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

Alonso Alfaro-Ureña† and Paolo Zacchia‡

Preview (extended abstract) version, July 2023

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 comments. We acknowledge, and express gratitude for, the fundamental role of Isabela Manelici and especially Jose Vasquez at coordinating our collaboration on this project. We owe special thanks to Santiago Campos-Rodríguez and César Ulate for their invaluable research assistance. This project was enabled by financial support primarily from Charles University's PRIMUS/21/HUM/022 grant. Financial support from the University's UNCE project (UNCE/HUM/035) is also gratefully acknowledged. The views expressed in this paper are the authors' only and are not necessarily shared by the BCCR.

†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 intermediate buyers is a key driver of firm 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 formed bears implications towards industrial policy,[1] the study of agglomeration economies,[2] firms in developing economies,[3] and more. In these settings, factors as diverse as policies targeting key 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, for use with data that register firm-to-firm transactions.[4] Specifically, we are interested in explanatory variables which, as in the previous examples, display a *dyadic variation*: that depend on the specific buyer-seller pair. With respect to conventional models of multinomial choice, this particular problem presents a central challenge: the prices of the alternative choices available to each buyer are typically unobserved in transaction data. In addition, prices are endogenous, as they depend on the marginal costs of direct sup-

[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 to specialized suppliers that are close in space generates economies of scale (Duranton and Puga, 2004; Moretti, 2011). However, there is to the best of our knowledge scant evidence on this particular mechanism. In our empirical application, we study the incidence of travel times on buyer-seller matching.

[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), Japan (Carvalho et al., 2021), Costa Rica (Alfaro-Ureña et al., 2022), Ecuador (Adão et al., 2022), Turkey (Demir et al., 2023), Uganda (Almunia et al., 2023) and others.

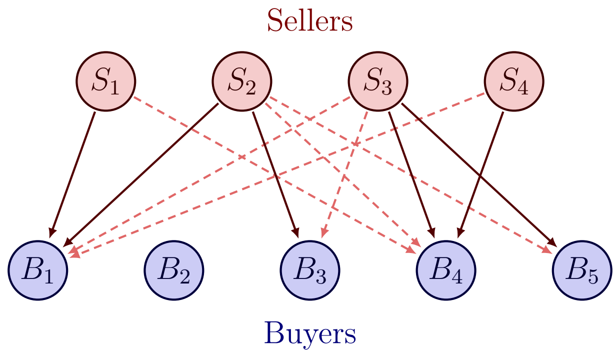
pliers, which are themselves recursively dependent on the more upstream structure of the production network (better suppliers-of-suppliers make direct suppliers more productive). Since the explanatory variables are likely to co-vary with firm marginal costs, the construction of a consistent estimator 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 conditional approach *à la* Chamberlain (1980), which is motivated by theoretical hypotheses embedded within a model of production network formation where the explanatory variables are treated as *matching frictions* between buyers and sellers, akin to iceberg trade costs in models of trade. 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 supplied by a seller in 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 simple stylized fact that we document using Costa Rican data: in the vast majority of cases, firms purchase their inputs from only one supplier of a given four-digits sector (which we treat as a proxy for the type of sourced task) in the same year.

This insight enables a transformation of the multinomial logit likelihood function that allows consistent estimation of the parameters of interest. However, this transformation differs substantively from those typical of conditional logit models for panel data, which express the conditional probability for the observed sequence of outcomes *over time*. Instead, our transformation expresses the conditional probability that *in the same time period* (year), and in a specific section of the production network: the one restricted to sellers from 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. In other words, in the transformation the observed subnetworks are pitted against alter-

[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 normalized cost shares conditional on the industry of both the buyer and the seller appears bimodal, which we interpret as an indication that some buyers source multiple tasks via a single transaction.

native configurations of the same subnetworks that display the same out-degree and in-degree sequences,[6] respectively for buyer 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 sharing the same degree sequences. One can think, for example, of the solid, darker edges to represent the observed subnetwork, whereas the dashed, lighter ones would 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 possibly even easier while also more intuitive, like a plain multinomial logit with fixed effects that

[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).

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 of Costa Rica affects the proability 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 the 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 less economically significant and statistically robust than those obtained from a naive multinomial logit that neglects the issue of seller fixed effects. In addition, the estimates about travel times and dependence on a previous period's connection that are obtained via our approach are statistically significant and display the expected sign, but are much smaller in magnitude than those from the naive logit.

Because of its aims, our paper connects with a number of fairly diverse strands of literature. We consider our contribution primarily an adaptation of multinomial logit models that condition on fixed effects (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 of 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. Furthermore, we show that replacing the time dimension with the network dimension in the likelihood transformation allows to identify the effect of the lagged dependent variable (presence of a connection in the previous time period)

[7]Our application thus speaks to the extant literature about the economic effects of infrastructures. Whereas previous studies typically focus on the effect upon regions (Faber, 2014) or firms (Holl, 2016) our application is, to the best of our knowledge, the first one to focus on firm-to-firm connections.

[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). This type of models inspire the estimators adversarial to ours in the Monte Carlo simulation.

under conventional assumptions, whereas as shown by Honoré and Kyriazidou (2000), in the standard model this is only possible under stronger conditions.

This paper also relates to the literature on the econometrics of network formation, that 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 preferences are determined by their profit functions, yet multiple equilibria are still possible as in typical models of production network formation. We thus outline the assumptions under which our approach is robust to the particular equilibrium selection.

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

[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 gives rise to multiple equilibria, which complicates the 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). We deem a buyer-initiated protocol more appropriate for the sake of developing an empirical framework to study production network formation, as it ties more naturally with models of discrete choice. Adding a dimension of *seller choice*, e.g. under seller capacity constraints, would be a valuable direction for future work.

warrant our distinct approach. First, the model by Panigrahi (2023) assumes a Cobb-Douglas firm production function, which we find too restrictive; we assume a constant elasticity of substitution (CES) technology instead. Second, he handles the seller fixed costs through an imputation approach which relies on the first order conditions of his model holding exactly; our sufficient statistic approach is lighter on the assumptions, and closer to standard models of discrete choice.

The complete version of this paper will be structured as follows. Section 2 will introduce the data and some novel stylized facts that we use to motivate our approach. Section 3 will detail on our conceptual framework and the resulting estimator. Section 4 will describe the Monte Carlo simulation and its results. Section 5 will develop our empirical application. Finally, Section 6 will conclude.

References

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, 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.

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.

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.

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

Panigrahi, Piyush (2023) "Endogenous Spatial Production Networks: Quantitative Implications for Trade and Productivity," Working paper.

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