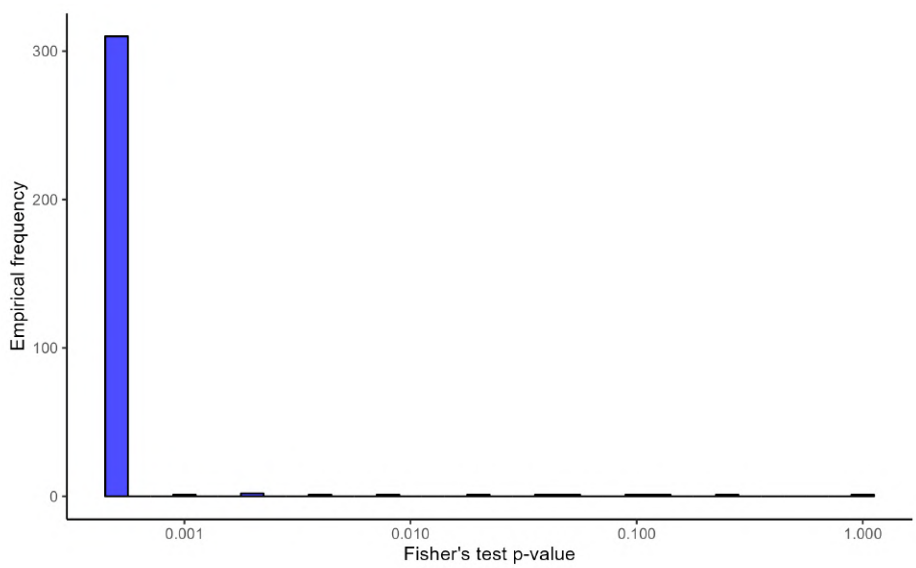


Figure A.2: Distribution of Fisher test p-values across four-digits sectors

Notes. This histogram displays the empirical distribution of the p-values from the Fisher exact tests discussed in the text, one for each four-digit sector. The horizontal axis is cast on a logarithmic scale. *Source*: Revec.

B Proposition proofs

This appendix provides analytical derivations of the propositions from Section 3.

Proof of Proposition 1

We start from characterizing the probability mass function of the production network $\boldsymbol{G}$ *conditional* on the sequence of seller fixed effects γ_j and the dyadic characteristics $\boldsymbol{z}_{ij}$. As we discuss in Appendix C, linkages are typically not pairwise independent and the model can feature multiple equilibria. However, under our assumptions, the only source of dependence lies in the recursive structure of sellers' fixed effects. Hence, by conditioning on these one can simplify the statistical characterization of the network. Let $\boldsymbol{\gamma} = \{\gamma_j\}_{j \in \mathcal{I}}$. We write the conditional mass function of the network as:

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

where the second equality follows from Assumption 4 and from conditioning on both $\boldsymbol{\gamma}$ and $\boldsymbol{Z}$, while the third one follows from Assumption 2 and the definition of g_{ij}.

From (B.1) we can derive the mass function of the out-degree sequence $\boldsymbol{d}$:

$$
\begin{aligned}
f_d\left(\boldsymbol{d}\middle|\, \boldsymbol{\gamma}, \boldsymbol{Z}\right) &= \sum_{\boldsymbol{H} \in \mathcal{H}} f_G\left(\boldsymbol{H}\middle|\, \boldsymbol{\gamma}, \boldsymbol{Z}\right) \\
&= \sum_{\boldsymbol{H} \in \mathcal{H}} \left[\prod_{i \in \mathcal{I}} \prod_{j \in \mathcal{I}} \left(\frac{\exp\left(\gamma_j + \boldsymbol{z}_{ij}' \boldsymbol{\beta}\right)}{\sum_{\ell \in \mathcal{S}_{s(j)}} \exp\left(\gamma_\ell + \boldsymbol{z}_{i\ell}' \boldsymbol{\beta}\right)}\right)^{h_{ij}}\right],
\end{aligned}
\tag{B.2}
$$

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

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

Every matrix $\boldsymbol{H} \in \mathcal{H}$ corresponds bijectively with a unique collection $\{\boldsymbol{H}_s\}_{s=1}^S$ where $\boldsymbol{H}_s \in \mathcal{H}_s$ for $s = 1, \dots, S$, with $\boldsymbol{H}$ resulting from (re-)joining the columns of all $\boldsymbol{H}_s$ 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:

$$\begin{aligned} f_{G|d}\left(\boldsymbol{G}|\,\boldsymbol{d},\boldsymbol{\gamma},\boldsymbol{Z}\right) &= \frac{f_{(G,d)}\left(\boldsymbol{G},\boldsymbol{d}|\,\boldsymbol{\gamma},\boldsymbol{Z}\right)}{f_d\left(\boldsymbol{d}|\,\boldsymbol{\gamma},\boldsymbol{Z}\right)} \\ &= \frac{f_G\left(\boldsymbol{G}|\,\boldsymbol{\gamma},\boldsymbol{Z}\right)}{f_d\left(\boldsymbol{d}|\,\boldsymbol{\gamma},\boldsymbol{Z}\right)} \\ &= \frac{\exp\left(\sum_{j\in\mathcal{I}} D_j\gamma_j\right)\exp\left(\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}} g_{ij}\boldsymbol{z}_{ij}'\boldsymbol{\beta}\right)}{\sum_{\boldsymbol{H}\in\mathcal{H}}\left[\exp\left(\sum_{j\in\mathcal{I}} D_j\gamma_j\right)\exp\left(\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}} h_{ij}\boldsymbol{z}_{ij}'\boldsymbol{\beta}\right)\right]} \\ &= \frac{\prod_{s=1}^{S}\exp\left(\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{S}_s} g_{ij}\boldsymbol{z}_{ij}'\boldsymbol{\beta}\right)}{\sum_{\boldsymbol{H}\in\mathcal{H}}\left[\prod_{s=1}^{S}\exp\left(\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{S}_s} h_{ij}\boldsymbol{z}_{ij}'\boldsymbol{\beta}\right)\right]} \\ &= \prod_{s=1}^{S}\left(\frac{\exp\left(\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{S}_s} g_{ij}\boldsymbol{z}_{ij}'\boldsymbol{\beta}\right)}{\sum_{\boldsymbol{H}_s\in\mathcal{H}_s}\exp\left(\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{S}_s} h_{ij}\boldsymbol{z}_{ij}'\boldsymbol{\beta}\right)}\right) \end{aligned} \tag{B.3}$$

where: the second equality follows from the fact that $\boldsymbol{G}$ implies $\boldsymbol{d}$, 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 (B.1) and (B.2); 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 (B.3); note that we drop $\boldsymbol{\gamma}$ from the list of conditioned variables as it is implicit from conditioning on $\boldsymbol{d}$.

Proof of Proposition 2

This is largely an application of existing results to our particular case. We will 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 $s = 1, \dots, S$:

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

where $\widetilde{\mathcal{H}}_s$ is some collection of, possibly repeated, elements of $\mathcal{H}_s$ (such as $\mathcal{H}_s$, or $\mathcal{H}_s^*$), while $\boldsymbol{Z}_s$ is the subset of $\boldsymbol{Z}$ restricted to those elements $\boldsymbol{z}_{ij}$ such that $s\left(j\right) = s$. The

RSL estimator thus maximizes the following log-likelihood function:

$$\log \mathscr{L}^* \left(\boldsymbol{\beta} \middle| \boldsymbol{d}, \boldsymbol{G}, \boldsymbol{Z} \right) = \frac{1}{S} \sum_{s=1}^{S} \log \varPhi \left(\boldsymbol{G}_s \middle| \mathcal{H}_s^*; \boldsymbol{Z}_s; \boldsymbol{\beta} \right). \tag{B.4}$$

We show that the maximum of (B.4) converges to the true value of $\boldsymbol{\beta}$, which we write as $\boldsymbol{\beta}_\circ$ (not to be mistaken for β_0), as $S \to \infty$.

Following McFadden (1978), the probability limit of (B.4) for $S \to \infty$, accounting for $\mathcal{H}_s^*$ being random and for sectors not being identically distributed, is:

$$\operatorname{plim} \log \mathscr{L}^* \left(\boldsymbol{\beta} \middle| \boldsymbol{d}, \boldsymbol{G}, \boldsymbol{Z} \right) = \lim_{S \to \infty} \frac{1}{S} \sum_{s=1}^{S} L \left(\boldsymbol{G}_s, \mathcal{H}_s, \mathcal{H}_s^*, \boldsymbol{Z}_s; \boldsymbol{\beta} \right), \tag{B.5}$$

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

$$\begin{aligned} & L \left(\boldsymbol{G}_s, \mathcal{H}_s, \mathcal{H}_s^*, \boldsymbol{Z}_s; \boldsymbol{\beta} \right) \equiv \\ & \equiv \int_{\mathbb{Z}_s} \left[\sum_{\mathcal{H}_s^* \in \mathbb{H}_s} \sum_{\boldsymbol{G}_s \in \mathcal{H}_s} \mathbb{P} \left(\mathcal{H}_s^* \middle| \boldsymbol{G}_s; \boldsymbol{Z}_s \right) \varPhi \left(\boldsymbol{G}_s \middle| \mathcal{H}_s; \boldsymbol{Z}_s; \boldsymbol{\beta}_\circ \right) \log \varPhi \left(\boldsymbol{G}_s \middle| \mathcal{H}_s^*; \boldsymbol{Z}_s; \boldsymbol{\beta} \right) \right] dF_Z \left(\boldsymbol{Z}_s \right) \\ & = \int_{\mathbb{Z}_s} \left[\sum_{\mathcal{H}_s^* \in \mathbb{H}_s} \frac{\sum_{\boldsymbol{H}_s \in \mathcal{H}_s^*} \exp \left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} h_{ij} \boldsymbol{z}_{ij}' \boldsymbol{\beta}_\circ \right)}{\sum_{\boldsymbol{H}_s \in \mathcal{H}_s} \exp \left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{S}_s} h_{ij} \boldsymbol{z}_{ij}' \boldsymbol{\beta}_\circ \right)} \cdot \right. \\ & \qquad \left. \cdot \sum_{\boldsymbol{G}_s \in \mathcal{H}_s} \mathbb{P} \left(\mathcal{H}_s^* \middle| \boldsymbol{G}_s; \boldsymbol{Z}_s \right) \varPhi \left(\boldsymbol{G}_s \middle| \mathcal{H}_s^*; \boldsymbol{Z}_s; \boldsymbol{\beta}_\circ \right) \log \varPhi \left(\boldsymbol{G}_s \middle| \mathcal{H}_s^*; \boldsymbol{Z}_s; \boldsymbol{\beta} \right) \right] dF_Z \left(\boldsymbol{Z}_s \right), \end{aligned} \tag{B.6}$$

where $\mathbb{Z}_s$ is the support of $\boldsymbol{Z}_s$; $F_Z \left(\boldsymbol{Z}_s \right)$ is the cumulative distribution function of $\boldsymbol{Z}_s$; 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 $\mathbb{P} \left(\mathcal{H}_s^* \middle| \boldsymbol{G}_s; \boldsymbol{Z}_s \right)$ terms reduce to a constant; thus, (B.5) is understood as a weighted average of terms of the kind:

$$\sum_{\boldsymbol{G}_s \in \mathcal{H}_s} \varPhi \left(\boldsymbol{G}_s \middle| \cdot\, ; \boldsymbol{\beta}_\circ \right) \log \varPhi \left(\boldsymbol{G}_s \middle| \cdot\, ; \boldsymbol{\beta} \right)$$

where $\sum_{\boldsymbol{G}_s \in \mathcal{H}_s} \varPhi \left(\boldsymbol{G}_s \middle| \cdot\, ; \boldsymbol{\beta} \right) = 1$. Hence, the maximum of (B.5) is unique and equals $\boldsymbol{\beta}_\circ$; moreover, by standard arguments the maximum of (B.4) converges in probability to that of (B.5); therefore, the RSL estimator is consistent. This derivation is easily extended to the case in which uniform conditioning does not hold but $\mathbb{P} \left(\mathcal{H}_s^* \middle| \boldsymbol{G}_s; \boldsymbol{Z}_s \right)$ 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 $\boldsymbol{Z}$ is:

$$\mathscr{I}_S\left(\widetilde{\boldsymbol{\beta}}_{RSL}\middle|\,\boldsymbol{Z}\right)=\frac{1}{S}\sum_{s=1}^{S}\mathbb{E}_{\mathbb{H}_s}\left[\sum_{\boldsymbol{H}_s\in\mathcal{H}_s}\mathit{\Phi}\left(\boldsymbol{H}_s\middle|\mathcal{H}_s;\boldsymbol{Z}_s;\boldsymbol{\beta}_\circ\right)\boldsymbol{Z}_\circ\left(\boldsymbol{H}_s\right)\boldsymbol{Z}_\circ'\left(\boldsymbol{H}_s\right)-\right.$$
$$\left.-\sum_{(\boldsymbol{H}_s,\boldsymbol{K}_s)\in\mathcal{H}_s^2}\mathit{\Phi}\left(\boldsymbol{H}_s\middle|\mathcal{H}_s^*;\boldsymbol{Z}_s;\boldsymbol{\beta}_\circ\right)\mathit{\Phi}\left(\boldsymbol{K}_s\middle|\mathcal{H}_s^*;\boldsymbol{Z}_s;\boldsymbol{\beta}_\circ\right)\boldsymbol{Z}_\circ\left(\boldsymbol{H}_s\right)\boldsymbol{Z}_\circ'\left(\boldsymbol{K}_s\right)\right],\quad\text{(B.7)}$$

where $\boldsymbol{Z}_\circ\left(\boldsymbol{H}_s\right)\equiv\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{S}_s}h_{ij}\boldsymbol{z}_{ij}'\boldsymbol{\beta}_\circ$ and all expectations are taken over the $\mathbb{H}_s$ sets. The above is derived from (9) noting that the $\mathit{\Phi}\left(\boldsymbol{H}_s\right|\cdot)$ probabilities are not identically distributed across the S subnetworks, but are conditionally (on $\boldsymbol{Z}$) independent per Assumption 4. Since the information matrix of the unfeasible maximizer of (7) also follows (B.7) 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:

$$\boldsymbol{M}_s\left(\boldsymbol{Z}_s\right)\equiv\sum_{(\boldsymbol{H}_s,\boldsymbol{K}_s)\in\mathcal{H}_s^2}\boldsymbol{Z}_\circ\left(\boldsymbol{H}_s\right)\boldsymbol{Z}_\circ'\left(\boldsymbol{K}_s\right)\mathit{\Pi}\left(\boldsymbol{H}_s,\boldsymbol{K}_s\right|\mathcal{H}_s^*;\boldsymbol{Z}_s;\boldsymbol{\beta}_\circ\right),$$

where:

$$\mathit{\Pi}\left(\boldsymbol{H}_s,\boldsymbol{K}_s\right|\mathcal{H}_s^*;\boldsymbol{Z}_s;\boldsymbol{\beta}_\circ\right)\equiv\mathbb{E}_{\mathbb{H}_s}\left[\mathit{\Phi}\left(\boldsymbol{H}_s\middle|\mathcal{H}_s^*;\boldsymbol{Z}_s;\boldsymbol{\beta}_\circ\right)\mathit{\Phi}\left(\boldsymbol{K}_s\middle|\mathcal{H}_s^*;\boldsymbol{Z}_s;\boldsymbol{\beta}_\circ\right)\right]-$$
$$-\mathit{\Phi}\left(\boldsymbol{H}_s\middle|\mathcal{H}_s;\boldsymbol{Z}_s;\boldsymbol{\beta}_\circ\right)\mathit{\Phi}\left(\boldsymbol{K}_s\middle|\mathcal{H}_s;\boldsymbol{Z}_s;\boldsymbol{\beta}_\circ\right),\quad\text{(B.8)}$$

are symmetric positive matrices for $s=1,\dots,S$. This follows from some algebraic analysis as the (B.8) 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'Haultfœuille 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 setup. Before proceeding, it is useful to establish some more notation. For all firms $j\in\left(\mathcal{I}^d\cup\mathcal{I}^f\right)$, let $\chi_{jt}=\gamma_{jt}/\rho_n$, where $n=d$ if $j\in\mathcal{I}^d$ while $n=f$ if $j\in\mathcal{I}^f$. Let $\nu_{ikt}\equiv\beta_0-\log W_t+m_{is(k)t}$. Lastly, let $K_{is}=\left|\{k\in\mathcal{K}_i:s\left(k\right)=s\}\right|$ be the number of tasks which firm i can only source from external suppliers of sector s. 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 the dyadic covariates $\boldsymbol{z}_{ijt}$, the rescaled seller effects χ_{jt}, and ν_{ikt}, is obtained by Assumptions 3, 4a and 5 and for the reasons outlined in the discussion of (13) and (14) in the text as:

$$\mathbb{P}\left(j^*(i,k,t) = j, j \in \mathcal{I}^n \middle| \{(\chi_\ell, \boldsymbol{z}_{i\ell t})\}_{\ell \in \left(\mathcal{S}^d_{z(k)} \cup \mathcal{S}^f_{z(k)}\right)}, \nu_{ikt}\right) = \Psi_{istn} \Psi_{ijt|n}$$

where:

$$\Psi_{iktn} \equiv \mathbb{P}\left(j^*(i,k,t) \in \mathcal{I}^n \middle| \cdot\right) = \frac{\left[\sum_{\ell \in \mathcal{S}^n_{z(k)t}} \exp\left(\chi_{\ell t} + \boldsymbol{z}'_{i\ell t}\boldsymbol{\varphi}^n\right)\right]^{\rho_n}}{\exp(\nu_{ikt}) + \sum_{\widetilde{n} \in \{d,f\}} \left[\sum_{\ell \in \mathcal{S}^{\widetilde{n}}_{z(k)t}} \exp\left(\chi_{\ell t} + \boldsymbol{z}'_{i\ell t}\boldsymbol{\varphi}^{\widetilde{n}}\right)\right]^{\rho_{\widetilde{n}}}}$$

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

$$\Psi_{ijkt|n} \equiv \mathbb{P}\left(j^*(i,k,t) = j \middle| j^*(i,k,t) \in \mathcal{I}^n; \cdot\right) = \frac{\exp\left(\chi_{jt} + \boldsymbol{z}'_{ijt}\boldsymbol{\varphi}^n\right)}{\sum_{\ell \in \mathcal{S}^n_{z(k)t}} \exp\left(\chi_{\ell t} + \boldsymbol{z}'_{i\ell t}\boldsymbol{\varphi}^n\right)}$$

is the probability that j is chosen *conditional* on the choice falling within the $\mathcal{I}^n$ set. Similarly,

$$\Psi_{iktm} \equiv \mathbb{P}\left(j^*(i,k,t) = \varnothing \middle| \cdot\right) = \frac{\exp(\nu_{ikt})}{\exp(\nu_{ikt}) + \sum_{\widetilde{n} \in \{d,f\}} \left[\sum_{\ell \in \mathcal{S}^{\widetilde{n}}_{z(k)t}} \exp\left(\chi_{\ell t} + \boldsymbol{z}'_{i\ell t}\boldsymbol{\varphi}^{\widetilde{n}}\right)\right]^{\rho_{\widetilde{n}}}}$$

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 task pairs (k, k') such that $s(k) = s(k')$.

By extending (B.1), one obtains the joint mass function of the networks (domestic and foreign) at $t = 1, \dots, T$, for $\boldsymbol{\chi}^n_t = \{\chi_{jt}\}_{j \in \mathcal{I}^n_t}$ and $\boldsymbol{\nu}_t = \{\nu_{i1t}, \dots, \nu_{iSt}\}_{i \in \mathcal{I}^d_t}$, as:

$$\begin{aligned}
& f_{G,G}\left(\boldsymbol{G}^d_t, \boldsymbol{G}^f_t \middle| \boldsymbol{\chi}^d_t, \boldsymbol{\chi}^f_t, \boldsymbol{Z}^d_t, \boldsymbol{Z}^f_t, \boldsymbol{\nu}_t\right) = \\
& = \prod_{i \in \mathcal{I}^d_t} \prod_{k \in \mathcal{K}_i} \left[\left(\Psi_{iktm}\right)^{\mathbb{1}[j^* = \varnothing]} \prod_{j \in \mathcal{I}^d_t} \left(\Psi_{iktd} \Psi_{ijkt|d}\right)^{\mathbb{1}[j^* = j]} \prod_{j' \in \mathcal{I}^f_t} \left(\Psi_{iktf} \Psi_{ij'kt|f}\right)^{\mathbb{1}[j^* = j']}\right] \\
& = \prod_{i \in \mathcal{I}^d_t} \prod_{j \in \mathcal{I}^d_t} \Psi^{g_{ijt}}_{ijkt|d} \prod_{i \in \mathcal{I}^d_t} \prod_{j' \in \mathcal{I}^f_t} \Psi^{g_{ij't}}_{ij'kt|f} \prod_{s=1}^{S} \prod_{i \in \mathcal{I}^d_t} \left(\Psi_{istm}\right)^{K_{is} - B^d_{ist} - B^f_{ist}} \left(\Psi_{istd}\right)^{B^d_{ist}} \left(\Psi_{istf}\right)^{B^f_{ist}},
\end{aligned} \tag{B.9}$$

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 (B.9) follows from Assumption 4a, and in particular on the $\boldsymbol{\varepsilon}_{ikt}$ 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 sectors (hence the replacement of the k subscript with the s ones) 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 $\boldsymbol{b}_t^d$ and $\boldsymbol{b}_t^f$, for $n = d, f$ and $t = 1, \dots, T$, is derived as:

$$\begin{aligned} f_{b,b}\left(\boldsymbol{b}_t^d, \boldsymbol{b}_t^f \middle| \boldsymbol{\chi}_t^d, \boldsymbol{\chi}_t^f, \boldsymbol{Z}_t^d, \boldsymbol{Z}_t^f, \boldsymbol{\nu}_t\right) &= \prod_{i\in\mathcal{I}_t^d}\prod_{k\in\mathcal{K}_i} \left(\Psi_{iktm}\right)^{\mathbb{1}[j^*=\varnothing]} \left(\Psi_{iktd}\right)^{\mathbb{1}\left[j^*\in\mathcal{I}^d\right]} \left(\Psi_{iktf}\right)^{\mathbb{1}\left[j^*\in\mathcal{I}^f\right]} \\ &= \prod_{s=1}^{S}\prod_{i\in\mathcal{I}_t^d} \left(\Psi_{istm}\right)^{K_{is}-B_{ist}^d-B_{ist}^f} \left(\Psi_{istd}\right)^{B_{ist}^d} \left(\Psi_{istf}\right)^{B_{ist}^f}. \end{aligned} \tag{B.10}$$

Hence, the joint mass function of the networks, conditional on the in-degree sequences, can be obtained as the ratio between (B.9) and (B.10), which in turn can be factorized as follows, as the two networks of type d and f are conditionally independent:

$$\begin{aligned} f_{G,G|b,b}\left(\boldsymbol{G}_t^d, \boldsymbol{G}_t^f \middle| \boldsymbol{b}_t^d, \boldsymbol{b}_t^f, \boldsymbol{\chi}_t^d, \boldsymbol{\chi}_t^f, \boldsymbol{Z}_t^d, \boldsymbol{Z}_t^f, \boldsymbol{\nu}_t\right) &= \frac{f_{G,G}\left(\boldsymbol{G}_t^d, \boldsymbol{G}_t^f \middle| \cdot\right)}{f_{b,b}\left(\boldsymbol{b}_t^d, \boldsymbol{b}_t^f \middle| \cdot\right)} \\ &= \prod_{n\in\{d,f\}} f_{G|b}\left(\boldsymbol{G}_t^n \middle| \boldsymbol{b}_t^n, \boldsymbol{\chi}_t^n, \boldsymbol{Z}_t^n\right), \end{aligned}$$

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

$$f_{G|b}\left(\boldsymbol{G}_t^n \middle| \boldsymbol{b}_t^n, \boldsymbol{\chi}_t^n, \boldsymbol{Z}_t^n\right) = \prod_{i\in\mathcal{I}_t^d}\prod_{j\in\mathcal{I}_t^n}\left(\frac{\exp\left(\chi_{jt} + \boldsymbol{z}_{ijt}'\boldsymbol{\varphi}^n\right)}{\sum_{\ell\in\mathcal{S}_{z(k)t}^n}\exp\left(\chi_{\ell t} + \boldsymbol{z}_{i\ell t}'\boldsymbol{\varphi}^n\right)}\right)^{g_{ijt}}. \tag{B.11}$$

Note that the set of conditioned variables on the left-hand side of (B.11) only includes variables of type n. This is admissible as one can show that for $n = d$, the $\boldsymbol{b}_t^d$ vector is sufficient for the 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 $\boldsymbol{d}_t^n$ of type n, which is a sufficient statistic for the vector of rescaled fixed effects $\boldsymbol{\chi}_t^n$, for $n = d, f$; hence, given a proper definition of $\mathcal{H}_{st}^n$, (15) is obtained from a derivation specular to (B.3); the details are omitted.

Proof of Proposition 4

This is an application of standard results in econometrics; one can derive the result in two interrelated ways. The first is to acknowledge that the model delivers a categorical distribution representation of the probability to draw a specific subnetwork $\boldsymbol{H}_{st} \in \mathcal{H}_{st}^{n}$ for any $\mathcal{H}_{st}^{n}$ and for $n = d, f$, with the property that, given $\boldsymbol{\varphi}_{\circ}^{n} \equiv \boldsymbol{\beta}_{\circ}/\rho_{n}$:

$$\mathbb{E}\left[\mathbb{1}\left[\boldsymbol{G}_{st} = \boldsymbol{H}_{st}\right] \middle| \boldsymbol{Z}_{st}; \boldsymbol{\beta}_{\circ}\right] = \Phi\left(\boldsymbol{H}_{st}^{n} \middle| \mathcal{H}_{st}; \boldsymbol{Z}_{st}^{n}; \boldsymbol{\varphi}_{\circ}^{n}\right) \tag{B.12}$$

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:

$$\widehat{\boldsymbol{\varphi}}_{USL}^{n} = \underset{\boldsymbol{\varphi} \in \mathbb{R}^{Q}}{\arg\max} \prod_{t=1}^{T} \prod_{s=1}^{S} \frac{\exp\left(\sum_{i \in \mathcal{I}_{t}^{d}} \sum_{j \in \mathcal{S}_{st}^{n}} g_{ij} \boldsymbol{z}_{ij}' \boldsymbol{\varphi}\right)}{\sum_{\boldsymbol{H}_{st} \in \mathcal{H}_{st}^{n}} \exp\left(\sum_{i \in \mathcal{I}_{t}^{d}} \sum_{j \in \mathcal{S}_{st}^{n}} h_{ij} \boldsymbol{z}_{ij}' \boldsymbol{\varphi}\right)}, \tag{B.13}$$

satisfies the conditions of Theorem 5.4 and Corollary 5.5 by White (1994), which are based on Gourieroux et al. (1984). Consequently, $\operatorname{plim} \widehat{\boldsymbol{\varphi}}_{USL}^{n} = \boldsymbol{\varphi}_{\circ}^{n}$. In addition, one can adapt the proof of Proposition 2 to the longitudinal dimension; this would show that $\operatorname{plim} \widehat{\boldsymbol{\varphi}}_{USL}^{n} = \operatorname{plim} \widehat{\boldsymbol{\varphi}}_{RSL}^{n} = \boldsymbol{\varphi}_{\circ}^{n}$.

The second approach leverages the Method of Moments interpretation of QMLE. Again by adapting the proof of Proposition 2, it is easy to show that:

$$\mathbb{E}\left[\nabla_{\boldsymbol{\varphi}} \log \Phi\left(\boldsymbol{G}_{st} \middle| \mathcal{H}_{st}^{n*}; \boldsymbol{Z}_{st}; \boldsymbol{\varphi}_{\circ}^{n}\right)\right] = \boldsymbol{0} \tag{B.14}$$

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 function of the data-dependence structure specific to a setting.

C Extended discussion of the model

This appendix elaborates on selected features of both the theoretical and econometric sides of the framework developed in Section 3 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 equlibrium

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 the sake of simplicity, we restrict the discussion to the "static" framework from Section 3.1 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:

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

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 upperscript 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 $i \in \mathcal{I}$:

$$X_i^c = P_i^{-\psi} \left(\delta_i P \right)^{\psi - 1} E \tag{C.2}$$

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:

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

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:

$$\frac{P_{ijk} X_{ijk}}{C_i Y_i} = \frac{\left(\alpha_k / P_{ijk} \right)^{\sigma - 1}}{\Theta_i} \tag{C.4}$$

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 et al. (2017), the *sourcing capability* of firm i. Similarly:

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

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

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

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 $\mathcal{J}_i$; $\alpha_{ijk} \equiv \alpha_k \exp(\varepsilon_{ijk})$; μ_j^c is the markup charged by firm j on final consumers, while

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

is a firm's "average" (quantity-weighted) markup. Note that the dyadic frictions τ_{ij} do not appear explicitly in (C.6) since, due to their "iceberg" nature, they are extra costs borne by buyers not transferred to sellers; however, they 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 3 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:

$$j^{**}(i,k) = \underset{j\in\left\{\mathcal{S}_{z(k)}\setminus j^*(i,k)\right\}}{\arg\max} \gamma_j - \log\tau_{ij} + \varepsilon_{ijk} \tag{C.8}$$

we obtain:

$$\mu_{ijk} = \frac{C_{j^{**}}}{C_{j^*}} \frac{\tau_{ij^{**}}}{\tau_{ij^*}} \exp\left(\varepsilon_{ij^*k} - \varepsilon_{ij^{**}k}\right). \tag{C.9}$$

Final consumers are charged a monopolistically competitive markup:

$$\mu_i^c = \frac{\psi}{\psi - 1}, \tag{C.10}$$

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:

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

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

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

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 C.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* $\mathcal{G}^*$*:*

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

f. *a consumer price index* P*; and* g. *an aggregate expenditure* E*; such that equations* (2)*,* (5) (C.2)*,* (C.3)*,* (C.4)*,* (C.5)*,* (C.6)*,* (C.9)*,* (C.10)*,* (C.11)*,* (C.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 C.2. *In a* limit economy *all sectors grow indefinitely in size:* $|\mathcal{S}_s| \to \infty$ *for* $s = 1, \dots, S$*.*

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

Definition C.3. *An* acyclic input-output sectoral structure *is one where the* S *sectors can be ordered in such a way that buyers can only source their inputs from sectors that precede the buyer's in the order. Thus, there is an order* $s_{(1)}, \dots, s_{(S)}$ *such that for all* $i \in \mathcal{I}$*, if* $s(i) = s_{(1)}$ *it is* $\mathcal{K}_i = \varnothing$*; while if* $s(i) = s_{(r)}$ *for any* $r = 2, \dots, R$*, it is* $z(k) = s_{(q)}$ *with* $q < r$ *for all* $k \in \mathcal{K}_i$*.*

This definition implies a "tree-like" structure of all potential production networks that is determined by the technology characteristic of a sector; the order specified in the definition moves from upstream to downstream. Note that the first sector in the order only uses the labor input. While unlikely to hold in the real world, this concept is a useful abstraction that can approximate the technological input-output structure of actual production.

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:

$$\mathbf{c} = \mathbf{A}\left(\mathbf{l} + \mathbf{\Gamma c}\right), \tag{C.13}$$

where $\mathbf{c} = \left(C_1^{1-\sigma}, \dots, C_N^{1-\sigma}\right)$, $\mathbf{A} = \operatorname{diag}\left(A_1^{\sigma-1}, \dots, A_N^{\sigma-1}\right)$, $\mathbf{l} = W^{1-\sigma}\left(\alpha_{01}^{\sigma-1}, \dots, \alpha_{0N}^{\sigma-1}\right)$ while $\mathbf{\Gamma}$ is an $N \times N$ matrix whose Γ_{ij} elements, for $i, j = 1, \dots, N$, read as:

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

for some given $\mathcal{G}$. In what follows we write $\mathbf{\Gamma}^*$ as the version of $\mathbf{\Gamma}$ such that in (C.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 C.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 $\mathbf{c} = \left(\mathbf{I} - \mathbf{A\Gamma}^*\right)^{-1}\mathbf{Al}$, 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 (C.11) together with (C.6) allows one to solve for the total expenditures E via:

$$E = WL + \sum_{i \in \mathcal{I}} \frac{1}{\psi}\left(\frac{\left(\psi - 1\right)\delta_i A_i P}{\psi}\right)^{\psi-1}\Theta_i E.$$

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

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 C.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 $\mu_{ij^*(i,k)k} \to 1$ for all $i \in \mathcal{I}$ and $k \in \mathcal{K}_i$.*

Proof. Given (C.8) and (C.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:

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

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

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

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 it based on our proof.

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

Proof. The proof is constructive and inductive. Start from firms that belong to sector $s_{(1)}$: as they only decide on how much labor to hire and what combination of price and quantity to set in the final consumers market, their choices are uniquely determined. Move to firms in sector $s_{(2)}$: these must additionally buy inputs from firms in sector $s_{(1)}$. Yet, since the marginal costs in the latter are determined in the previous step, the choices of sector $s_{(2)}$ are also uniquely determined. Firms in sector $s_{(3)}$ source from both sectors $s_{(1)}$ and $s_{(2)}$, and so forth. Iterate until sector $s_{(S)}$ to obtain a candidate equilibrium *limit* network $\mathcal{G}^*$. Thus, make two observations. First, the $\mathbf{A\Gamma}^*$ matrix associated with $\mathcal{G}^*$ has spectral radius which is less then one; hence, the conditions of Lemma C.1 are satisfied and the economy's equilibrium is fully determined as long as markups are all unitary. This is in turn ensured by Lemma C.2. Consequently, the unique equilibrium of this particular limit economy is identified. $\square$

This analysis does not rule out equilibrium multiplicity (for examples, see Dhyne et al., 2023), nor degenerate cases without equilibria where firms would always want to update their supplier choices for some tasks as they observe other firms to do so. We note, however, that neither case is problematic for our empirical approach, since this treats firm choices as *conditional* on the actual marginal costs (firm fixed effects) that are realized in the data. We find it nonetheless useful to formalize the conditions under which our framework also exhibits desirable theoretical properties.

On the distribution of the errors

This section expands the discussion about Assumptions 4 and 4a of our framework, which specify distributional assumptions for the ε_{ijk} shocks of our model.

On other distributions. As it is well known, the multinomial logit probabilities that are foundational to our model may be obtained via other means. For example, if instead of (3) prices were assumed as:

$$P_{ijk} = \frac{\mu_{ijk} C_j \tau_{ij}}{\upsilon_{ijk}}, \tag{C.15}$$

where υ_{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 υ_{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 υ_{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 υ_{ijk} has shape parameter equal to π); then (6) is replaced by:

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

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 (C.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 4a summarized in Section 3.2 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 $P_{ijkt} \propto \mu_{ijkt} C_{jt} \tau_{ijt} \exp(-\omega_{ijkt})$. We assume that in each time period, these shocks are independent across buyers and tasks. However, it is:

$$\mathbb{C}\mathrm{ov}\left(\omega_{ijkt}, \omega_{i\ell kt}\right) = \begin{cases} \varrho_d & \text{if } (j,\ell) \in \left(\mathcal{I}^d\right)^2, \\ \varrho_f & \text{if } (j,\ell) \in \left(\mathcal{I}^f\right)^2, \\ 0 & \text{otherwise;} \end{cases} \tag{C.17}$$

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:

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

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, $T_0 \to \infty$ implies convergence in distribution of ε_{ijkt} to the distribution specified by Assumption 4a 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 4a might be modified so as to accommodate the following joint distribution of the error term:

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

where the parameters $\rho_{nst} \in (0, 1]$ vary over $n = d, f$, $s = 1, \ldots, S$ and $t = 1, \ldots, T$. This is conducive to a random parameter treatment of our estimation framework as discussed in section 3.3 of the text.

Allowing for structural dynamics

As we discussed in Section 3, allowing for structural dynamics would make our model misspecified by construction. To appreciate why, suppose that:

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

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:

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

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 (C.20) within the framework developed in the paper. Write the count of suppliers that have been fulfulling any task of a firm i for at least two years up to time t as:

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

and consider the following replacement for Assumption 5.

Assumption 5a. *The set of dyadic characteristics $\{\boldsymbol{z}_{ijt}\}_{j\in\mathcal{I}}$ 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:*

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

where $\widetilde{\phi} \geq \phi(\sigma-1)$, 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 (C.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 (C.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 $j^*\left(i, k, t+u \middle| \mathcal{G}_{-(ik)t+u}\right)$ 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 C.2. *Let $\mathcal{G}_{-(ik)t+u} = \mathcal{G}_{-(ik)}$ be constant for all $u \in \mathbb{N}_0$. Under assumptions 1a, 2, 3, 4a, 5a, and structural dynamics as specified in (C.20), for all $u \in \mathbb{N}$ it is:*

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

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 $j^* = j^*(i, k, t) \neq j^*\left(i, k, t+1 \middle| \mathcal{G}_{-(ik)}\right) = j^\dagger$. As both $\boldsymbol{z}_{ijt}$ 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:

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

The second inequality (third line) follows from (12) and from $\mathcal{G}_{-(ik)}$ being fixed, and the third inequality (fourth line), from Assumption 5a. 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. □

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 5a 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 in one case, not statistically significant: at least in this particular case, this diminishes concerns about the defensibility of this approach.

D Supplement to the empirical application

This appendix elaborates on some miscellaneous aspects of our empirical application discussed in Section 5 of the text.

Classification of transactions by number of tasks. As motivated in Section 3.3, we implement a mixture-of-distributions approach to infer the unobserved number of tasks associated with each transaction. To this end, we calculate "normalized shares" as:

$$Normalized\ share_{ijt} = \frac{Share_{ijt} - avg\left(Share_{ijt}\right)}{sd\left(Share_{ijt}\right)}$$

where $Share_{ijt} = Transaction_{ijt}/Revenue_{it}$, while $avg\left(\cdot\right)$ and $sd\left(\cdot\right)$ are the empirical mean and standard deviation calculated among those transaction shares such that the four-digits sector of buyers is $s\left(i\right)$, that of sellers is $s\left(j\right)$, and time is t (these are the same quantities used to calculate the red-colored kernel density of Figure 2 in the text, see footnote 16). We thus provide an approximate normalized counterpart to equation 17, though with buyer revenue instead of total costs in the shares' denominator: the two quantities are proportional to one another in equilibrium, but the former is more accurately measured, as we cannot observe transactions that are too small. We drop un-normalized shares larger than one: likely transactions for long-term investments.

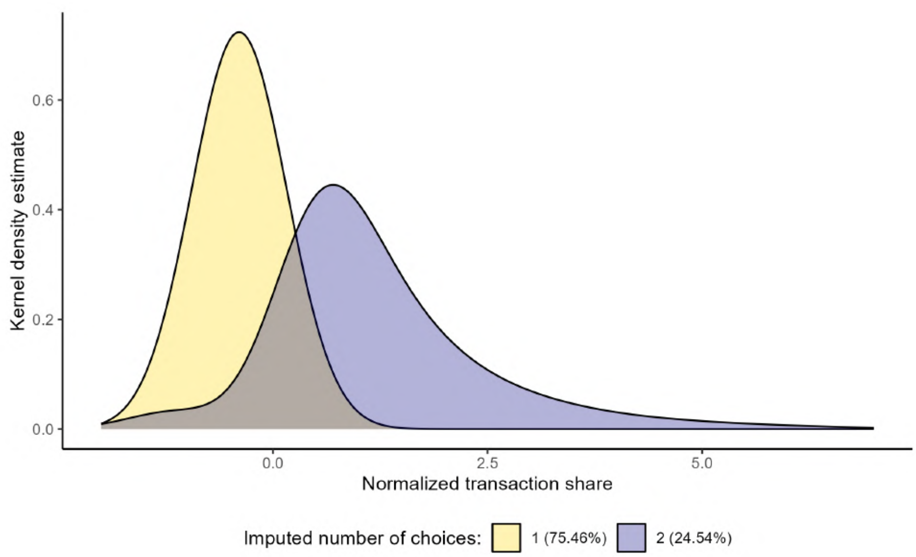


Figure D.1: Frequency of transactions by number of imputed choices or "tasks"

Notes. 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 (both defined at the four-digits level) and year, as in Figure 2 in the text. Two separate estimates are reported, one per group of transactions with given imputed number of "task" choices (1 versus 2). The legend reports the share of each group (in percentage) over the total. *Source*: Revec.

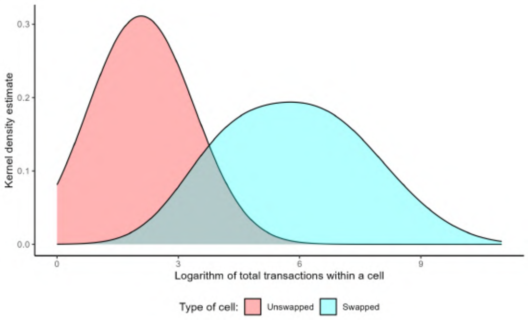
We fit a simple Gaussian mixture model with two components on the normalized shares, which we cluster according to their highest predicted component probability. The results are displayed in Figure D.1: 75.46 per cent of all transactions are classified as "one task," 24.54 per cent as "two tasks;" the figure reports kernel density estimates separately for both groups: they appear nicely unimodal, suggesting that two is the appropriate number of mixture components. We use the imputed value of $g_{ijt} \in \{1, 2\}$ to inform our RSL estimates. Clearly, under this approach some transactions may be misclassified, especially over the interval where the two densities overlap the most. In some unreported Monte Carlo simulations, we show that if the simulated production network has density comparable to the real one, this approach carries only a minuscule bias to the RSL estimates, if any.

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.2 dispels this concern: it displays 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 sharing the same seller sector and year that are exposed to *Ruta 27*) is distributionally the same for both groups. Hence, our sample cut rule does not select on the treatment. Furthermore, Figure D.2 shows that retained subnetworks feature buyer-seller pairs that are, on average, more distant in space (a mechanical effect due to the higher chance that small subnetworks are entirely confined within Costa Rica's Greater Metropolitan Area) and have about the same share of dyads that carry over from the previous year as the dropped subnetworks, albeit with a lower dispersion.

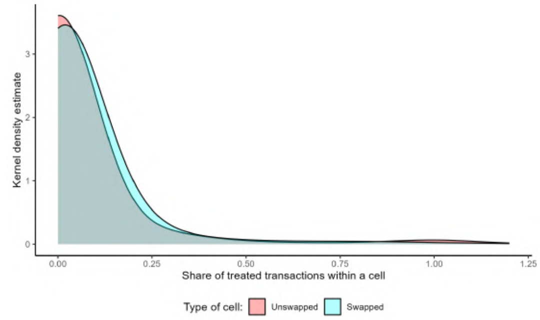
Naive multinomial logit estimates. We compare how RSL estimator fares against a seemingly simpler model for buyer choice that neglects the issue of seller fixed effects (hence "naive"). Thus, we estimate a number of specifications, symmetric to those of our RSL estimates from Table 4, of a multinomial 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;
- 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.

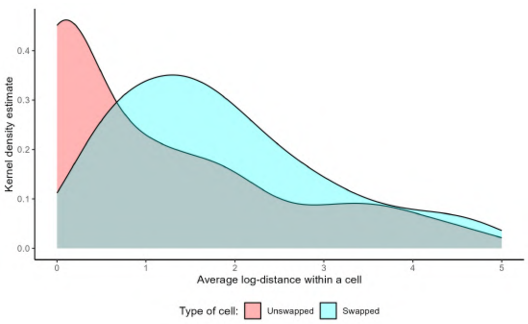
To let the BCCR computers handle the estimation problem, we are compelled to use only a random sample of about 10 per cent of all eligible transactions.



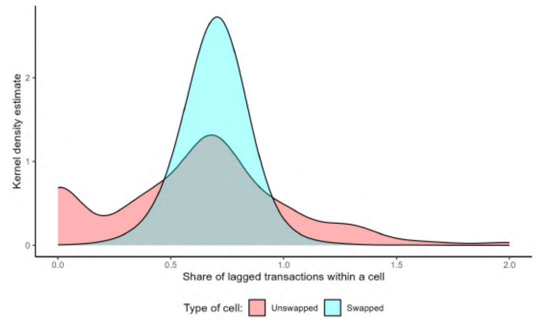
(a) Logarithm of the number of transactions



(b) Share of "treated" dyads



(c) Average logarithm of dyadic distance



(d) Share of dyads transacting at $t-1$

Figure D.2: Analysis of subnetworks (un-)selected for RSL estimation

Notes. This figure reports kernel density estimates (Gaussian kernel, 0.1 bandwidth size) about four variables, one per each subfigure, that are measured at the *subnetwork* level, where subnetworks are as discussed in Section 5. Each subfigure reports two estimates: one for the subnetworks used to inform the RSL estimates of Table 4 ("swapped," cyan) and one for the unused subnetworks ("unswapped," red). In subfigure (d) "shares" are calculated using, in the denominator, the number of tasks imputed for each transaction, averages are then calculated at the cell level; hence the support running from 0 through 2. *Source*: Revec.

The results from this estimation exercise are reported in Table D.1. Remarkably, for most parameters point estimates are larger by one or more orders of magnitude than the corresponding RSL estimates from Table 4, and are statistically significant at the 1 per cent level throughout. Signs are usually preserved, except for the logarithmic "size ratio" control. This suggests that neglecting seller fixed effects can seriously bias one's estimates; interpretations vary by explanatory variable. Thus, our dyadic binary treatment is arguably positively correlated with the seller fixed effects because *Ruta 27* connects the most productive and developed areas of Costa Rica; similarly, 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 seemingly associated with the "lag connection" dummy must be attributed to the latter clearly selecting sellers that are even *ex ante* more convenient.

Table D.1: Empirical application: "naive" multinomial logit estimates

Panel A: discrete treatment specification

	(1)	(2)	(3)	(4)
Treatment (*"Ruta 27"*)	0.406***	0.290***	0.237***	0.259***
	(0.016)	(0.020)	(0.020)	(0.030)
log(Size ratio)		0.812***	0.674***	0.547***
		(0.002)	(0.003)	(0.004)
log(Trade relative exposure)			0.073***	0.069***
			(0.001)	(0.001)
Transaction at $t-1$				5.588***
				(0.028)
Same province dummies	YES	YES	YES	YES
Distance decile dummies	YES	YES	YES	YES
Akaike information criterion	633301.12	404700.32	394235.10	179478.00
Number of choices	197,203	197,203	197,203	197,203

Panel B: continuous regressor specification

	(5)	(6)	(7)	(8)
log(Travel time)	−0.404***	−0.380***	−0.375***	−0.331***
	(0.002)	(0.003)	(0.003)	(0.004)
log(Size ratio)		0.822***	0.678***	0.550***
		(0.002)	(0.003)	(0.004)
log(Trade relative exposure)			0.076***	0.072***
			(0.001)	(0.001)
Transaction at $t-1$				5.592***
				(0.027)
Same province dummies	YES	YES	YES	YES
Akaike information criterion	646974.99	410568.55	399075.27	181548.50
Number of choices	197,203	197,203	197,203	197,203

Notes. This table reports estimates of a "naïve" multinomial logit model that explains observed transactions using the same dyadic variables as in the RSL estimates from Table 4 in the text, but without attempting to control for seller fixed effects. As in Table 4, this table displays estimates for both specifications under consideration, one per panel; the two panels follow the same structure as those from Table 4. Estimates are based on a random sample of 10 per cent of all post-2008 transactions used to compute the summary statistics from Table 3 in the text, reported here as the "number of choices" associated with each set of estimates; the set of alternatives used to construct the multinomial logit probabilities is also obtained via random sampling as described in the notes of Table 3. Estimates are weighted by each transaction's imputed number of "tasks," as described in this Appendix. Asterisk sequences *, ** and *** denote significance at the 10, 5, and 1 per cent level, respectively. *Source*: Revec.