Endogenous Production Networks and Firm Dynamics

Anuraag Aekka — World Bank

Gaurav Khanna — UCSD

May 2025

Updated regularly, please click here for latest version.

Abstract

We study the role of firm-to-firm matching in shaping firm dynamics and aggregate productivity. Using transaction-level data from a large Indian state, we document lifecycle patterns of customer and supplier networks. We find that younger firms have fewer customers and suppliers, lower sales and intermediate expenditures, and higher input costs. Motivated by these patterns, we develop a model of endogenous network formation where heterogeneous firms slowly match with partners over time using a random search technology. Firms in the decentralized equilibrium search inefficiently due to standard search externalities. This inefficiency turns out to be central for understanding how search technology shapes aggregate productivity. Calibrating the model to the Indian data, we find that most of the gains from technology improvement come from improvement in allocative efficiency, rather than in technical efficiency.

JEL Codes: D24, D61, D62, E22, F14, L14

We are highly grateful to Venky Venkateswaran and Michael Waugh for their generous advice and support. We also thank Jess Benhabib, Corina Boar, Giulia Brancaccio, Francisco Buera, Gian Luca Clementi, Niklas Engbom, Simon Gilchrist, Hugo Hopenhayn, Boyan Jovanavic, Ricardo Lagos, Guido Menzio, Virgiliu Midrigan, Alessandra Peter, Paul Scott, Daniel Xu, and seminar participants at NYU, Washington University in St. Louis, San Francisco Fed, BREAD, BSE Summer Forum, IIM Ahmedabad, and SED for valuable comments.

1 Introduction

Firm growth drives economic prosperity. But firms grow, to a large degree, through matching with trading partners (Afrouzi et al., 2023; Einav et al., 2022; Argente et al., 2023). As firms form new relationships with customers and suppliers–and dissolve existing ones–they reconfigure the production network. These matching decisions determine the network's structure, which, under production complementarities or gains from variety, influences aggregate productivity. Importantly, the technology governing firm-to-firm matching varies systematically across countries, reflecting differences in information infrastructure (Jensen, 2007; Aker, 2010; Goyal, 2010) and legal institutions (Boehm and Oberfield, 2020; Boehm, 2022). As such, cross-country differences in matching frictions, and firms' endogenous responses to them, are first-order determinants of firm dynamics, aggregate productivity, and economic development.

How does firm-to-firm matching technology shape firm dynamics, production network structure, and aggregate productivity? To answer this question, we develop a novel model of endogenous network formation where firms slowly match with customers and suppliers over time using a random search technology. The model rationalizes lifecycle patterns of customer and supplier networks, which we document in new firm-to-firm transaction data from India. Firms in the model make inefficient search choices due to standard externalities. Firms do not internalize the match surplus they generate for partners they match with, nor the congestion for other firms searching for partners.

Our new insight is that this *inefficiency* is central to understanding how search technology shapes aggregate productivity. When firms can scale their customer and supplier networks more easily, high-productivity firms grow faster and occupy a greater share of links in the production network, leading to higher aggregate productivity. Calibrating the model to match moments of the Indian data, we find that most of the gains from such technology improvement come from improvements in allocative efficiency (distance from efficient frontier), rather than in technical efficiency (position of efficient frontier).

We start our analysis by documenting lifecycle patterns of customer and supplier networks using firm-to-firm trade data from a large Indian state.[1] The data cover the near universe of transactions, where at least one partner in the transaction lies within the state. We find that younger firms have fewer customers and suppliers, and lower sales and intermediate expenditure. We then view the data through the lens of a CES demand system. This allows us to map trade flows between firms to the input costs firms face and the output prices they

[1]The state is twice Chile's population and three times Belgium's, both popular sources of similar data.

charge. We find that younger firms face higher input costs.

Conditioning on firm sales, we find that younger firms have fewer customers and suppliers, face higher input costs, and charge lower output prices. These patterns are difficult to generate in models that feature only dynamics in idiosyncratic productivity. The premise of the customer capital literature, that firms must spend resources to match with trading partners due to the presence of search and matching frictions (e.g., Arkolakis, 2010; Drozd and Nosal, 2012; Gourio and Rudanko, 2014), provides a natural way to rationalize them.

Motivated by these patterns, we develop a novel model of endogenous network formation where firms slowly match with trading partners over time using a random search technology. In the model, a continuum of monopolistically competitive firms produce differentiated varieties using labor and varieties produced by other firms. The economy features search and matching frictions, which entail that a firm is only able to trade with partners it is connected to. In order to connect with trading partners, in each period, firms exert costly search effort, both for matching with customers and matching with suppliers, and matches are made via an aggregate matching function. The cost of search effort is strictly convex, creating an incentive for firms to spread out search efforts over time. At the firm level, these search efforts give rise to the lifecycle patterns observed in the data. At the aggregate level, these search efforts give rise to an endogenous production network.

In general, introducing dynamic customer and supplier networks into an endogenous production network model adds significant complexity. When firms decide how much search effort to exert, they have to consider how potential partners will evolve over time. However, how a given customer (supplier) evolves over time depends itself on how the customer's (supplier's) own customers and suppliers evolve over time. This logic continues *ad infinitum*, such that an individual firm's evolution depends on the evolution of all firms upstream and downstream of it. In our setting, we maintain tractability by making assumptions on the network formation technology and firm productivity process, which imply that, in equilibrium, the age and productivity of a partner become a sufficient type to describe it.

We use the model to study how search technology shapes firm dynamics and aggregate productivity. We study the normative properties of our model and find that the decentralized equilibrium is inefficient due to standard vertical and search externalities. Firms set prices too high and underutilize intermediate inputs due to double marginalization. In addition, firms search inefficiently due to the presence of search externalities. When choosing search effort, firms do not internalize the surplus they generate for partners they match with. Furthermore, firms fail to internalize how their customer (supplier) search effort reduces matching probabilities for other firms searching for customers (suppliers).

Calibrating the model to the Indian data, we find that inefficient pricing and search choices generate large aggregate productivity losses. Aggregate productivity in the efficient allocation is 15% greater than in the decentralized equilibrium. Correcting only the double marginalization inefficiency increases aggregate productivity by 5.2%. Thus, most of the productivity gain in the efficient allocation comes from correcting search inefficiencies. These inefficiencies are central to understanding how differences in search technology map to aggregate productivity differences.

In the model, the search technology is summarized by two parameters, capturing the level and curvature of search costs. When firms are able to scale their trading partners more easily, due to lower curvature in search costs, search effort responds more to idiosyncratic productivity differences. As a result, high-productivity firms grow faster relative to low-productivity firms and make up a greater share of links in the production network. This generates higher aggregate productivity. On the other hand, when the level of search costs is lower, all firms search more and match with more partners. The resulting network features more links, generating higher aggregate productivity as firms benefit from gains from variety.

In the calibrated model, the curvature of search costs is disciplined by the elasticity between the number of customers and sales. In our data, we estimate this elasticity to be 0.33. For comparison, Arkolakis et al. (2023) find a higher elasticity in Chilean firm-to-firm data, suggesting that Chilean firms are able to scale their customer and supplier networks more easily. Using this moment as a plausible estimate of technological variation, we recalibrate our model to target the Chilean moment, holding fixed the total number of links. At the firm level, we find that lifecycle growth increases by 10% relative to the baseline. At the aggregate level, we find that aggregate productivity increases by 2%.

The inefficiencies present in our environment play a central role in generating this productivity gain. In inefficient economies, technology changes affect aggregate productivity by affecting both technical efficiency (position of the efficient frontier) and allocative efficiency (distance of the decentralized equilibrium from the efficient frontier). We find that roughly 85% of the 2% gain in aggregate productivity is due to improvement in allocative efficiency, while only 15% is due to improvement in technical efficiency.

Stagnant firm growth has been posited to be a crucial determinant of the differences in economic prosperity across countries. Hsieh and Klenow (2014) document lower lifecycle growth and aggregate productivity in India and Mexico in comparison to the US. Existing work has studied the importance of various factors in generating these patterns, including financial frictions (Cole et al., 2016) and managerial delegation frictions (Akcigit et al., 2021). Our results suggest that firm-to-firm matching technology may be a complementary driver of these patterns. Slow growth of firms in developing countries may be, in part, due to stronger search and matching frictions, which make scaling customer and supplier networks more difficult. These frictions, in turn, reduce aggregate productivity by reducing the prevalence of high-productivity firms in the production network.

Furthermore, our results suggest that much of the gains from improving firm-to-firm matching technology can be realized through policy which corrects search inefficiency. This is useful, given that improving firm-to-firm matching technology, for example, through improving information technology or legal institutions, appears to be prohibitively costly. In particular, our results imply that subsidizing search for high-productivity firms can generate significant aggregate productivity gains. Conversely, policies that subsidize search in an untargeted manner or those that subsidize search for low-productivity firms can lead to significant losses in productivity.

As for the level of search costs, the elasticity of aggregate output with respect to the level of search costs depends only on the intermediate share, the elasticity of substitution across inputs, and the curvature of search costs. Under our baseline calibration, a 10% reduction in the level of search costs leads to a 0.9% increase in aggregate productivity.

Related Work. This paper contributes to several strands of the literature studying endogenous production networks. This literature has taken seriously the idea that the production network is an endogenous object arising out of individual firm decisions and has documented numerous facts on firm-to-firm trade (e.g., Chaney, 2014; Oberfield, 2018; Bernard et al., 2019; Acemoglu and Azar, 2020; Taschereau-Dumouchel, 2020; Eaton et al., 2022; Bernard et al., 2022). Empirically, we contribute to this literature by adding facts on the lifecycle dimension of firm-to-firm links.

Theoretically, we relate to models in this literature where trading partners evolve over time. Lim (2018) studies a model where the value of relationships varies over time due to idiosyncratic shocks, and so the dynamics of customers and suppliers arise out of the dynamics of idiosyncratic shocks. Huneeus (2020) builds on this framework by introducing adjustment costs which prevent firms from readjusting customers and suppliers. In Boehm et al. (2024), potential suppliers arrive randomly according to a Poisson process. We differ from existing work by modeling dynamics that arise from a random search technology. This technological assumption provides a natural way to rationalize the lifecycle patterns we document in the data, and is consistent with evidence on search and matching frictions in firm-to-firm trade (e.g., Brancaccio et al., 2020; Miyauchi, 2024; Bergquist et al., 2024). We contribute by studying how *inefficiencies* from search externalities shape firm dynamics and aggregate productivity. In particular, we find that inefficiency in search is central to understanding how technological differences generate productivity differences.

In modeling the production network as being formed through search and matching, we relate to Arkolakis et al. (2023) and Demir et al. (2023). We extend these static models by introducing *dynamics* in customer and supplier networks. As discussed above, this introduces significant complexity as the evolution of an individual firm depends on the evolution of all firms upstream and downstream from it. We contribute by developing a tractable model of random search in a dynamic network setting.

The normative properties of our model are similar to those of other endogenous production network models. Though we differ by featuring dynamic customer and supplier networks, the underlying inefficiencies in our model are the same as those in models that rely on search and matching as the network formation technology (Arkolakis et al., 2023; Demir et al., 2023). The inefficiencies in our model are related to those in endogenous network models that rely on other network formation technologies (e.g., Lim, 2018; Huneeus, 2020), as even in these models, firms do not fully internalize how their network formation choices affect their partners. We contribute by demonstrating that inefficiency in search effort over the lifecycle generates quantitatively large aggregate productivity losses, and that changes in allocative efficiency are central for understanding how technological differences map to aggregate productivity differences.

This paper also contributes to the literature on customer capital. Theoretically, this literature has posited that firms must spend resources to match with trading partners due to the presence of search and matching frictions (e.g., Arkolakis, 2010; Drozd and Nosal, 2012; Gourio and Rudanko, 2014). This is also the case in our model. We differ from the customer capital literature by embedding firms in an endogenous network. As a result, search efforts of firms have implications for upstream and downstream partners and the structure of the production network. Empirically, this literature has documented the importance of customer growth in explaining sales growth. Examples of recent work include Einav et al. (2022), Fitzgerald et al. (2023), Afrouzi et al. (2023), and Argente et al. (2023). Our lifecycle patterns align with the findings of this literature, that firms grow through expanding their customer networks. We also add complementary patterns with respect to supplier networks.

The paper is organized as follows. In Section 2, we describe our data and document lifecycle patterns of customer and supplier networks. In Section 3, we describe our model and discuss its normative properties. In Section 4, we describe our calibration strategy and compare lifecycle patterns generated by the model against the data. In Section 5, we use the model to quantitatively study how search technology shapes firm dynamics and aggregate productivity.

2 Data and Lifecycle Facts

2.1 Data

Data on firm-to-firm relationships comes from daily transactions between firms in a large Indian state and trading partners in India and abroad. In April 2018, the state tax authority created an E-Way Bill System to improve tax compliance. Under the new system, any transaction with value exceeding 50,000 Rs (700 USD) must be reported electronically using the system. The system generates a waybill which the transporter must carry during shipment. The waybill contains the Permanent Account Number (Tax ID) of the supplier and customer, the 4-digit HSN code of the product, and the value of the shipment.

We use the sample of transactions that occurred between April 2018 and March 2020 (before the start of the Covid-19 pandemic). We drop any firms that are born after 2017 to ensure we observe a full year of transactions for every firm in the sample. As our empirical strategy relies on a connected set of firms, we keep the largest connected set (giant component) in each HSN. This leaves us with 5,000,000 links between 390,000 firms.

We add data on firm age from two separate sources. The first source is state records on various types of registration (e.g., year of incorporation, year of registration with state tax authority, etc.). We call the first year a firm is registered with the state as the registration year of the firm. The second source is a large online platform connecting buyers and sellers called IndiaMART. IndiaMART is the largest online B2B marketplace in India and contains rich information about firms on the platform. Included in this information is the year of establishment. We assign the birth year of a firm as the minimum between the year of establishment and the year of registration. In total, we can assign birth years to 170,000 firms in our sample.

Due to the introduction of a new state value added tax system in 2005, there is a large mass of firms with ages 11-14. Under the new law, many previously existing firms registered with the state authority. For this reason, in our empirical analysis, we will bin firms 11 and older into a single group.

2.2 Lifecycle Facts on Customers and Suppliers

In this section, we document lifecycle patterns of customer and supplier networks. Specifically, we define four age groups, {Age 1-3, Age 4-6, Age 7-10, Age 11+}, and estimate lifecycle patterns using using the following specification:

$$y_{i,h,t} = \sum_{a=2}^{4} \gamma_a \mathbf{1}(age(i,t) \in g_a) + \delta_{o(i,h,t),h,t} + \epsilon_{i,h,t} \tag{1}$$

Here, $y_{i,h,t}$ denotes firm i's outcome in product HSN h and year t; $\mathbf{1}(age(i,t) \in g_a)$ is an indicator which equals 1 if the age of firm i in year t, $age(i,t)$, is in age group g_a; and $o(i,h,t)$ is the location of the firm for HSN h in year t.[2] Our estimates of interest are age effects γ_a. We include $\delta_{o(i,h,t),h,t}$ to control for differences in firm outcomes across location-HSN-year. We omit the first age category, $a = 1$, so the age effects estimate outcomes of firms in a given age group relative to the youngest age group in a location-HSN-year. We refer to firms in the youngest age group as "entrants." Note that outcomes are within HSN, e.g. number of customers within an HSN or intermediate expenditure within an HSN.

Fact 1 *Younger firms have fewer customers and lesser sales.*

Taking $y_{i,h,t}$ as firm i's log number of customers and log sales in HSN h, we document lifecycle patterns of the number of customers and sales. Figure 1 plots the estimated age effects, γ_a, from Equation 1. Again, as we omit the youngest age group, our age effects estimate the log number of customers and log sales of firms in a given age group relative to entrants who sell products in the same HSN. Firms that are 11+ years old have 38% more customers within an HSN than entrants in the same HSN. Firms that are 11+ have 40% greater sales within an HSN than entrants in the same HSN.

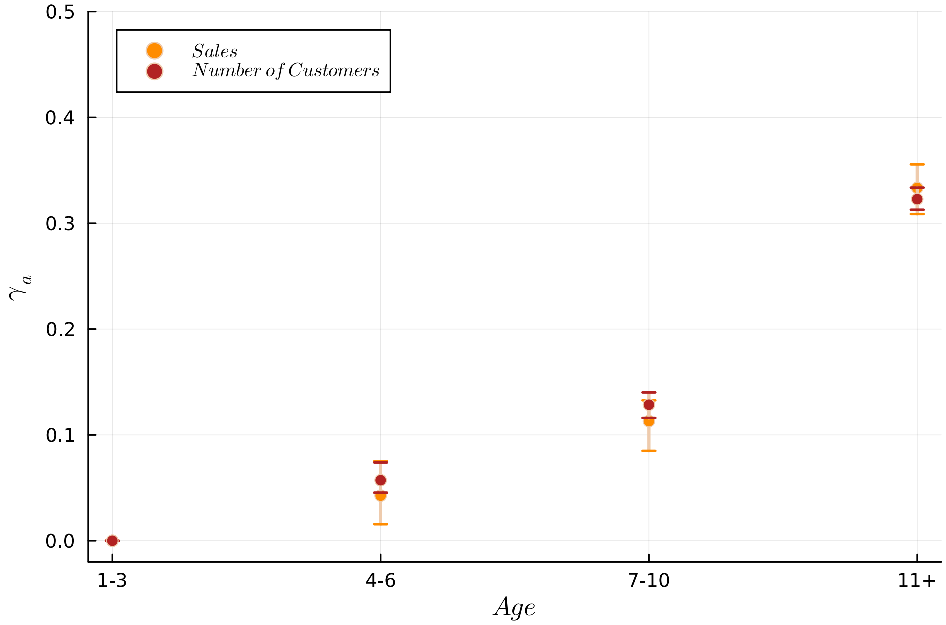
Fact 2 *Younger firms have fewer suppliers and lesser intermediate expenditure.*

Taking $y_{i,h,t}$ as firm i's log number of suppliers and log intermediate expenditure in HSN h, we document lifecycle patterns of number of suppliers and intermediate expenditure. Figure 2 plots the estimated age effects, γ_a, from Equation 1. Again, as we omit the youngest age category, our age effects estimate the log number of suppliers and log intermediate expenditure of firms in a given age category relative to entrants who purchase intermediate inputs from the same HSN. Firms that are 11+ years old have 12% more suppliers in an HSN than entrants who purchase inputs from the same HSN. Firms that are 11+ have 61% greater intermediate expenditure in an HSN than entrants who purchase inputs from the same HSN.

We next document lifecycle patterns of input costs and output prices. In order to do so, we impose additional restrictions on pricing and input demand.

[2]We assign the location of the firm in year t as the state from which the firm ships the greatest value of HSN h goods.

Figure 1: Sales and Number of Customers



Note: We plot estimated age effects, γ_a, from Equation 1 along with 95% bootstrap confidence intervals, taking $y_{i,h,t}$ to be log number of customers and log sales. The age effects estimate the log number of customers (within an HSN) and log sales (within an HSN) of firms in a given age category relative to entrants who sell products in the same HSN.

Assumption 1 ***(CES Demand System):***

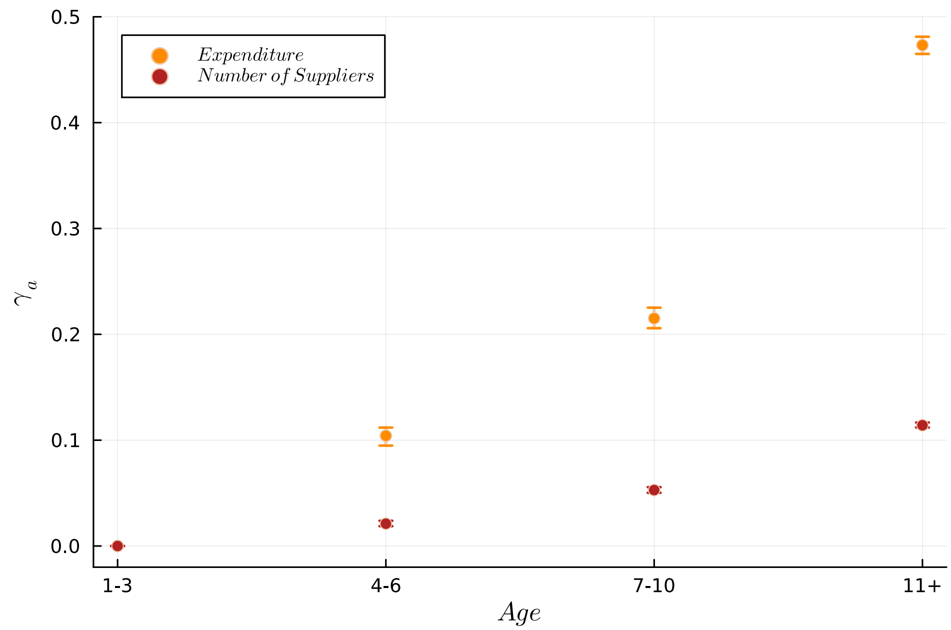
1. *For each HSN h, supplier i charges constant price, $p_{i,h,t}$, and provides constant quality, $q_{i,h,t}$, to all of its customers in year t.*
2. *For each HSN h, customer j minimizes the cost of sourcing an h intermediate good, where intermediate good h is a composite of varieties purchased from different suppliers:*

$$\min_{\nu_{i,j,h,t}} \sum_{i \in G_{j,h,t}} p_{i,h,t}\nu_{i,j,h,t} \qquad s.t. \quad \left(\sum_{i \in G_{j,h,t}} (q_{i,h,t}\nu_{i,j,h,t})^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \geq x_{j,h,t}$$

Here $x_{j,h,t}$ denotes the quantity of intermediate good h customer j requires in year t, $G_{j,h,t}$ denotes the set of h suppliers customer j is connected to in year t, and $\nu_{i,j,h,t}$ denotes the quantity of supplier i's variety customer j demands in year t. Varieties are imperfect substitutes with constant elasticity of substitution, $\sigma \geq 1$.

Solving the customer's cost minimization problem, the share of HSN h expenditure customer j spends on supplier i (log expenditure share) is given by a log-linear function of the supplier's quality-adjusted output price and the customer's quality-adjusted input cost:

Figure 2: Intermediate Expenditure and Number of Suppliers



Note: We plot estimated age effects, γ_a, from Equation 1 along with 95% bootstrap confidence intervals, taking $y_{i,h,t}$ to be log number of suppliers and log intermediate expenditure. The age effects estimate the log number of suppliers (within an HSN) and log intermediate expenditure (within an HSN) of firms in a given age category relative to entrants who purchase inputs from the the same HSN.

$$e_{i,j,h,t} = (1-\sigma)log\left(\frac{p_{i,h,t}}{q_{i,h,t}}\right) - (1-\sigma)log\left(c_{j,h,t}\right) \ , \tag{2}$$

where

$$c_{j,h,t} = \left(\sum_{i \in G_{j,h,t}} \left(\frac{p_{i,h,t}}{q_{i,h,t}}\right)^{1-\sigma}\right)^{1/(1-\sigma)}$$

Assumption 1 allows one to map trade flows between firms to input costs firms face and output prices they charge. This comes at the cost of imposing additional restrictions on pricing and input demand. However, this proves useful when one is interested in studying quality-adjusted costs and prices, but lacks information on product quality; as is the case in our setting.

Moreover, we find support for Assumption 1 in our data. The waybills that comprise our transaction data also have an entry for the quantity of goods shipped. This information is not required by the tax authority, so it is sometimes missing in the waybills. However, for the subset of transactions for which we can observe both transaction values and quantities, we can construct unit values. In Appendix A.2, we decompose variation in unit values and find that most of the variation can be explained by supplier fixed effects. In other words, variation within supplier across customers plays a small role in explaining the total variation in unit values. Additionally, in Appendix A.2, we find that alternative theories of pricing add little over Assumption 1 in terms of explaining variation.

Guided by Equation 2, we estimate the following regression equation:[3]

$$e_{i,j,h,t} = \psi_{i,h,t} + \phi_{j,h,t} + \varepsilon_{i,j,h,t} \tag{3}$$

We project log expenditure share onto a supplier fixed effect, $\psi_{i,h,t}$, and a customer fixed effect $\phi_{j,h,t}$. Under Assumption 1, $\psi_{i,h,t}$ corresponds to the supplier's quality-adjusted output price, and $\phi_{j,h,t}$ corresponds to the customer's quality-adjusted input cost.

We estimate Equation 3 and use $\phi_{j,h,t}$ and $\psi_{i,h,t}$ to document lifecycle patterns of input costs and output prices. We refer to these patterns as *CES Facts* since interpreting $\phi_{j,h,t}$ and $\psi_{i,h,t}$ as input costs and output prices requires Assumption 1.

CES Fact 1 *Younger firms face higher input costs.*

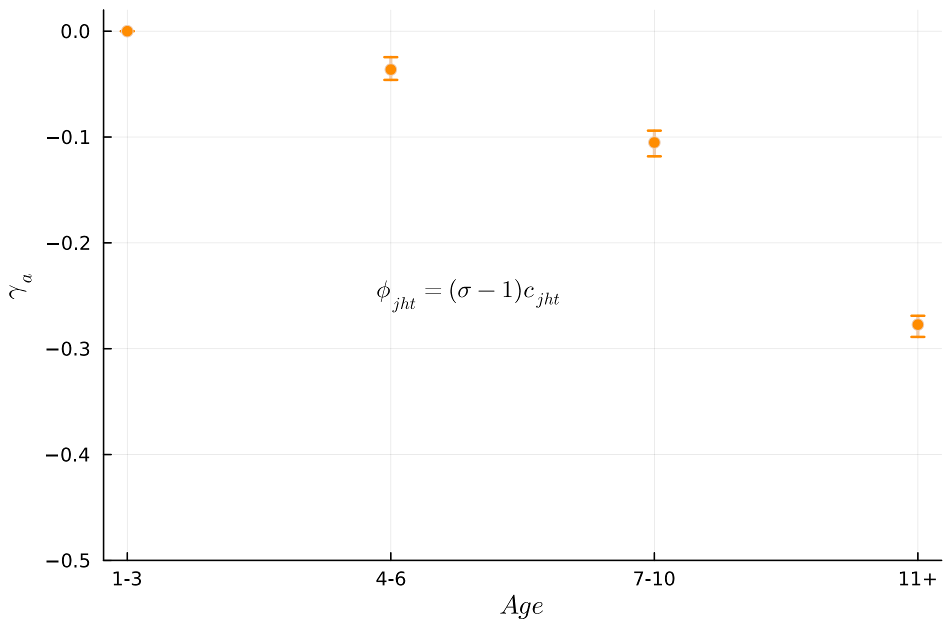
Taking $y_{j,h,t}$ in Equation 1 to be $\hat{\phi}_{j,h,t}$ estimated in Equation 3, we document lifecycle patterns of input costs. Figure 3 plots the estimated age effects, γ_a, from Equation 1. Again, as we omit the youngest age category, our age effects estimate the customer fixed effect $\phi_{j,h,t}$ of firms in a given age category, relative to entrants who purchase intermediate inputs from the same HSN. Firms that are 11+ have customer fixed effects that are 0.28 lower than entrants who purchase inputs from the same HSN. Noting that $\phi_{j,h,t}$ corresponds to $(\sigma - 1)log\,(c_{j,h,t})$ under Assumption 1, if $\sigma = 4.30$, this implies that firms that are 11+ have 8% lower input costs than entrants.

Taking $y_{i,h,t}$ in Equation 1 to be $\hat{\psi}_{i,h,t}$ estimated in Equation 3, we document lifecycle patterns of output prices. Figure 4 plots the estimated age effects, γ_a, from Equation 1. Again, as we omit the youngest age category, our age effects estimate the supplier fixed effect $\psi_{i,h,t}$ of firms in a given age category, relative to entrants who sell products in the same HSN. We

[3]Estimating Equation 3 using OLS poses a threat to identification. To obtain unbiased estimates, the assignment of suppliers to customers must be exogenous with respect to $\varepsilon_{i,j,h,t}$, an assumption referred to as "exogenous mobility" in the labor literature (Abowd et al., 1999; Card et al., 2013). In Appendix A.1, we argue that in the case this assumption is violated, our estimated differences between young and old firms represent lower bounds. That is, the true differences in input costs and output prices between young and old firms are larger.

In order for $\psi_{i,h,t}$ to be identified, a firm must have at least 2 customers in year t. Similarly, in order for $\phi_{j,h,t}$ to be identified, a firm must have at least 2 suppliers in year t. Thus, we will only be able to recover fixed effects for firms in our sample which meet this criteria. Furthermore, the estimated fixed effects can only be compared within a connected set. Thus, for each HSN, we isolate the largest connected set.

Figure 3: Expenditure Share Customer Effect $\phi_{j,h,t}$



Note: We plot estimated age effects, γ_a, from Equation 1 along with 95% bootstrap confidence intervals, taking $y_{j,h,t}$ to be $\hat{\phi}_{j,h,t}$ estimated in Equation 3. The age effects estimate customer fixed effects $\phi_{j,h,t}$ of firms in a given age category, relative to entrants who purchase intermediate inputs from the same HSN. Under Assumption 1, $\phi_{j,h,t}$ corresponds to $(\sigma - 1)log\,(c_{j,h,t})$.

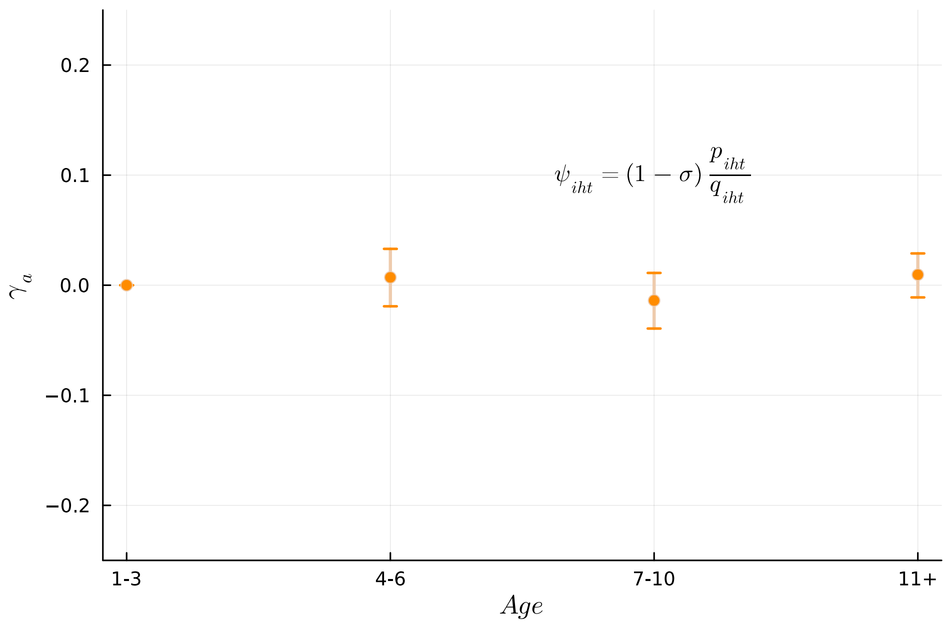
do not find that $\psi_{i,h,t}$ varies systematically with firm age.

Summarizing the patterns documented thus far, we find that younger firms have fewer customers and suppliers, lesser sales and intermediate expenditure, and higher input costs. The premise of the customer capital literature, that firms must spend resources to accumulate trading partners over time (Gourio and Rudanko, 2014; Drozd and Nosal, 2012), provides a natural way to rationalize these patterns. In the next section, we develop a model of endogenous network formation that builds on this premise.

These patterns, however, are also consistent with alternative theories of firm dynamics in which firm growth is driven by the evolution of idiosyncratic productivity (Hopenhayn, 1992). Introducing period-by-period firm-to-firm matching into such models could also give rise to similar patterns. In the case that more productive firms match with more partners, a positive correlation between the number of trading partners and age could just reflect a positive correlation between productivity and age.

In order to understand the extent to which our patterns can be accounted for by this explanation, we document patterns controlling for firm size. For downstream patterns, we

Figure 4: Expenditure Share Supplier Effect $\psi_{i,h,t}$



Note: We plot estimated age effects, γ_a, from Equation 1 along with 95% bootstrap confidence intervals, taking $y_{j,h,t}$ to be $\hat{\psi}_{i,h,t}$ estimated in Equation 3. The age effects estimate supplier fixed effects $\psi_{i,h,t}$ of firms in a given age category, relative to entrants who sell products in the same HSN. Under Assumption 1, $\psi_{i,h,t}$ corresponds to $(1-\sigma)log\left(\frac{p_{i,h,t}}{q_{i,h,t}}\right)$.

estimate the following regression equation:

$$y_{i,h,t} = \sum_{a=2}^{7} \gamma_a \mathbf{1}(age(i,t) \in g_a) + \beta_h s(i,h,t) + \delta_{o(i,h,t),h,t} \ , \tag{4}$$

where the only difference from Equation 1 is that we now control for log firm sales, $s(i,h,t)$. We allow the coefficient on sales to vary by the HSN of the product. For upstream patterns, we estimate the following regression equation,

$$y_{i,h,t} = \sum_{a=2}^{7} \gamma_a \mathbf{1}(age(i,t) \in g_a) + \beta_{h,h^d(i,t)} s^d(i,t) + \delta_{o(i,h,t),h,t} \ , \tag{5}$$

where the difference from Equation 1 is that we introduce the log downstream sales of the firm, $s^d(i,t)$.[4] We allow the coefficient on downstream sales to vary by the upstream HSN for which $y_{i,h,t}$ is observed, h, and the downstream HSN in which the firm sells, $h^d(i,t)$.[5]

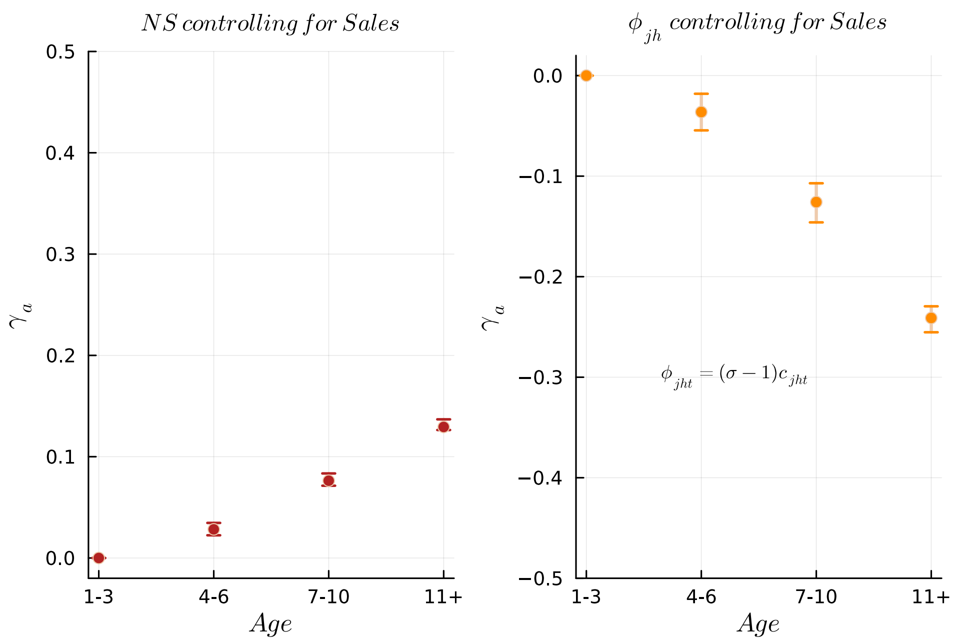
CES Fact 2 *Conditional on sales, younger firms have fewer suppliers and face higher input*

[4] $exp(s^d(i,t))$ sums sales of firm i across all downstream HSNs.
[5] We assign the downstream HSN of a firm as the HSN that makes up the greatest share of its sales.

costs.

In Figure 5, we plot estimated age effects from Equation 5. The left panel takes number of suppliers as $y_{i,h,t}$. The right panel takes $\hat{\phi}_{i,h,t}$ estimated in Equation 3 as $y_{i,h,t}$. As seen in the figure, conditional on having the same downstream sales, firms which are 11+ have 14% more suppliers than entrants and 0.24 lower estimated customer fixed effects, $\hat{\phi}_{i,h,t}$. Under Assumption 1, if $\sigma = 4.30$, this implies 7% lower input costs.

Figure 5: Upstream Patterns Controlling for Firm Size



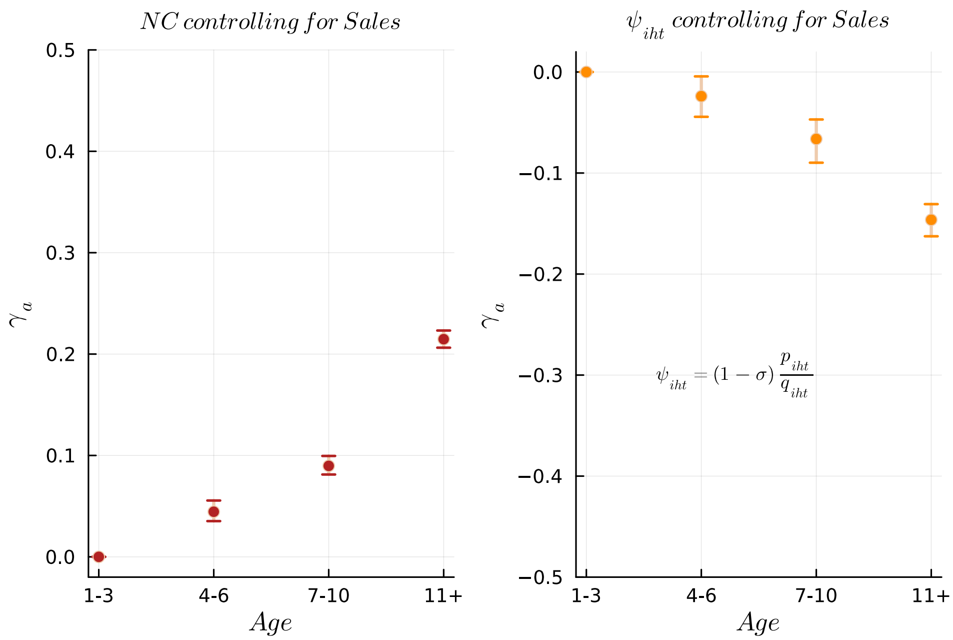
Note: We plot estimated age effects, γ_a, from Equation 5 along with 95% bootstrap confidence intervals, taking $y_{j,h,t}$ to be number of suppliers in the left panel and $\hat{\phi}_{j,h,t}$ estimated in Equation 3 in the right panel. The age effects estimate $y_{j,h,t}$ of firms in a given age category, relative to entrants who purchase intermediate inputs from the same HSN. Under Assumption 1, $\phi_{j,h,t}$ corresponds to $(\sigma - 1)log\,(c_{j,h,t})$.

CES Fact 3 *Conditional on sales, younger firms have fewer customers and charge lower output prices.*

In Figure 6, we plot the estimated age effects from Equation 4. The left panel takes number of customers as $y_{i,h,t}$. The right panel takes $\hat{\psi}_{i,h,t}$ estimated in Equation 3 as $y_{i,h,t}$. As seen in the figure, conditional on having the same sales, firms that are 11+ have 24% more customers than entrants and 0.15 lower estimated supplier fixed effects $\hat{\psi}_{i,h,t}$. Under Assumption 1, if $\sigma = 4.30$, this implies 4% *higher* output prices. Put differently, comparing firms of the same size, younger firms achieve these sales, despite having fewer customers, by charging lower output prices and selling more on the intensive margin.

Notice in the case that a firm's customer and supplier networks are completely determined

Figure 6: Downstream Patterns Controlling for Firm Size



Note: We plot estimated age effects, γ_a, from Equation 4 along with 95% bootstrap confidence intervals, taking $y_{i,h,t}$ to be number of customers in the left panel and $\hat{\psi}_{i,h,t}$ estimated in Equation 3 in the right panel. The age effects estimate $y_{i,h,t}$ of firms in a given age category, relative to entrants who sell products in the same HSN. Under Assumption 1, $\psi_{i,h,t}$ corresponds to $(1-\sigma)log\left(\frac{p_{i,hsn}}{q_{i,hsn}}\right)$.

by their idiosyncratic productivity, there should be no variation in these objects, conditional on sales. In contrast, we find that conditioning on sales, customer and supplier networks are systematically related with firm age. These patterns imply that dynamics of customer and supplier networks cannot be explained by idiosyncratic productivity alone. An alternative theory is required.

3 Model of Network Formation

In this section, we develop a develop a model of endogenous network formation in which firms slowly match with trading partners over time using a random search technology. In Section 4, we show that the model rationalizes the lifecycle patterns documented in the previous section. We use the model to quantitatively study how search technology shapes firm dynamics and aggregate productivity in Section 5.

3.1 Environment

Time is discrete. The economy is inhabited by a representative household which supplies labor and a continuum of monopolistically competitive firms that produce differentiated varieties. Each firm is endowed with a permanent productivity and operates a constant returns to scale technology that uses labor and varieties produced by other firms as inputs. The economy features search and matching frictions, which entail that a firm can only trade with other firms it has links with. In order to connect with trading partners, firms, in each period, exert costly search effort and match with new customers and suppliers via a random search technology.

Firms. There exists a continuum I of monopolistically competitive firms. Firm $i \in I$ produces a differentiated variety using a constant returns to scale technology:

$$y_t(i) = \kappa z(i) l_t(i)^{\alpha} x_t(i)^{1-\alpha}$$

Here, $z(i)$ denotes the firm's permanent productivity, $l_t(i)$ denotes the quantity of labor it uses in time period t, and $x_t(i)$ denotes the quantity of an intermediate bundle it uses. The intermediate bundle is a composite of varieties the firm purchases from its suppliers.

$$x_t(i) \equiv \left(\int_{G_{i,t}} \nu_t(i,k)^{(\sigma-1)/\sigma} dk \right)^{\sigma/(\sigma-1)}$$

$G_{i,t}$ denotes the set of suppliers of firm i in time period t, and $\nu_t(i,k)$ denotes the quantity of goods purchased from supplier k. Varieties are imperfect substitutes with constant elasticity of substitution, $\sigma \geq 1$. There exist search and matching frictions which entail that a firm can only trade with suppliers it has matched with. Firms exert costly search effort and match with new suppliers using a random search technology described further below. This search effort leads to the endogenous evolution of its supplier set over time. Given the set of suppliers in period t, firms minimize the cost of satisfying downstream demand $\bar{y}_t(i)$:

$$min_{l_t(i),\nu_t(i,k)}\, l_t(i) + \tilde{p}_t(i,k)\nu_t(i,k) \quad s.t. \quad \kappa z(i) l_t(i)^{\alpha} x_t(i)^{1-\alpha} \geq \bar{y}_t(i)\ , \tag{6}$$

where the wage is set as the numeraire and $\tilde{p}_t(i,k)$ denotes the price supplier k charges firm i. Solving the cost minimization problem, the price of firm i's intermediate bundle is given by:

$$c_t(G_{i,t}) = \left(\int_{G_{i,t}} \tilde{p}_t(k,i)^{1-\sigma} dk \right)^{1/(1-\sigma)}$$

We will refer to this price, $c(G_{i,t})$, as the "input cost" of the firm. The marginal cost of firm i is given by:

$$mc_t(i) = \frac{c_t(i)^{1-\alpha}}{z(i)}$$

Firms sell their variety to the representative household and to other firms. The set of firms that are customers of firm i is denoted by $H_{i,t}$. Again, the presence of search and matching frictions entails that a firm can only trade with customers it has matched with. Firms exert costly search effort and match with new customers using the random search technology described further below.

Representative Household. There exists a representative household that inelastically supplies its labor endowment $L = 1$, owns all firms, and consumes differentiated varieties. The household does not face any frictions in matching with firms, so that all firms are exogenously connected to the household. Household utility is given by a composite of varieties it consumes:

$$X_t = \left(\int_I \nu_t^f(i)^{(\sigma-1)/\sigma} di \right)^{\sigma/(\sigma-1)} ,$$

where $\nu_t^f(i)$ denotes final consumption of variety i. Varieties are imperfect substitutes with constant elasticity of substitution, $\sigma \geq 1$. Note that the household and firms share the same elasticity of substitution.

The household's price index is given by:

$$\mathcal{P}_t \equiv \left(\int_I p_t^f(i)^{1-\sigma} di \right)^{\frac{1}{1-\sigma}} ,$$

where $p_t^f(i)$ is the price firm i charges the household. The representative household owns all firms. Household income is the sum of labor income and firm profits:

$$E_t = 1 + \int_I \pi_t(i) di$$

Search and Matching Frictions. There exist search and matching frictions which entail that a firm can only trade with other firms it has matched with. Firms enter without any firm-to-firm links and slowly match with partners over time using a random search technology.

Specifically, in period t, firm i exerts search effort for finding customers, $u_t(i)$, and search effort for finding suppliers, $v_t(i)$. Exerting effort $u_t(i)$ costs the firm $\varphi_h(u_t(i))$ units of labor and exerting effort $v_t(i)$ costs the firm $\varphi_g(v_t(i))$ units of labor. We assume $\frac{\partial \varphi_h(u)}{\partial u} > 0$,

$\frac{\partial^2 \varphi_h(u)}{\partial u^2} > 0$, $\frac{\partial \varphi_g(v)}{\partial v} > 0$, and $\frac{\partial^2 \varphi_g(v)}{\partial v^2} > 0$. The assumption of strictly convex search costs creates an incentive for firms to spread out search efforts over time, generating the feature that firms slowly match with partners over time. This assumption captures the premise of the customer capital literature which posits that firms must spend resources to accumulate trading partners over time (Gourio and Rudanko, 2014; Drozd and Nosal, 2012). It will allow us to rationalize the lifecycle patterns documented in Section 2.

Given the search efforts of firms, the measure of new links formed in time period t is given by an aggregate matching function $\mathcal{M}(\mathcal{U}_t, \mathcal{V}_t)$, where $\mathcal{U}_t \equiv \int_I u_t(i)di$ is the aggregate customer search effort and $\mathcal{V}_t \equiv \int_I v_t(i)di$ is the aggregate supplier search effort. We assume the matching function exhibits constant returns to scale.[6]

Thus, exerting customer search effort $u_t(i)$ results in a measure of new customers:

$$u_t(i)\frac{\mathcal{M}(\mathcal{U}_t, \mathcal{V}_t)}{\mathcal{U}_t}$$

Similarly, exerting supplier search effort $v_t(i)$ results in a measure of new suppliers:

$$v_t(i)\frac{\mathcal{M}(\mathcal{U}_t, \mathcal{V}_t)}{\mathcal{V}_t}$$

Note that the firm matches with a measure of new customers and a measure of new suppliers, rather than discrete numbers. This implies that each firm is connected to a continuum of customers and a continuum of suppliers. Conditional on matching with a customer, the probability the customer is in measurable set $\mathcal{C} \subset I$ is proportional to the supplier search effort of firms in that set. Thus, the measure of new customers in set $\mathcal{C}$ is given by:

$$u_t(i)\frac{\mathcal{M}(\mathcal{U}_t, \mathcal{V}_t)}{\mathcal{U}_t}\frac{\int_{\mathcal{C}} v_t(k)dk}{\mathcal{V}_t}$$

Similarly, conditional on matching with a supplier, the probability the supplier is in measurable set $\mathcal{C} \subset I$ is proportional to the customer search effort of firms in that set. Thus, the measure of new suppliers in set $\mathcal{C}$ is given by:

$$v_t(i)\frac{\mathcal{M}(\mathcal{U}_t, \mathcal{V}_t)}{\mathcal{V}_t}\frac{\int_{\mathcal{C}} u_t(k)dk}{\mathcal{U}_t}$$

[6] There is limited empirical work estimating firm-to-firm matching functions. The assumption of constant returns to scale is in line with Krolikowski and McCallum (2021). It is also in line with empirical work estimating matching functions between workers and jobs in the labor market (Petrongolo and Pissarides, 2001). However, there is also some work suggesting increasing returns to scale in firm-to-firm matching functions (Miyauchi, 2024; Eaton et al., 2022).

At the end of every period, a share δ of existing relationships are exogenously destroyed.

In modeling the production network as arising from a search-and-matching technology, our model relates to Demir et al. (2023) and Arkolakis et al. (2023). We extend these models in introducing dynamics in the customer and supplier sets of firms.

Let $H_{i,t}(z', a')$ denote the measure of productivity z', age a' customers a firm is matched with in time period t and let $G_{i,t}(z', a')$ denote the measure of productivity z', age a' suppliers a firm is matched with in time period t.[7] Furthermore, let $\tilde{v}_t(z', a')$ denote a conjecture about the supplier search effort of (z', a') firms and $\tilde{u}_t(z', a')$ denote a conjecture about the customer search effort of (z', a') firms. These conjectures, combined with the matching and separation processes, imply the following laws of motion for customer and supplier sets:

$$H_{i,t}(z', a') = u_t(i) \frac{\mathcal{M}(\mathcal{U}_t, \mathcal{V}_t)}{\mathcal{U}_t} \frac{\tilde{v}_t(z', a') n_t(z', a')}{\mathcal{V}_t} + (1 - \delta) H_{i,t-1}(z', a' - 1) \tag{7}$$

$$G_{i,t}(z', a') = v_t(i) \frac{\mathcal{M}(\mathcal{U}_t, \mathcal{V}_t)}{\mathcal{V}_t} \frac{\tilde{u}_t(z', a') n_t(z', a')}{\mathcal{U}_t} + (1 - \delta) G_{i,t-1}(z', a' - 1) \ , \tag{8}$$

where $n_t(z', a')$ denotes the measure of (z', a') firms in the economy. Notice that these laws of motion conjecture that all (z', a') firms exert the same customer search effort and the same supplier search effort. These conjectures will be true in equilibrium.

Intuitively, the measure of (z', a') customers a firm is connected to in period t is equal to the measure of new (z', a') matches made this period plus the measure of $(z', a' - 1)$ links from the previous period which survive. Similarly, the measure of (z', a') suppliers a firm is connected to in period t is equal to the measure of new (z', a') matches made this period plus the measure of $(z', a' - 1)$ links from the previous period which survive. We will refer to the laws of motion described in Equation 7 and Equation 8 by:

$$H_{i,t} = \Gamma_t^h(u_t(i), H_{i,t-1})$$

$$G_{i,t} = \Gamma_t^g(v_t(i), G_{i,t-1})$$

The choice of indexing customers and suppliers by (z', a') is a deliberate one. Firms require information on potential customers and suppliers when choosing search effort. With respect to suppliers, firms require information on the output prices charged by potential suppliers, as well as how these prices will evolve over time. This depends not only on the productivity, z', of a given supplier, but also on the supplier's set of suppliers. With respect to customers,

[7] This is an abuse of notation, as before $G_{i,t}$ and $H_{i,t}$ were defined as sets.

firms require information on the input demand of potential customers, as well as how this demand will evolve over time. Again, this depends not only on the productivity, z', of a given customer, but also on the customer's customers and the customer's suppliers. Due to our assumptions on the network formation technology and the firm productivity process, in equilibrium, all (z', a') suppliers will charge identical output prices, and all (z', a') customers will have identical input demand. They will also exert identical search efforts and evolve in an identical manner. Thus, (z', a') serves as a sufficient type to identify customers and suppliers.

Timing. Timing in the model is as follows:

1. Measure n_t^e of new firms enter and draw permanent productivity.
2. Firms exert search effort and match with new customers and suppliers.
3. Firms produce and sell to the household and other firms.
4. Share $(1-\beta)$ of firms exogenously exit and share δ of relationships are exogenously destroyed.

3.2 Decision Problems

This section describes the decision problems faced by firms and the representative household.

Static Demand. In each period, firms sell to other firms and the representative household. As every firm and the representative household is connected to a continuum of suppliers, we assume the market structure is monopolistic competition.

Let $p_{i,t}(z', a')$ denote the price firm i charges a (z', a') customer in time period t, $\tilde{c}_t(z', a')$ denote a conjecture about a (z', a') customer's input cost, and $\tilde{m}_t(z', a')$ denote a conjecture about a (z', a') customer's total intermediate expenditure. The production technology implies that sales to a (z', a') customer are given by:

$$r_{i,t}(z', a') = \left(\frac{p_{i,t}(z', a')}{\tilde{c}_t(z', a')}\right)^{1-\sigma} \tilde{m}_t(z', a')$$

Similarly, sales to the representative household are given by:

$$r_{i,t}^f = \left(\frac{p_t^f(i)}{\tilde{\mathcal{P}}_t}\right)^{1-\sigma} \tilde{E}_t ,$$

where $\tilde{\mathcal{P}}_t$ and $\tilde{E}_t$ are conjectures about the household's price index and income. The conjectures about other firms and the household will be true in equilibrium.

Summing sales to downstream firms and the household, total sales for firm i are given by:

$$s_t(z(i), G_{i,t}, H_{i,t}) = \left(\frac{p_t^f(i)}{\tilde{\mathcal{P}}_t}\right)^{1-\sigma} \tilde{E}_t + \int_{Z,A} \left(\frac{p_{it}(z', a')}{\tilde{c}_t(z', a')}\right)^{1-\sigma} \tilde{m}_t(z', a') H_{i,t}(z', a') dz' da'$$

Dynamic Firm Problem. Firms maximize the discounted sum of profits. The firm's problem is described by the value function in Equation 9. A firm with talent z, previous-period customer set G_{t-1} and previous-period supplier set H_{t-1} chooses customer search effort u_t, supplier search effort v_t, price to charge the household p_t^f, and price to charge a (z', a') customer $p_t(z', a')$ to maximize the sum of current-period profits and value next period.

$$\begin{aligned} V_t(z, G_{t-1}, H_{t-1}) = \max_{u_t, v_t, p_t^f, p_t} & \frac{p_t^f - mc_t(z, G)}{p_t^f} \left(\frac{p_t^f}{\tilde{\mathcal{P}}_t}\right)^{1-\sigma} \tilde{E}_t \\ & + \int_{Z,A} \frac{p_t(z', a') - mc_t(z, G)}{p_t(z', a')} \left(\frac{p_t(z', a')}{\tilde{c}_t(z', a')}\right)^{1-\sigma} \tilde{m}_t(z', a') H_t(z', a') dz' da' \\ & - \varphi_h(u_t) - \varphi_g(v_t) + \beta V_{t+1}(z, G_t, H_t) \end{aligned} \tag{9}$$

where:

$$G_t = \Gamma_t^g(v_t, G_{t-1}), \quad H_t = \Gamma_t^h(u_t, H_{t-1})$$

$$mc_t(z, G) = \frac{1}{z} \left(\int_{Z,A} \tilde{p}_t(z', a')^{1-\sigma} G_t(z', a') dz' da'\right)^{\frac{1-\alpha}{1-\sigma}}$$

Modeling both pricing choices and search choices together introduces a potentially complicated problem. For example, in certain environments, firms may initially offer low prices to accumulate customers faster. Pricing choices can interact with search choices in a non-trivial way to shape the evolution of trading partners. Our assumptions on the network formation process provide tractability here.

First, notice that the price a firm charges does not affect the evolution of its customer or supplier network in Equations 7 and 8. This implies that a firm faces a static pricing problem and charges the standard monopolistic competition markup over its marginal cost. This feature in our model, that firms use non-price activities to accumulate demand, is in line with recent empirical evidence. Argente et al. (2023) find that markups in the consumer food sector do not systematically vary with a firm's age in a market. Fitzgerald et al. (2023) find that following successful entry into export markets, exporters have post-entry dynamics of quantities, but no post-entry dynamics of markups.

Second, the fact that all customers have the same constant elasticity of substitution implies that every firm charges the same markup $\mu = \frac{\sigma}{\sigma-1}$ to all of its customers. Taking this into account, we can rewrite the firm's value function as:

$$
\begin{aligned}
&V_t(z, G_{t-1}, H_{t-1}) = \max_{u_t, v_t} \frac{\mu - 1}{\mu} s_t(z, G_t, H_t) - \varphi_h(u_t) - \varphi_g(v_t) + \beta V_{t+1}(z, G_t, H_t) \\
&\text{where:} \\
&s_t(z, G_t, H_t) = \left(\tfrac{\mu c_t(G_t)^{1-\alpha}}{z\tilde{\mathcal{P}}_t}\right)^{1-\sigma} \tilde{E}_t + \textstyle\int_{Z,A} \left(\tfrac{\mu c_t(G_t)^{1-\alpha}}{z\tilde{c}_t(z',a')}\right)^{1-\sigma} \tfrac{1-\alpha}{\mu} \tilde{s}_t(z', a') H_t(z', a') dz' da' \\
&\qquad c_t(G_t) = \left(\textstyle\int_{Z,A} \left(\tfrac{\mu \tilde{c}_t(z',a')^{1-\alpha}}{z'}\right)^{1-\sigma} G_t(z', a') dz' da'\right)^{1/(1-\sigma)} \\
&\qquad G_t = \Gamma_t^g(v_t, G_{t-1}), \quad H_t = \Gamma_t^h(u_t, H_{t-1})
\end{aligned}
\tag{10}
$$

Relative to Equation 9, there are no longer explicit pricing choices as the firm charges a constant markup μ over its marginal cost. The firm still requires conjectures about the search efforts of other firms, $\tilde{u}_t(z', a')$ and $\tilde{v}_t(z', a')$. However, now the firm only requires conjectures about other firms' input costs, $\tilde{c}(z', a')$, and sales, $\tilde{s}_t(z', a')$, using the fact that all firms charge a constant markup μ over their marginal costs.

Solving the firm's problem, a firm's optimal customer search effort satisfies:

$$
\begin{aligned}
\frac{\mu - 1}{\mu} \Bigg(\sum_{\tau=0}^{\infty} \sum_{z',a'} ((1-\delta)\beta)^{\tau} \left(\frac{\mu c_{t+\tau}(G_{t+\tau})^{1-\alpha}}{z\tilde{c}_{t+\tau}(z', a'+\tau)}\right)^{1-\sigma} \frac{1-\alpha}{\mu} \tilde{s}_{t+\tau}(z', a'+\tau) \\
\times \frac{\mathcal{M}(\mathcal{U}_t, \mathcal{V}_t)}{\mathcal{U}_t} \frac{\tilde{v}_t(z', a') n_t(z', a')}{\mathcal{V}_t} \Bigg) = \frac{\partial \varphi_h(u_t)}{\partial u_t}
\end{aligned}
\tag{11}
$$

The term on the left is the marginal benefit of customer search effort. This is equal to the discounted sum of future profits generated by additional matches. Future profits generated from matching with a (z', a') customer depend on the path of the customer's sales, which governs their intermediate expenditure, and the path of the firm's output price relative to the customer's input cost, which governs the firm's share in the customer's intermediate expenditure. This marginal benefit is set equal to the marginal cost of search, which is the term on the right.

A firm's optimal supplier search effort satisfies:

$$\frac{\mu-1}{\mu}\left(\sum_{\tau=0}^{\infty}\sum_{z',a'}((1-\delta)\beta)^{\tau}\left(\frac{\mu\tilde{c}_{t+\tau}(z',a'+\tau)^{1-\alpha}}{z'c_{t+\tau}(G_{t+\tau})}\right)^{1-\sigma}(1-\alpha)s_{t+\tau}(z,G_{t+\tau},H_{t+\tau})\right.$$
$$\left.\times\frac{\mathcal{M}\left(\mathcal{U}_t,\mathcal{V}_t\right)}{\mathcal{V}_t}\frac{\tilde{u}_t(z',a')n_t(z',a')}{\mathcal{U}_t}\right)=\frac{\partial\varphi_g(v_t)}{\partial v_t}\quad(12)$$

The term on the left is the marginal benefit of supplier search effort. This is equal to the discounted sum of future profits generated by additional matches. Matching with suppliers reduces the marginal cost of the firm, allowing the firm to charge lower prices to their customers and sell more on the intensive margin. This marginal benefit is equated to the marginal cost of search, which is the term on the right.

We emphasize here that the presence of constant markups does not preclude competition among suppliers. As seen through Equation 12, when firms choose supplier search effort, they take into account how the output price they charge compares to the input cost of their customers. In other words, how their output price compares to the output prices charged by other suppliers. This informs their decision to search for suppliers and reduce their marginal cost. Competition incentivizes firms to reduce their price in this setting, but the channel through which they do this is through reducing their marginal cost, rather than their markup.

3.3 Equilibrium

Definition 1 *An equilibrium is a set of optimal search policies* $\left\{u_t(z,G,H),\, v_t(z,G,H)\right\}_{t=0}^{\infty}$; *optimal labor and intermediate input policies* $\{l_t(z,G,H),\ \nu_t(z,G,H)\}$; *conjectures about other firms' search efforts* $\left\{\tilde{u}_t(z,a),\tilde{v}_t(z,a)\right\}_{t=0}^{\infty}$, *input costs* $\left\{\tilde{c}_t(z,a)\right\}_{t=0}^{\infty}$, *sales* $\left\{\tilde{s}_t(z',a')\right\}_{t=0}^{\infty}$, *and customer and supplier sets* $\left\{\tilde{H}_{z,a,t}(z',a'),\tilde{G}_{z,a,t}(z',a')\right\}_{t=0}^{\infty}$; *and conjectures about aggregate objects,* $\left\{\tilde{\mathcal{P}}_t,\tilde{E}_t\right\}_{t=0}^{\infty}$ *such that:*

1. *Optimal search policies solve firm problem (10).*
2. *Optimal labor and intermediate input policies solve cost minimization problem (6).*
3. *Conjectures about customer and supplier sets are consistent with laws of motion:*

$$\tilde{H}_{z,a,t}=\Gamma_t^h\left(\tilde{u}_t(z,a),\tilde{H}_{z,a-1,t-1}\right),\quad\tilde{G}_{z,a,t}(z',a')=\Gamma_t^g\left(\tilde{v}_t(z,a),\tilde{G}_{z,a-1,t-1}\right)$$

4. *Conjectures about search efforts of other firms are consistent with optimal policies:*

$$\tilde{u}_t(z,a) = u_t(z, \tilde{G}_{z,a-1,t-1}, \tilde{H}_{z,a-1,t-1}), \quad \tilde{v}_t(z,a) = v_t(z, \tilde{G}_{z,a-1,t-1}, \tilde{H}_{z,a-1,t-1})$$

5. *Conjectures about input costs and sales are consistent with customer and supplier sets:*

$$\tilde{c}_t(z,a) = c_t(\tilde{G}_{z,a,t}), \quad \tilde{s}_t(z,a) = s_t\left(z, \tilde{G}_{z,a,t}, \tilde{H}_{z,a,t}\right)$$

6. *Conjectures about household objects are consistent with optimal policies:*

$$\tilde{\mathcal{P}}_t = \left(\int \left(\mu z(i)^{-1} c_t(G_{it})^{1-\alpha}\right)^{1-\sigma} di\right)^{1/(1-\sigma)}, \quad \tilde{E}_t = 1 + \frac{\mu - 1}{\mu} \int s_t\left(z(i), G_{it}, H_{it}\right) di$$

7. *Labor market clears:*

$$1 = \int_I \Big(l_t(z_i, G_{it}, H_{it}) + \varphi_h(u_t(z_i, G_{it}, H_{it})) + \varphi_g(v_t(z_i, G_{it}, H_{it}))\Big) di$$

For the remainder of this paper, we will focus on stationary equilibria.

Definition 2 *A stationary equilibrium is defined as an equilibrium in which optimal policies and conjectures are time-independent.*

Even in a static setting, solving for equilibria in endogenous network models is a complex problem. A firm's choice to form links depends on payoff-relevant attributes of potential trading partners. However, these attributes of trading partners depend themselves on the attributes of their own trading partners. Thus, in equilibrium, one needs to solve for two fixed points. First, a fixed point in payoff-relevant attributes, as the attributes of a firm depend on the attributes of its partners. Second, a fixed point in the link formation choices of firms, as the link formation choices depend on the attributes of potential trading partners. The endogenous production network literature has made progress here by considering environments in which these fixed points can be easily characterized and solved for.

Our setting further complicates this problem by introducing dynamics in customer and supplier networks. As in the static setting, a firm's decision to form links depends on the payoff-relevant attributes of potential trading partners. However, as trading partners themselves match with new customers and suppliers over time, their payoff-relevant attributes evolve over time. Furthermore, as the payoff-relevant attributes of trading partners depend on the attributes of their own trading partners, the evolution of their attributes depends on how their trading partners' attributes evolve over time. Again, in equilibrium, we need to find a fixed point in attributes and a fixed point in link formation choices, but here, as firms evolve over time, it cannot be a fixed point in attributes and choices of individual firms. Instead, our assumptions on the network formation technology and the firm productivity process provide a state, (z, a), such that these fixed points can be easily characterized and solved for.[8]

3.4 Constrained Inefficiency

In this section, we discuss the normative properties of the model. These properties turn out to be important for understanding how technology differences map to productivity differences in Section 5.

Consider the problem of a constrained planner who chooses household consumption of varieties, labor and intermediate input choices of firms, and search efforts of firms, to maximize consumption in a stationary equilibrium, imposing that the planner has available the same random search technology as decentralized firms:

$$\max_{\nu^f, l, \nu, u, v} \left(\sum_{z,a} \nu^f(z,a)^{\frac{\sigma-1}{\sigma}} n(z,a) \right)^{\frac{\sigma}{\sigma-1}}$$

subject to the following constraints:

$$\nu^f(z,a) + \sum_{z',a'} \nu_{z',a'}(z,a) H_{z,a}(z',a') \leq y(z,a) \quad \forall z, a$$

$$\sum_{z,a} l(z,a)n(z,a) + \sum_{z,a} \varphi_h(u(z,a))n(z,a) + \sum_{z,a} \varphi_g(v(z,a))n(z,a) \leq L$$

$$y(z,a) = \kappa z l(z,a)^{\alpha} \left(\sum_{z',a'} \nu(z,a,z',a')^{(\sigma-1)/\sigma} G_{z,a}(z',a') \right)^{\frac{\sigma(1-\alpha)}{\sigma-1}}$$

$$H_{z,a}(z',a') = u(z,a) \frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{U}} \frac{v(z',a')n(z',a')}{\mathcal{V}} + (1-\delta)H_{z,a-1}(z',a'-1) \quad \forall z, a, z', a'$$

$$G_{z,a}(z',a') = v(z,a) \frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{V}} \frac{u(z',a')n(z',a')}{\mathcal{U}} + (1-\delta)G_{z,a-1}(z',a'-1) \quad \forall z, a, z', a'$$

Here, $y(z,a)$ denotes the output of a firm with productivity z and age a, $u(z,a)$ and $v(z,a)$ denote its customer and supplier search efforts, $l(z,a)$ denotes its labor input, and $\nu(z,a,z',a')$ denotes its intermediate usage of a productivity z', age a' supplier's variety. Household

[8] In Appendix B.1, we describe the exact solution algorithm.

consumption of a (z, a) firm's variety is denoted by $\nu^f(z, a)$.

The first constraint imposes that the sum of household consumption of a variety and usage of it by other firms in production must be less than the output of that variety. The second constraint imposes that aggregate labor usage must be less than aggregate labor supply. The third constraint is just the production function of the firm. The last two constraints impose that the planner uses the same random search technology as firms in the decentralized equilibrium.

We discuss the planner's solution in detail in Appendix B.3. Here, we briefly highlight the inefficiencies present in the decentralized equilibrium.

First, firms in the decentralized equilibrium set prices too high and underutilize intermediate inputs as they fail to internalize a vertical externality (double marginalization). The price a supplier charges its customer affects the marginal cost and so the profits of the customer. Similarly, the price the customer charges its own customers affects demand for its variety, which in turn, affects its demand for inputs and so the profits of the supplier. As both customer and supplier in the decentralized equilibrium choose prices to maximize private profits rather than joint profits, they set prices too high and use too little intermediate input relative to the planner's solution.

Second, firms search inefficiently as they fail to internalize search externalities. When a firm exerts more search effort, it generates more matches, which creates surplus for both the firm and the partners it matches with. While firms internalize their private surplus, they do not internalize the surplus generated for partners. This externality has been termed the "thickness" and "composition" externalities in the search literature, and the "match creation" externality in some of the endogenous network literature. In addition, firms fail to internalize how their customer (supplier) search effort affects matching probabilities for other firms searching for customers (suppliers). This externality has been termed the "congestion" externality. In the case the match creation externality exactly offsets the congestion externality for every firm in the decentralized economy, the decentralized search efforts are efficient. However, under the decentralized surplus-splitting rule (i.e. constant markups), this is not possible and so search effort is inefficient. In particular, it will turn out that low-productivity firms search too much relative to high-productivity firms.

It is important to emphasize here that the inefficiency of search effort depends critically on the decentralized surplus-splitting rule.[9] It is because firms in the decentralized equilibrium charge constant markups that the private surplus from matching differs from the social

[9] See Brancaccio et al. (2023) for a detailed discussion of efficiency in random search models with heterogeneous agents.

surplus, and decentralized firms' search efforts do not coincide with the planner's.

In Appendix A.2, we decompose variation in unit values for the subset of transactions for which we observe them. The quantity of goods shipped is not required by the tax authority, so it is sometimes missing in the waybills. We find that most of the variation in unit values can be explained by supplier fixed effects, in line with a model of firms charging constant markups. Additionally, we find that alternative theories of surplus-splitting, which imply that markups should be correlated with quantity or bilateral market share, add little in terms of explaining variation. That is, there does not appear to be an alternative theory of surplus-splitting that fits the data significantly better.

In our quantitative analysis, we will find that these inefficiencies lead to quantitatively large steady-state aggregate productivity losses. In addition, they turn out to be central for understanding how technology differences map to productivity differences. Most of the difference in productivity comes from a difference in allocative efficiency (distance to efficient frontier), rather than technical efficiency (position of efficient frontier).

4 Calibration

In this section, we estimate structural parameters of the model to match moments of the Indian data and compare lifecycle patterns generated by the model against the data.

4.1 Calibration

We estimate structural parameters of the model to match moments of the Indian data.

Structural Parameters. We first impose some functional forms. We assume a standard form for search costs:

$$\varphi_g(v) = \frac{\xi}{\zeta} v^{\zeta}, \quad \varphi_h(u) = \frac{\xi}{\zeta} u^{\zeta}$$

The level parameter ξ governs the level of costs, while the curvature parameter ζ governs how easily firms can scale their customer and supplier networks. We impose that the level and curvature parameters are the same for both customer and supplier search effort.

We assume that the aggregate matching function is Cobb-Douglas, where γ governs the elasticity of matches with respect to customer search.

$$\mathcal{M}(\mathcal{U}, \mathcal{V}) = \mathcal{U}^{\gamma} \mathcal{V}^{1-\gamma}$$

Finally, we specify the firm's productivity process. Upon entry, a firm draws a productivity of 1 with probability p_{low}, and draws a productivity of $\bar{z} > 1$ with probability $(1 - p_{low})$

$$F_e(z) = \begin{cases} 1 \text{ with prob. } p_{low} \\ \bar{z} \text{ with prob. } (1 - p_{low}) \end{cases}$$

In addition, in the quantitative version of our model, we allow for productivity to exogenously evolve over time at rate ρ:

$$z_{i,t} = \rho z_{i,t-1}$$

This allows the model to better quantitatively match the lifecycle patterns we document in Section 2. This does not cause any issue in the exposition of previous sections, as long as z is now taken to be *productivity upon entry* rather than *permanent productivity*.

This results in ten parameters that require estimation. The model parameters are listed in Table 1. Most of these parameters are common to many models and will be estimated using standard moments. The two which are less common are the level and curvature of search costs. We briefly discuss how these parameters affect equilibrium objects in our setting. This will inform our strategy for estimating them.

How do search cost parameters affect equilibrium objects? We first explore the role of the level parameter ξ. In Appendix B.2, we show that the level parameter only affects the level of equilibrium objects (e.g., firm sales, number of partners) and does not affect any cross-sectional or lifecycle patterns. Thus, in our calibration, we will use this parameter to normalize the equilibrium to a given level.

Next, we explore the role of the curvature parameter ζ. Taking the ratio of first order conditions (i.e. Equation 11) for two firms i and j, we can express their relative customer search efforts in Equation 13:

$$\frac{u_{i,t}}{u_{j,t}} = \left[\frac{\frac{\mu-1}{\mu} \sum_{s=0}^{\infty} \sum_{z',a'} ((1-\delta)\beta)^s \left(\frac{\mu c(G_{i,t+s})^{1-\alpha}}{z_{i,t+s}\tilde{c}(z',a'+s)} \right)^{1-\sigma} \frac{1-\alpha}{\mu} \tilde{s}(z', a'+s) \frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{U}} \frac{\tilde{v}(z',a')n(z',a')}{\mathcal{V}}}{\frac{\mu-1}{\mu} \sum_{s=0}^{\infty} \sum_{z',a'} ((1-\delta)\beta)^s \left(\frac{\mu c(G_{j,t+s})^{1-\alpha}}{z_{j,t+s}\tilde{c}(z',a'+s)} \right)^{1-\sigma} \frac{1-\alpha}{\mu} \tilde{s}(z', a'+s) \frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{U}} \frac{\tilde{v}(z',a')n(z',a')}{\mathcal{V}}} \right]^{\frac{1}{\zeta-1}} \tag{13}$$

The relative search efforts of firms i and j depends on the ratio of expected profits generated from matching with additional customers. Notice that the only differences in customer search effort across firms arise from differences in marginal cost. Firms that expect to have lower marginal costs today and in the future exert more search effort, as conditional on matching with a customer, they sell more on the intensive margin. The curvature of the cost function, ζ, governs the elasticity between search effort and marginal cost. When ζ is low, small differences in marginal costs lead to large differences in search effort. On the other hand, when ζ is high, all firms exert similar search effort, despite large differences in marginal costs.

This elasticity between search effort and marginal cost governs the correlation between the number of customers and sales. Thus, we will be able to use this correlation to calibrate the curvature parameter.

An advantage of this moment is that it has been reported for other countries and has been used similarly to calibrate search cost curvature in other papers. In particular, Arkolakis et al. (2023) use the same moment to calibrate a search cost curvature using data from Chile. Thus, we can use variation in this moment across countries to get plausible estimates of technological variation.

As firm-to-firm links arise out of search effort in our setting, differences in search technology lead to differences in firm dynamics and production network structure. In Section 5, we utilize variation in this moment to quantitatively study how differences in search technology generate differences in firm dynamics and aggregate productivity. We find that both lifecycle growth and aggregate productivity are higher under lower curvature of search costs, as search efforts respond more to differences in idiosyncratic productivity.

Estimation Procedure. We assign existing estimates to a subset of our parameters. We use the labor share from the 2019 Penn World Table to set $\alpha = 0.52$. We set $\sigma = 4.30$, referring to Baqaee et al. (2023). We set $\gamma = 0.5$, which is in line with Krolikowski and McCallum (2021).

The remaining parameters are jointly estimated to match moments in our firm-to-firm data. Though all parameters are jointly estimated, there exists an intuitive mapping between parameters and target moments. We set exogenous survival rate $\beta = 0.93$ to match the fraction of firms which are 11+ in age.[10] We set exogenous separation rate $\delta = 0.50$ to match the 1-year survival rate of links. As discussed above, we normalize ξ such that the average number of customers is equal to 1, and estimate ζ by targeting the correlation between the number of customers and sales.[11] This moment in our data is 0.33. We estimate the productivity of the high type $\bar{z}$ and probability of drawing the low type p_{low} by targeting the interquartile range and skewness of log sales. Finally, we estimate exogenous growth in idiosyncratic productivity, ρ, by targeting the lifecycle growth of sales.

We summarize our estimation results in Table 1. We match all moments well.

[10] We target the share of Age 11+ firms due to the issue described in Section 2.1.

[11] To be clear, in the model, firms have "measures" of partners rather than "numbers", though at times we imprecisely use the two terms interchangeably.

Table 1: Estimation Results

Parameter	Description	Value	Target Moment	Data	Model
α	labor share	0.52	Penn World Table 2019	-	-
σ	elasticity of substitution	4.30	Baqaee et al. (2023)	-	-
γ	matching function elasticity	0.50	Krolikowski and McCallum (2021)	-	-
β	exogenous firm survival rate	0.93	population share of 11+ Age Firms	0.46	0.46
δ	exogenous separation rate	0.49	1-year survival rate of links	0.47	0.47
$\bar{z}$	initial productivity of high type	1.63	IQR of log sales	2.89	2.89
p_{low}	probability of low type	0.57	skewness of log sales	0.28	0.28
ρ	exogenous productivity growth	1.0007	lifecycle growth of log sales	0.33	0.33
ξ	level of search costs	0.97	normalization (average NC = 1)	-	-
ζ	curvature of search costs	3.10	elasticity between NC and Sales	0.33	0.33

Note: We summarize results from our calibration strategy. We estimate values for structural parameters to jointly match moments of the Indian data. Column 1 lists structural parameters. Column 2 provides descriptions. Column 3 displays estimated parameter values. Column 4 lists target moments. Columns 5 and 6 display values of target moments in the data and model, respectively.

4.2 Model Validation: Lifecycle Patterns

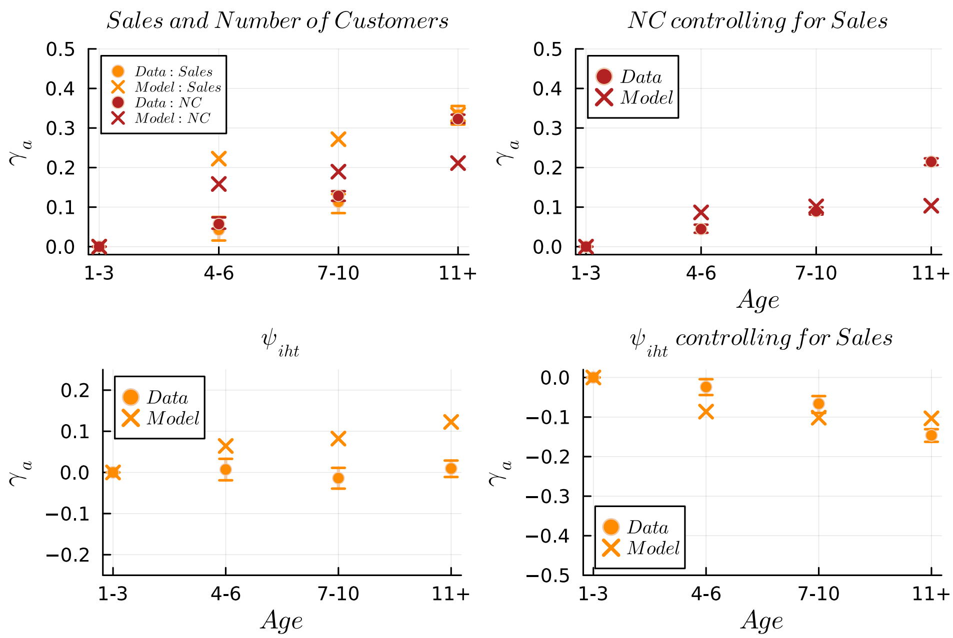
In Figure 7 and Figure 8 we plot lifecycle patterns from the calibrated model against the data. Figure 7 plots downstream patterns, while Figure 8 plots upstream patterns. Except for the lifecycle growth of sales, all patterns here are untargeted.

Qualitatively, the model output matches the data well. Older firms have more customers and suppliers, greater sales and intermediate expenditure, and lower input costs. Importantly, the model also matches patterns conditional on firm size. Conditioning on sales, older firms have more customers and charge higher output prices. Conditioning on sales, older firms have more suppliers and face lower input costs. As discussed in Section 2, it is difficult to generate these size-conditional patterns in a model featuring only dynamics in idiosyncratic productivity. In contrast, the frictional model presented here naturally rationalizes these patterns.

Quantitatively, the model is somewhat less successful, undershooting some patterns and overshooting others. However, given the relative simplicity of the model, we do not expect to fully capture the richness of the true data generating process.[12]

[12] In the data, we find no systematic variation of supplier fixed effects, ψ_{iht}, with age; while in the model, there is an increasing relationship. This deviation is not an inherent feature of the model. Under alternative calibrations, the model can generate flat, and even declining, relationships between age and ψ_{iht}.

Figure 7: Model Output (Downstream Patterns)



Note: We plot downstream patterns generated by the calibrated model against the data. The circle markers display the data moments documented in Section 2, while the cross markers display model moments. Other than the lifecycle growth of sales, all downstream patterns are untargeted in the calibration strategy.

5 Quantitative Analysis

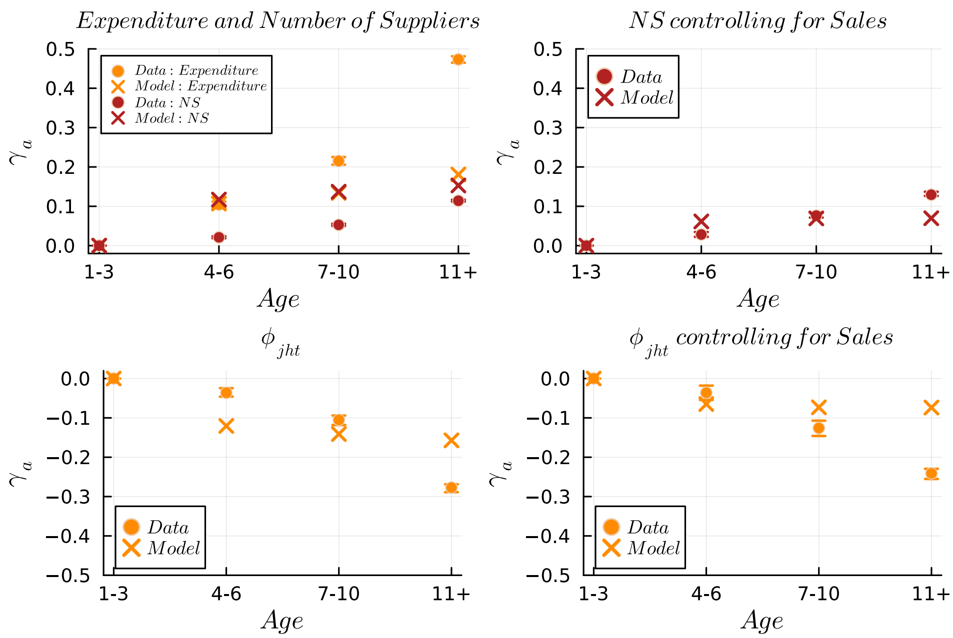
In this section, we use the calibrated model to study how search technology shapes firm dynamics and aggregate productivity.

5.1 Efficient Allocation

Having calibrated the model to the Indian data, we solve the constrained planner's problem introduced in Section 3.4 and discuss the efficient allocation below. We find large productivity losses from inefficient production and search choices. Aggregate productivity is 15% greater in the efficient allocation, relative to the decentralized equilibrium.

In Figure 9, we compare the efficient supplier and customer search efforts to the decentralized efforts. The two left panels display search efforts, while the right panels display the corresponding number of partners. As seen in the figures, all firms exert too much search effort. This is especially true for firms that draw low productivity upon entry ("low-productivity firms"). While search effort in the efficient allocation is roughly similar in magnitude for firms that draw high productivity upon entry ("high-productivity firms"), the planner exerts almost no search effort for low-productivity firms. As a result, while high-productivity firms

Figure 8: Model Output (Upstream Patterns)



Note: We plot upstream patterns generated by the calibrated model against the data. The circle markers display data moments documented in Section 2, while the cross markers display model moments. All upstream patterns are untargeted in the calibration strategy.

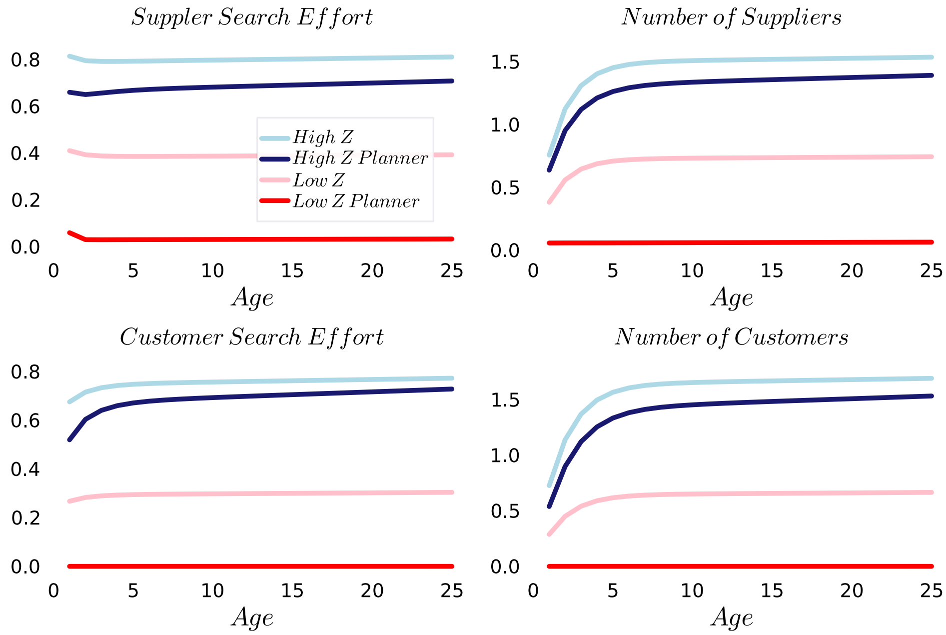
maintain similar numbers of partners, low-productivity firms have virtually no partners in the efficient allocation.

The efficient search efforts give rise to the efficient production network. In Figure 10, we compare the efficient network to the decentralized network. We plot the difference in the share of links between given supplier-customer pairs. The horizontal axis displays the type of the supplier, with the left half corresponding low-productivity suppliers of various ages, and the right half corresponding to high-productivity suppliers. The vertical axis displays the type of the customer, with the bottom half corresponding to low-productivity customers, and the top half corresponding to high-productivity customers. At each grid point in the figure, a positive value indicates that the efficient network has a greater share of links between these types, while a negative value indicates that the efficient network has a lesser share.

The efficient network features a greater share of links between high-productivity suppliers and high-productivity customers. These are the most productive matches in the economy. As decentralized search efforts depend on the private surplus from matching, rather than the social surplus, the decentralized production network features too few links between the most productive pairs.

We find large aggregate productivity differences between the efficient allocation and the

Figure 9: Efficient Search Efforts



Note: The two left panels display search efforts in the decentralized equilibrium and the efficient allocation. The right two panels plot the corresponding number of partners. "High Z" refers to firms which draw high productivity upon entry, while "Low Z" refers to firms which draw low productivity upon entry.

decentralized equilibrium. Final output (household consumption) is 15% greater in the efficient allocation. As the stock of labor is inelastic here, this gain entirely reflects an improvement in aggregate productivity stemming from the better coordination of search and production choices.

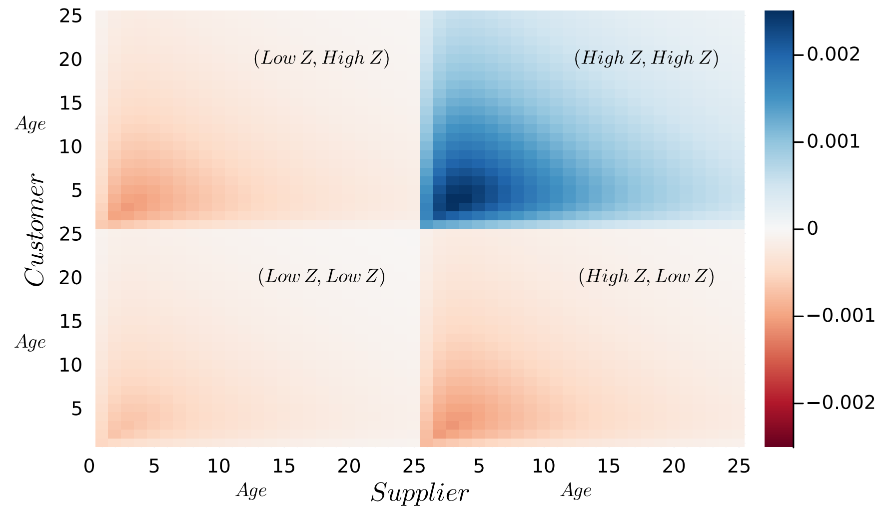
In order to understand the relative contributions of the double marginalization and search inefficiencies, we also solve a problem in which the planner takes the decentralized network as fixed, but can choose production choices for firms. Aggregate productivity increases by 5.2% in the solution to this "fixed network" planner's problem. Thus, most of the productivity gain in the full constrained planner's solution comes from correcting search inefficiencies.

5.2 Search Technology and Aggregate Productivity

Next, we study how differences in search technology map to differences in firm dynamics and aggregate productivity. The normative properties discussed in the previous section will be central for understanding how technology shapes productivity differences.

In our model, the search technology is governed by the parameters of the search cost. Thus, we study comparative statics with respect to these parameters.

Figure 10: Efficient Network vs. Decentralized Network



Note: We plot the difference between the decentralized production network and the efficient production network. The horizontal axis corresponds to suppliers of varying ages who enter with either low productivity (left half) or high productivity (right half). The vertical axis corresponds to customers of varying ages who enter with either low productivity (bottom half) or high productivity (top half). Each grid point displays the difference in the share of links between the corresponding types in the efficient allocation and the decentralized equilibrium. A positive value indicates that the efficient network has a greater share of links between the corresponding types, while a negative value indicates the opposite.

We begin by studying the role of the curvature of search costs ζ. As discussed in Section 4, ζ is disciplined by the elasticity between the number of customers and sales. In our data, we find this moment to be 0.33. This moment, however, is reported to be 0.42 in Arkolakis et al. (2023) for Chile.[13] We use the variation in this moment as a plausible estimate of technological variation. Through the lens of our model, these moments imply that Chilean firms are able to scale their customer and supplier networks more easily than Indian firms due to the lower curvature of search costs.

In a counterfactual exercise, we recalibrate ζ in our model to match the Chilean moment, holding the total measure of links fixed by adjusting ξ. All other parameters are held fixed. The counterfactual calibration is summarized in Table 2. The elasticity between the number of customers and sales implies $\zeta = 2.40$ in Chile, compared to $\zeta = 3.10$ in India.

In Figure 11, we compare the lifecycle trajectories of firms in the Chilean counterfactual to the lifecycle trajectories of firms under our baseline calibration. Under the lower curvature

[13] Again, similar to our moment, this is an average across sectors.

Table 2: Chile ζ Counterfactual

Parameter	Value: India	Value: Chile	Target Moment	Model: India	Model: Chile
α	0.52	0.52	Penn World Table 2019	-	-
σ	4.30	4.30	Baqaee et al. (2023)	-	-
γ	0.50	0.50	Krolikowski and McCallum (2021)	-	-
β	0.93	0.93	population share of 11+ Age firms	0.46	0.46
δ	0.49	0.49	1-year survival rate of links	0.47	0.47
z_{high}	1.63	1.63	IQR of log sales	2.89	3.61
p_{low}	0.57	0.57	skewness of log sales	0.28	0.28
ρ	1.0007	1.0007	lifecycle growth of log sales	0.33	0.37
ξ	0.97	0.58	normalization (average NC = 1)	-	-
ζ	3.10	2.40	elasticity between NC and sales	0.33	0.42

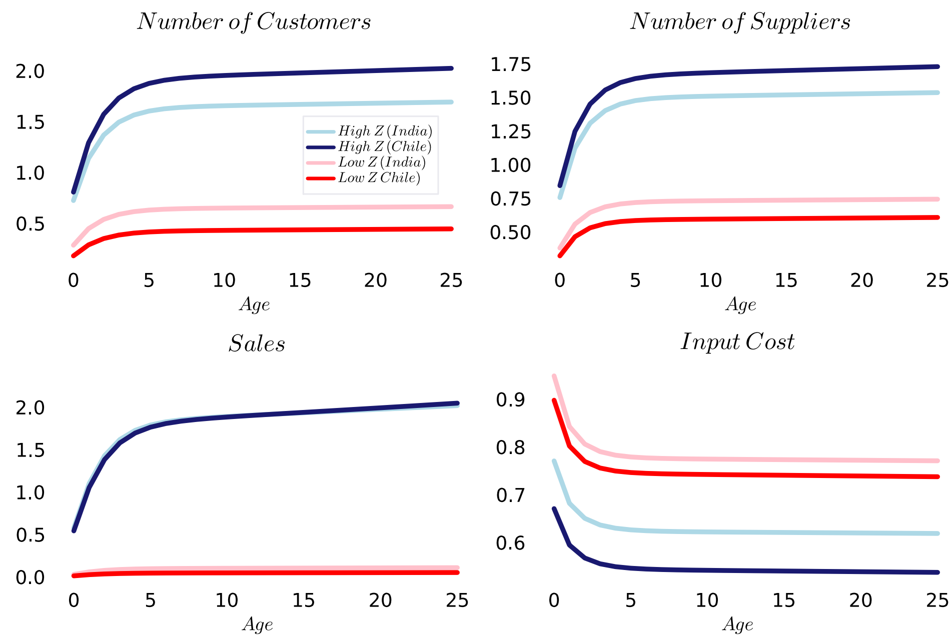
Note: We summarize our counterfactual calibration, in which we recalibrate ζ to match the Chilean moment. Column 1 lists structural parameters. Columns 2 and 3 display parameter values under the baseline Indian and counterfactual Chilean calibrations. Column 4 lists target moments. Columns 5 and 6 display values of the target moments under the baseline and counterfactual calibrations.

corresponding to Chile, firm search efforts respond more to idiosyncratic productivity differences. At the firm level, lifecycle growth of sales is 10% higher in the counterfactual economy. Notice, this is despite idiosyncratic productivity growth being identical in both economies. Moreover, high-productivity firms grow relatively faster compared to low-productivity firms. Lifecycle growth of sales is 13% higher for high-productivity firms in the counterfactual economy, compared to 7% higher for low-productivity firms.

As the production network arises from search effort, the response of firms to different technologies has implications for the production network. In Figure 12, we document how the structure of the production network in the counterfactual economy differs from the baseline. The figure plots the difference in the share of links between given supplier-customer pairs. The horizontal axis displays the type of the supplier, with the left half corresponding low-productivity suppliers of various ages, and the right half corresponding to high-productivity suppliers of various ages. The vertical axis displays the type of the customer, with the bottom half corresponding to low-productivity customers of various ages, and the top half corresponding to high-productivity customers of various ages. At each grid point in the figure, a positive value indicates that the counterfactual network has a greater share of links between these types, while a negative value indicates that the counterfactual network has a lesser share. Note that by construction, the total measure of links in each economy will be the same.

The counterfactual network has a greater share of links between high-productivity suppliers and high-productivity customers, and a lesser share between all other pairs. As search efforts respond more to idiosyncratic productivity differences under the Chilean technology, high-

Figure 11: Counterfactual Lifecycle Trajectories



Note: We plot lifecycle trajectories for firms in our baseline economy, and for firms in the counterfactual economy calibrated to match the Chilean moment.

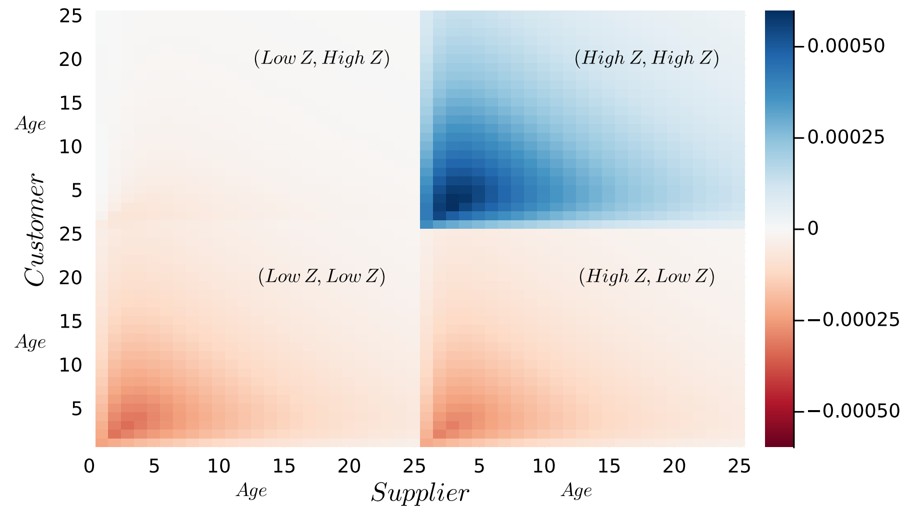
productivity firms grow faster relative to low-productivity firms and make up a greater share of links in the production network.

We find that under the Chilean technology, aggregate productivity rises by 2.1%. As the total number of links is held fixed, this difference is due to the production network being more concentrated in links between high-productivity firms.

Much attention has been paid to understanding differences in firm dynamics and aggregate productivity across countries (Hsieh and Klenow, 2014; Bento and Restuccia, 2017; Akcigit et al., 2021). Our results suggest that firm-to-firm matching technology may be an important driver of these differences. Slow growth of firms in developing countries may be, in part, due to stronger search and matching frictions which make scaling customer and supplier networks more difficult. These frictions, in turn, reduce aggregate productivity by reducing the prevalence of high-productivity firms in the production network.

Importantly, as the decentralized economy is inefficient, differences in search technology affect aggregate productivity by affecting both technical efficiency (productivity in efficient allocation) and allocative efficiency (distance of decentralized allocation from efficient allocation). In Table 3, we compare aggregate productivity under both the baseline Indian technology and the counterfactual Chilean technology, for both decentralized and efficient

Figure 12: Production Network: Chilean Technology vs. Indian Technology



Note: We plot the difference between the production network in the baseline economy and the network in the counterfactual economy calibrated to match the Chilean moment. The horizontal axis corresponds to low-productivity (left half) and high-productivity (right half) suppliers of varying ages. The vertical axis corresponds to low-productivity (bottom half) and high-productivity (top half) customers of varying ages. Each grid point displays the difference between the share of links between the corresponding types in the counterfactual economy and the baseline economy. A positive value indicates that the counterfactual economy has a greater share of links between the corresponding types, while a negative value indicates that the counterfactual economy has a lesser share.

allocations. We normalize by aggregate productivity in the Indian decentralized allocation. The second column displays productivities in the decentralized allocations, while the third column displays productivities in the efficient allocations. The last column computes the productivity difference between the efficient and decentralized allocations for each technology. The last row computes the productivity difference between the two decentralized allocations and the two efficient allocations.

We find that productivity differences between the efficient allocations are much smaller than those between the decentralized allocations. While aggregate productivity in the Chilean decentralized allocation is 2.1% greater than the Indian decentralized allocation, aggregate productivity in the Chilean efficient allocation is only 0.3% greater than in the Indian efficient allocation. Decomposing the productivity gain realized in our counterfactual, roughly 85% comes from an improvement in allocative efficiency, while 15% comes from an improvement in technical efficiency. In other words, though the efficient frontier does not shift significantly under the Chilean technology, firms in the decentralized economy are much closer to the efficient frontier. Aggregate productivity in the efficient allocation is 13% greater than in the decentralized allocation under the Chilean technology, compared to 15% under the Indian technology.[14]

Table 3: Technical vs. Allocative Efficiency

Technology	Decentralized Productivity	Efficient Productivity	% Difference
India	100.0	115.3	15.3
Chile	102.1	115.7	13.3
% Difference	2.1	0.3	

Note: This table compares aggregate productivity under the baseline Indian and counterfactual Chilean technologies. The first column displays productivity in the decentralized allocations under both technologies. The second column displays productivity in the efficient allocations under both technologies. We normalize by productivity in the Indian decentralized allocation. The last column displays the productivity difference between the efficient and decentralized economies for each technology. The last row displays the productivity difference between the two decentralized allocations and the two efficient allocations.

These results suggest that much of the gains from improving firm-to-firm matching technology can be realized through policy intervention. This is useful, given that improving firm-to-firm matching technology, for example, through improving information technology or legal institutions, appears to be prohibitively costly. In particular, our results imply that subsidizing search for high-productivity firms can generate significant aggregate productivity gains. Conversely, policies that subsidize search in an untargeted manner or those that subsidize search for low-productivity firms can lead to significant losses in productivity.

Finally, we study the role of the level of search costs, ξ. In Appendix B.2, we show that the elasticity of final output with respect to the level of search costs depends only on labor share, α, elasticity of substitution, σ, and curvature of search costs, ζ (holding all other parameters fixed); and is decreasing in all three. When ζ is greater, search efforts are less elastic and so respond less to changes in the level of search costs. When α is greater, intermediates are less important in production. As a result, gains from variety generated by reductions in the level of search costs matter less for final output. Finally, when σ is higher, the change in final output is lower due to lower gains from variety. Under our baseline calibration, a 10% reduction in the level of search costs leads to a 0.9% increase in aggregate productivity.

[14] As described in Section 5.1 we also solve a "fixed network" planner's problem, in which the planner takes the decentralized network as fixed, but chooses production choices for firms. We find that aggregate productivity rises by 5.3% in the "fixed network" solution under the Chilean technology, as compared to 5.2% under the Indian technology.

6 Conclusion

We study the role of firm-to-firm matching in shaping firm dynamics and aggregate productivity. Using transaction-level data from a large Indian state, we document lifecycle patterns of customer and supplier networks. The patterns support a theory of firm dynamics in which firms grow through matching with trading partners. Motivated by these patterns, we develop a model of endogenous network formation where heterogeneous firms slowly match with partners over time using a random search technology. Firms in the decentralized equilibrium search inefficiently due to standard search externalities. This inefficiency turns out to be central for understanding how search technology shapes aggregate productivity. Calibrating the model to the Indian data, we find that most of the gains from technology improvement come from improvement in allocative efficiency, rather than in technical efficiency.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333.

Acemoglu, D. and P. D. Azar (2020). Endogenous production networks. *Econometrica 88*(1), 33–82.

Afrouzi, H., A. Drenik, and R. Kim (2023). Concentration, market power, and misallocation: The role of endogenous customer acquisition.

Akcigit, U., H. Alp, and M. Peters (2021). Lack of selection and limits to delegation: firm dynamics in developing countries. *American Economic Review 111*(1), 231–275.

Aker, J. C. (2010). Information from markets near and far: Mobile phones and agricultural markets in niger. *American Economic Journal: Applied Economics 2*(3), 46–59.

Argente, D., D. Fitzgerald, S. Moreira, and A. Priolo (2023). How do firms build market share?

Arkolakis, C. (2010). Market penetration costs and the new consumers margin in international trade. *Journal of Political Economy 118*(6), 1151–1199.

Arkolakis, C., F. Huneeus, and Y. Miyauchi (2023). Spatial production networks.

Atalay, E., A. Hortacsu, J. Roberts, and C. Syverson (2011). Network structure of production. *Proceedings of the National Academy of Sciences 108*(13), 5199–5202.

Atkeson, A. and A. Burstein (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review 98*(5), 1998–2031.

Baqaee, D., A. Burstein, C. Duprez, and E. Farhi (2023). Supplier churn and growth: A micro-to-macro analysis.

Baqaee, D. R. and E. Farhi (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics 135*(1), 105–163.

Bento, P. and D. Restuccia (2017). Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics 9*(3), 267–303.

Bergquist, L. F., C. McIntosh, and M. Startz (2024). Search costs, intermediation, and trade: Experimental evidence from ugandan agricultural markets. Technical report, National Bureau of Economic Research.

Bernard, A. B., E. Dhyne, G. Magerman, K. Manova, and A. Moxnes (2022). The origins of firm heterogeneity: A production network approach. *Journal of Political Economy 130*(7), 1765–1804.

Bernard, A. B., A. Moxnes, and Y. U. Saito (2019). Production networks, geography, and firm performance. *Journal of Political Economy 127*(2), 639–688.

Boehm, J. (2022). The impact of contract enforcement costs on value chains and aggregate productivity. *Review of Economics and Statistics 104*(1), 34–50.

Boehm, J. and E. Oberfield (2020). Misallocation in the market for inputs: Enforcement and the organization of production. *The Quarterly Journal of Economics 135*(4), 2007–2058.

Boehm, J., R. South, E. Oberfield, and M. Waseem (2024). The network origins of firm dynamics: Contracting frictions and dynamism with long-term relationships.

Brancaccio, G., M. Kalouptsidi, and T. Papageorgiou (2020). Geography, transportation, and endogenous trade costs. *Econometrica 88*(2), 657–691.

Brancaccio, G., M. Kalouptsidi, T. Papageorgiou, and N. Rosaia (2023). Search frictions and efficiency in decentralized transport markets. *The Quarterly Journal of Economics 138*(4), 2451–2503.

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly journal of economics 128*(3), 967–1015.

Cavenaile, L., M. A. Celik, J. Perla, and P. Roldan-Blanco (2023). A model of product awareness and industry life cycles. Technical report.

Chaney, T. (2014). The network structure of international trade. *American Economic Review 104*(11), 3600–3634.

Cole, H. L., J. Greenwood, and J. M. Sanchez (2016). Why doesn't technology flow from rich to poor countries? *Econometrica 84*(4), 1477–1521.

Demir, B., A. C. Fieler, D. Y. Xu, and K. K. Yang (2023). O-ring production networks. *Journal of Political Economy, forthcoming*.

Dhyne, E., A. K. Kikkawa, M. Mogstad, and F. Tintelnot (2021). Trade and domestic production networks. *The Review of Economic Studies 88*(2), 643–668.

Dinlersoz, E. M. and M. Yorukoglu (2012). Information and industry dynamics. *American Economic Review 102*(2), 884–913.

Drozd, L. A. and J. B. Nosal (2012). Understanding international prices: Customers as capital. *American Economic Review 102*(1), 364–395.

Eaton, J., D. Jinkins, J. R. Tybout, and D. Xu (2022). Two-sided search in international markets.

Eaton, J., S. S. Kortum, and F. Kramarz (2022). Firm-to-firm trade: Imports, exports, and the labor market. Technical report, National Bureau of Economic Research.

Einav, L., P. J. Klenow, J. D. Levin, and R. Murciano-Goroff (2022). Customers and retail growth.

Fitzgerald, D., S. Haller, and Y. Yedid-Levi (2023). How exporters grow. *Review of Economic Studies, forthcoming*.

Foster, L., J. Haltiwanger, and C. Syverson (2016). The slow growth of new plants: Learning about demand? *Economica 83*(329), 91–129.

Gourio, F. and L. Rudanko (2014). Customer capital. *Review of Economic Studies 81*(3), 1102–1136.

Goyal, A. (2010). Information, direct access to farmers, and rural market performance in central india. *American Economic Journal: Applied Economics 2*(3), 22–45.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, 1127–1150.

Hottman, C. J., S. J. Redding, and D. E. Weinstein (2016). Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics 131*(3), 1291–1364.

Hsieh, C.-T. and P. J. Klenow (2014). The life cycle of plants in india and mexico. *The Quarterly Journal of Economics 129*(3), 1035–1084.

Huneeus, F. (2020). Production network dynamics and the propagation of shocks.

Jensen, R. (2007). The digital provide: Information (technology), market performance, and welfare in the south indian fisheries sector. *The quarterly journal of economics 122*(3), 879–924.

Kimball, M. S. (1995). The quantitative analytics of the basic neomonetarist model.

Kopytov, A., B. Mishra, K. Nimark, and M. Taschereau-Dumouchel (2023). Endogenous production networks under supply chain uncertainty.

Krolikowski, P. M. and A. H. McCallum (2021). Goods-market frictions and international trade. *Journal of International Economics 129*, 103411.

Lim, K. (2018). Endogenous production networks and the business cycle.

Luttmer, E. G. J. (2006). *Consumer search and firm growth.*

Miyauchi, Y. (2024). Matching and agglomeration: Theory and evidence from japanese firm-to-firm trade. *Econometrica 92*(6), 1869–1905.

Oberfield, E. (2018). A theory of input–output architecture. *Econometrica 86*(2), 559–589.

Petrongolo, B. and C. A. Pissarides (2001). Looking into the black box: A survey of the matching function. *Journal of Economic literature 39*(2), 390–431.

Taschereau-Dumouchel, M. (2020). Cascades and fluctuations in an economy with an endogenous production network. *Available at SSRN 3115854*.

A Empirical Appendix

A.1 Bias of Coefficients under Endogenous Mobility

In estimating firm input costs and output prices, we estimate the following equation:

$$e_{i,j,h,t} = \psi_{i,h,t} + \phi_{j,h,t} + \varepsilon_{i,j,h,t}$$

For the estimates of $\psi_{i,h,t}$ and $\phi_{j,h,t}$ to be unbiased, we require:

$$\mathbb{E}\left[s'_{i,h,t}\varepsilon_{i,j,h,t}\right] = 0 \quad \forall i,h,t$$

$$\mathbb{E}\left[d'_{j,h,t}\varepsilon_{i,j,h,t}\right] = 0 \quad \forall j,h,t \ ,$$

where S is the seller fixed effects design matrix and D is the customer fixed effects design matrix. These conditions are also known in the labor literature as the assumption of exogenous mobility (Abowd et al. (1999), Card et al. (2013)). In this context, the conditions imply that for every supplier, the average match-specific effect, $\varepsilon_{i,j,h,t}$, (across customers) is zero; and that for every customer, the average match-specific effect (across suppliers) is zero. This requires firms to match in a manner uncorrelated with the match-specific effects.

Suppose customers take into account both the supplier fixed effect and match-specific effect when matching with a supplier. In this case, customers match with suppliers that have high supplier effects or those that have high match-specific effects for them. This means that a supplier with a lower supplier effect requires a higher match-specific effect to match with a customer. This selection implies that OLS estimates of low supplier effects will be biased upwards.

Similarly, suppose suppliers take into account both the customer fixed effect and match-specific effect when matching with a supplier. In this case, suppliers match with customers which have high customer effects or those who have high match-specific effects for them. This means that a customer with a lower customer effect requires a higher match-specific effect to match with a supplier. This selection implies that OLS estimates of low customer effects will be biased upwards.

Selection on match-specific effects induces an upward bias on OLS estimates of low fixed effects, so that the estimates attenuate the true disparity in customer and supplier effects. Thus, under "endogenous mobility", our estimates underestimate the true differences in input costs and output prices between young and old firms.

A.2 Variation in Unit Values

In this section, we study variation in unit values. The waybills which comprise our transaction data also have entries for units[15] and quantities. This information is not required by the tax authority, so it is sometimes missing in the waybills.

Let $R_{i,j,hsn,u,y,m}$ denote the value of sales between supplier i and customer j within HSN hsn and unit $unit$ in year y and month m. Let $Q_{i,j,hsn,unit,y,m}$ denote the corresponding quantity of good shipped. Different from our main analysis, here we let HSN hsn refer to an 8-digit category. As we are interested in understanding the extent to which suppliers charge different prices for the same product, we study variation at the most granular product level available to us. This comes at the cost of losing 60% of our sample, as the Tax Authority only requires reporting up to the 4-digit level.

Define the log unit value supplier i charges customer j in an HSN-unit-year-month as:

$$p_{i,j,hsn,unit,y,m} \equiv log\left(\frac{R_{i,j,hsn,unit,y,m}}{Q_{i,j,hsn,unit,y,m}}\right)$$

As we are interested in studying variation in unit values within product, we compute partial R^2 of various "full" models relative to a "reduced" model specified in Equation 14.[16] The reduced model includes only product-unit-year-month effects:

$$p_{i,j,hsn,unit,y,m} = \chi_{hsn,unit,y,m} + \varepsilon_{i,j,hsn,unit,y,m} \tag{14}$$

In the "full" models, we also include supplier-HSN-unit-year-month fixed effects, $\psi_{i,hsn,unit,y,m}$, and covariates, $X_{i,j,hsn,unit,y,m}$:

$$p_{i,j,hsn,unit,y,m} = \psi_{i,hsn,unit,y,m} + \beta X_{i,j,hsn,unit,y,m} + \chi_{hsn,unit,y,m} + \varepsilon_{i,j,hsn,unit,y,m} \tag{15}$$

In Table 4, we display the partial R^2 from the reduced models. The first row displays the partial R^2 when Equation 15 includes only supplier fixed effects. We find that supplier fixed effects alone explain roughly 3/4 of the variation within $hsn, unit, y, m$. In other words, variation within suppliers across customers plays a small role in explaining the total variation in unit values. We take this as support for our assumption in the paper that firms charge the same price to all of their customers. In the second row, we show that including

[15] Examples of units include "boxes", "kilograms", and "bales".

[16] Partial R^2 is defined as $1 - \frac{SSE_{full}}{SSE_{reduced}}$ where SSE_{full} and $SSE_{reduced}$ are the sum of squared residuals of the full and reduced model, respectively. The partial R^2 gives the proportion of variation explained by the full model that cannot be explained by the reduced model.

buyer-state-seller-state effects slightly improves the variation explained by the specification.

We also test specifications motivated by alternative theories of pricing. In the third row, motivated by models in which sellers offer nonlinear pricing schedules to customers, we show results from a specification which includes log quantity in addition to seller effects and buyer-state-seller-state effects. In line with the literature, we find that firms offer quantity discounts. We find an estimate of $\hat{\beta} = -0.12$, implying that a 8% higher quantity is associated with a 1% lower unit value. However, the variation explained by the specification rises only slightly.

Finally, in the fourth row, motivated by models in which sellers with higher market shares charge higher markups (e.g. Kimball (1995), Atkeson and Burstein (2008)), we include log bilateral market share.[17] We find a slightly positive coefficient $\hat{\beta} = 0.003$, implying that a 10% increase in bilateral market share is associated with a 0.03% increase in unit value. The specification including bilateral market share does not perform better than the specification including only seller and location fixed effects.

Table 4: Explaining Variation in Unit Values

Specification	Partial R^2	Observations
$\psi_{i,hsn,unit,y,m}$	0.77	2,674,395
$\psi_{i,hsn,unit,y,m} + \iota_{loc(i),loc(j),hsn,unit,y,m}$	0.78	2,674,395
$\psi_{i,hsn,unit,y,m} + \beta q_{i,j,hsn,unit,y,m} + \iota_{loc(i),loc(j),hsn,unit,y,m}$	0.80	2,674,395
$\psi_{i,hsn,unit,y,m} + \beta e_{i,j,hsn,unit,y,m} + \iota_{loc(i),loc(j),hsn,unit,y,m}$	0.78	2,674,395

Note: We display partial R^2 of various specifications relative to the reduced model specified in Equation 14.

B Theoretical Appendix

B.1 Solution Algorithm

In this section, I describe our solution algorithm. A stationary equilibrium is a solution to 6 systems of equations. Equations 16 and 17 are fixed points in payoff-relevant attributes of firms. Payoff-relevant attributes depend on the structure of the network, as seen through $\{H_{z,a}\}$ and $\{G_{z,a}\}$ entering these equations. Equations 18 and 19 are the first order optimality conditions of customer search effort $u(z, a)$ and supplier search effort $v(z, a)$. Optimal search

[17] We define bilateral market share, $e_{i,j,hsn,unit,y,m}$, as the share of HSN hsn expenditure in period y, m which customer j spends on supplier i.

efforts of a (z, a) firm depend on its own payoff-relevant attributes, $c(z, a)$ and $s(z, a)$, and those of other firms, $c(z', a')$ and $s(z', a')$. Equations 20 and 21 are laws of motion for customer and supplier networks implied by the search technology. They depend on the optimal search efforts of firms $u(z, a)$ and $v(z, a)$.

$$s(z,a) = \left(\frac{\mu c(z,a)^{1-\alpha}}{z\mathcal{P}}\right)^{1-\sigma} E + \int_{Z,A} \left(\frac{\mu c(z,a)^{1-\alpha}}{zc(z',a')}\right)^{1-\sigma} \frac{1-\alpha}{\mu} s(z',a') H_{z,a}(z',a') \tag{16}$$

$$c(z,a) = \left(\int_{Z,A} \left(\frac{\mu c(z',a')^{1-\alpha}}{z'}\right)^{1-\sigma} G_{z,a}(z',a')\right)^{1/(1-\sigma)} \tag{17}$$

$$\frac{\mu-1}{\mu} \sum_{\tau=0}^{\infty} \sum_{z',a'} ((1-\delta)\beta)^{\tau} \left(\frac{\mu c(z,a+\tau)^{1-\alpha}}{zc(z',a'+\tau)}\right)^{1-\sigma} \frac{1-\alpha}{\mu} s(z',a'+\tau) \frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{U}} \frac{v(z',a')n(z',a')}{\mathcal{V}} = \frac{\partial \varphi_h(u(z,a))}{\partial u} \tag{18}$$

$$\frac{\mu-1}{\mu} \sum_{\tau=0}^{\infty} \sum_{z',a'} ((1-\delta)\beta)^{\tau} \left(\frac{\mu c(z',a'+\tau)^{1-\alpha}}{z'c(z,a+\tau)}\right)^{1-\sigma} (1-\alpha) s(z,a+\tau) \frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{V}} \frac{u(z',a')n(z',a')}{\mathcal{U}} = \frac{\partial \varphi_g(v(z,a))}{\partial v} \tag{19}$$

$$H_{z,a}(z',a') = u(z,a) \frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{U}} \frac{v(z',a')n(z',a')}{\mathcal{V}} + (1-\delta) H_{z,a-1}(z',a'-1) \tag{20}$$

$$G_{z,a}(z',a') = v(z,a) \frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{V}} \frac{u(z',a')n(z',a')}{\mathcal{U}} + (1-\delta) G_{z,a-1}(z',a'-1) \tag{21}$$

B.2 How does the level of search costs affect equilibrium objects?

In this section, we discuss how the level of search costs, ξ, affects equilibrium objects. Proposition 1 states that changing the level parameter from ξ_0 to ξ_1 uniformly scales equilibrium search efforts of firms by a factor $\left(\frac{\xi_1}{\xi_0}\right)^{-1/\zeta}$. As a result, the equilibrium sales, input costs, number of customers, and number of suppliers all scale by constant factors. However, as all firms scale by the same factor, there are no changes in patterns across the lifecycle nor across the cross-section of firms. In our calibration, we will use the level parameter to normalize the equilibrium to a given level, understanding that we can use Proposition 1 to calculate how the level of equilibrium objects would differ for alternative values of ξ.

Proposition 1 *Suppose search policies in an equilibrium with search cost $\varphi_0(x) = \frac{\xi_0}{\zeta} x^{\zeta}$ are given by:*

$$u(z, H_{-1}, G_{-1}), \quad v(z, H_{-1}, G_{-1})$$

In an equilibrium with cost function $\varphi_1(x) = \frac{\xi_1}{\zeta} x^\zeta$ (holding all other parameters fixed), equilibrium search policies are given given by:

$$\left(\frac{\xi_1}{\xi_0}\right)^{-1/\zeta} u(z, H_{-1}, G_{-1}), \quad \left(\frac{\xi_1}{\xi_0}\right)^{-1/\zeta} v(z, H_{-1}, G_{-1})$$

As a corollary of Proposition 1, we derive an expression that relates differences in ξ to differences in aggregate output.

Corollary 1 *Suppose final output in an equilibrium with search cost $\varphi_0(x) = \frac{\xi_0}{\zeta} x^\zeta$ is given by Y_0. Final output in an equilibrium with search cost $\varphi_1(x) = \frac{\xi_1}{\zeta} x^\zeta$ (holding all other parameters fixed) is given by:*

$$\frac{Y_1}{Y_0} = \left(\frac{\xi_1}{\xi_0}\right)^{-\frac{1-\alpha}{\zeta\alpha(\sigma-1)}}$$

Corollary 1 states that the elasticity of final output with respect to the level of search costs (holding all other parameters fixed) depends on only the labor share α, elasticity of substitution σ, and curvature of search costs ζ; and is decreasing in all three. When ζ is greater, search efforts are less elastic and so respond less to changes in the level of search costs. When α is greater, intermediates are less important in production. As a result, gains from variety generated by changes in ξ matter less for final output. Finally, when σ is higher, the change in final output is lower due to lowers gains from variety.

B.3 Constrained Inefficiency

In this section, we study the normative properties of our environment. We find that the decentralized equilibrium is inefficient as firms fail to internalize vertical externalities and search externalities. First, we describe the problem of a planner who chooses production choices for firms, taking the production network as given. Next, we describe the full problem of a planner who chooses both production and search choices for firms, imposing that the planner must use the same search technology as firms in the decentralized economy. Proceeding in this manner helps distinguish inefficiencies that arise from production choices versus search choices. We find that inefficient production and search choices generate quantitatively large aggregate losses.

B.3.1 Efficiency of Fixed Network Equilibrium

We first study the problem of a planner who chooses production choices for firms, taking the production network as given. The "fixed network" can be described as a collection of customer and supplier sets $\{H_{z,a}, G_{z,a}\}_{z,a}$. The problem of the planner is to choose consumption of varieties, $\nu^f(z,a)$, labor input of firms, $l(z,a)$, and intermediate inputs of firms, $\nu(z,a,z',a')$, to maximize aggregate consumption. Here, $\nu(z,a,z',a')$, denotes the quantity of a (z',a') variety a (z,a) firm uses. The planner's choices are subject to firm output and aggregate labor constraints.

$$\max_{\nu^f,l,\nu} \left(\sum_{z,a} \nu^f(z,a)^{\frac{\sigma-1}{\sigma}} n(z,a) \right)^{\frac{\sigma}{\sigma-1}}$$

$$\text{s.t.} \qquad \nu^f(z,a) + \sum_{z',a'} \nu(z',a',z,a) H_{z,a}(z',a') \le y(z,a) \quad \forall z,a$$

$$\sum l(z,a) n(z,a) \le L$$

$$y(z,a) = \kappa z l(z,a)^{\alpha} \left(\sum_{z',a'} \nu(z,a,z',a')^{(\sigma-1)/\sigma} G_{z,a}(z',a') \right)^{\frac{\sigma(1-\alpha)}{\sigma-1}}$$

$$\{H_{z,a}, G_{z,a}\}_{z,a} \quad \text{Exogenously Given}$$

The first constraint imposes that the sum of final consumption of a variety and usage of it by other firms in production must be less than the output of that variety. The second constraint imposes that aggregate labor input must be less than aggregate labor supply.

We find that the decentralized choices for consumption and labor input match the planner's choices, however intermediate input usage does not. In Equation 22, we express the decentralized firm's input demand. In Equation 23, we express the choice which satisfies the planner's first order condition.

$$\nu(z,a,z',a') = \left(\frac{\mu c(z',a')^{1-\alpha}}{z'} \right)^{-\sigma} c(z,a)^{\sigma-1} \frac{1-\alpha}{\mu} s(z,a) \tag{22}$$

$$\nu(z,a,z',a') = \mu^{\sigma} \left(\frac{\mu c(z',a')^{1-\alpha}}{z'} \right)^{-\sigma} c(z,a)^{\sigma-1} \frac{1-\alpha}{\mu} s(z,a) \tag{23}$$

Comparing the planner's choice to the decentralized choice, we see that the two diverge. Firms in the decentralized equilibrium underutilize intermediate inputs relative to the planner's allocation. The reason for this inefficiency is that firms fail to internalize a "vertical externality" and set prices inefficiently high (double marginalization). The price a supplier charges its customer affects the marginal cost and so the profits of the customer. Similarly, the price the customer charges its own customers affects demand for its variety, which in turn, affects its demand for inputs and so the profits of the supplier. As both customer and supplier in the decentralized equilibrium choose prices to maximize private profits rather than joint profits, they set prices too high and use too little intermediate input relative to the planner's solution.

It is useful to compare the setting here to a similar setting that lacks intermediate input usage. Suppose firms still sell to the household, but no longer sell to other firms (nor use intermediate inputs). If labor supply is inelastic, then this modified setting is efficient. Constant markups preclude distortions between firms. The inelasticity of labor supply precludes underutilization of labor. In contrast, in the fixed network setting, firms underproduce due to the underutilization of intermediate inputs. The point is that market power is not enough in this setting to generate inefficiency. This inefficiency requires both market power and intermediation chains.

B.3.2 Efficiency of Endogenous Network Equilibrium

We next study the efficiency of an endogenous network equilibrium. As in the fixed network problem, the planner chooses consumption of varieties $\nu^f(z,a)$, labor input $l(z,a)$, and intermediate inputs $\nu(z,a,z',a')$, to maximize aggregate consumption. In addition, the planner also chooses customer search effort $u(z,a)$, and supplier search effort $v(z,a)$. The production network then arises out of the same search technology available to decentralized firms. Again, we can describe the network as a collection of customer and supplier sets $\{H_{z,a}, G_{z,a}\}_{z,a}$. However, here, these sets are endogenous objects which arise out of the planner's search efforts.

$$\max_{\nu^f,l,\nu,u,v} \left(\sum_{z,a} \nu^f(z,a)^{\frac{\sigma-1}{\sigma}} n(z,a) \right)^{\frac{\sigma}{\sigma-1}}$$

$$\nu^f(z,a) + \sum_{z',a'} \nu_{z',a'}(z,a) H_{z,a}(z',a') \leq y(z,a) \quad \forall z,a$$

$$\sum_{z,a} l(z,a)n(z,a) + \sum_{z,a} \varphi_h(u(z,a))n(z,a) + \sum_{z,a} \varphi_g(v(z,a))n(z,a) \leq L$$

$$y(z,a) = \kappa z l(z,a)^\alpha \left(\sum_{z',a'} \nu(z,a,z',a')^{(\sigma-1)/\sigma} G_{z,a}(z',a') \right)^{\frac{\sigma(1-\alpha)}{\sigma-1}}$$

$$H_{z,a}(z',a') = u(z,a) \frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{U}} \frac{v(z',a')n(z',a')}{\mathcal{V}} + (1-\delta) H_{z,a-1}(z',a'-1) \quad \forall z,a,z',a'$$

$$G_{z,a}(z',a') = v(z,a) \frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{V}} \frac{u(z',a')n(z',a')}{\mathcal{U}} + (1-\delta) G_{z,a-1}(z',a'-1) \quad \forall z,a,z',a'$$

The first constraint imposes that the sum of the final consumption of a variety and its usage by other firms in production must be less than the output of that variety. The second constraint imposes that aggregate labor input must be less than aggregate labor supply. Different from the fixed network problem, however, here aggregate labor input is the sum of labor used for production and labor used for search. The third constraint is just the production function of the firm. The last two constraints impose that the planner uses the same search technology as firms in the decentralized equilibrium.

The planner's first order conditions with respect to consumption, labor input, and intermediate input usage are identical to the fixed network problem, so we omit further discussion, and instead focus on the first order conditions with respect to customer and supplier search. Let $\lambda(z,a)$ denote the shadow value of a (z,a) variety and W denote the shadow wage. The planner's choice for customer search effort $u(\hat{z},\hat{a})$ is given by:

$$\sum_{z,a}\sum_{s=0}^{\infty}(1-\delta)^s\left(\lambda(z,a+s)\frac{\partial y(z,a+s)}{\partial G_{z,a+s}(\hat{z},\hat{a}+s)}v(z,a)n(\hat{z},\hat{a})-\lambda(\hat{z},\hat{a}+s)\nu(z,a+s,\hat{z},\hat{a}+s)v(z,a)n(z,a)\right)\frac{\mathcal{M}\left(\mathcal{U},\mathcal{V}\right)}{\mathcal{U}\mathcal{V}}$$
$$+\frac{\partial\frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{U}}}{\partial u(\hat{z},\hat{a})}\frac{1}{\frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{U}}}\left[\Gamma_1-\Gamma_2\right]=W\frac{\partial\varphi_h(u(\hat{z},\hat{a}))}{\partial u(\hat{z},\hat{a})}n(\hat{z},\hat{a})$$

$$\Gamma_1=\sum_{z,a}\sum_{z',a'}\lambda(z,a)\frac{\partial y(z,a)}{\partial G_{z,a}(z',a')}G_{z,a}(z',a')$$

$$\Gamma_2=\sum_{z,a}\sum_{z',a'}\lambda(z,a)\nu(z',a',z,a)H_{z,a}(z',a')$$

The first term on the left is the marginal surplus generated by a $(\hat{z},\hat{a})$ firm connecting with more customers. This term is equal to the increase in customers' output due to adding a $(\hat{z},\hat{a})$ supplier, multiplied by the value of that output, minus the value of $(\hat{z},\hat{a})$ output used in production by the customers. The second term on the left is the marginal effect firm $(\hat{z},\hat{a})$ has on all other matches through congestion in the matching function. The planner equates the sum of these two terms to the marginal cost of search, which is the term on the right.

The planner's choice for supplier search effort $v(\hat{z},\hat{a})$ is given by:

$$\sum_{z,a}\sum_{s=0}^{\infty}(1-\delta)^s\left(\lambda(\hat{z},\hat{a}+s)\frac{\partial y(\hat{z},\hat{a}+s)}{\partial G_{\hat{z},\hat{a}+s}(z,a+s)}u(z,a)n(z,a)-\lambda(z,a+s)\nu(\hat{z},\hat{a}+s,z,a+s)u(z,a)n(\hat{z},\hat{a})\right)\frac{\mathcal{M}\left(\mathcal{U},\mathcal{V}\right)}{\mathcal{U}\mathcal{V}}$$
$$+\frac{\partial\frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{V}}}{\partial v(\hat{z},\hat{a})}\frac{1}{\frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{V}}}\left[\Gamma_1-\Gamma_2\right]=W\frac{\partial\varphi_g(v(\hat{z},\hat{a}))}{\partial v(\hat{z},\hat{a})}n(\hat{z},\hat{a})$$

The first term on the left is the marginal surplus generated by a firm $(\hat{z}, \hat{a})$ connecting with more suppliers. This term is equal to the increase in firm $(\hat{z}, \hat{a})$'s output due to adding suppliers, multiplied by the value of that output, minus the value of the additional intermediate inputs firm $(\hat{z},\hat{a})$ now uses. The second term on the left is the marginal effect firm $(\hat{z}, \hat{a})$ has on all matches through congestion in the matching function. The planner equates the sum of these two terms to the marginal cost of search, which is the term on the right.

The decentralized firm's optimal customer search effort satisfies Equation 24. In Equation 25, we express the search effort which satisfies the planner's first order condition.

$$\frac{\mu-1}{\mu}\sum_{\tau=0}^{\infty}\sum_{z',a'}((1-\delta)\beta)^{\tau}\left(\frac{\mu c(z,a+\tau)^{1-\alpha}}{zc(z',a'+\tau)}\right)^{1-\sigma}\frac{1-\alpha}{\mu}s(z',a'+\tau)\frac{\mathcal{M}\left(\mathcal{U},\mathcal{V}\right)}{\mathcal{U}}\frac{v(z',a')n(z',a')}{\mathcal{V}}=\frac{\partial\varphi_h(u)}{\partial u} \tag{24}$$

$$\begin{aligned}\frac{\mu^2-1}{\mu}\sum_{z',a'}\sum_{\tau=0}^{\infty}((1-\delta)\beta)^{\tau}\left(\frac{\mu c(z,a+\tau)^{1-\alpha}}{zc(z',a'+\tau)}\right)^{1-\sigma}\frac{1-\alpha}{\mu}s(z',a'+\tau)\frac{\mathcal{M}\left(\mathcal{U},\mathcal{V}\right)}{\mathcal{U}}\frac{v(z',a')n(z',a')}{\mathcal{V}}\\+\frac{\partial\frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{U}}}{\partial u}\frac{1}{\frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{U}}}\frac{\mathcal{P}}{\mu}\left[\Gamma_1-\Gamma_2\right]=\frac{\partial\varphi_h(u)}{\partial u}\end{aligned} \tag{25}$$

The decentralized firm's optimal supplier search effort satisfies Equation 26. In Equation 27, we express the search effort that satisfies the planner's first order condition.

$$\frac{\mu-1}{\mu}\sum_{\tau=0}^{\infty}\sum_{z',a'}((1-\delta)\beta)^{\tau}\left(\frac{\mu c(z',a'+\tau)^{1-\alpha}}{z'c(z,a+\tau)}\right)^{1-\sigma}(1-\alpha)s(z,a+\tau)\frac{\mathcal{M}\left(\mathcal{U},\mathcal{V}\right)}{\mathcal{V}}\frac{u(z',a')n(z',a')}{\mathcal{U}}=\frac{\partial\varphi_g(v)}{\partial v} \tag{26}$$

$$\begin{aligned}\frac{\mu^2-1}{\mu^2}\sum_{z',a'}\sum_{\tau=0}^{\infty}((1-\delta)\beta)^{\tau}\left(\frac{\mu c(z',a'+\tau)^{1-\alpha}}{z'c(z,a+\tau)}\right)^{1-\sigma}(1-\alpha)s(z,a+\tau)\frac{\mathcal{M}\left(\mathcal{U},\mathcal{V}\right)}{\mathcal{V}}\frac{u(z',a')n(z',a')}{\mathcal{U}}\\+\frac{\partial\frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{V}}}{\partial\mathcal{V}}\frac{1}{\frac{\mathcal{M}(\mathcal{U},\mathcal{V})}{\mathcal{V}}}\frac{\mathcal{P}}{\mu}\left[\Gamma_1-\Gamma_2\right]=\frac{\partial\varphi_g(v)}{\partial v}\end{aligned} \tag{27}$$

Firms search inefficiently as they fail to internalize search externalities. First, when a firm exerts more search effort, it generates more matches, which creates a surplus for both the firm and the partners it matches with. While firms internalize their private surplus, they do not internalize the surplus generated for partners. The planner, on the other hand, internalizes joint surplus. This externality appears in the first term on the left when comparing Equations 24 and 25, and when comparing Equations 26 and 27. This externality has been termed the “thickness” and “composition” externalities in the search literature, and the “match creation” externality in some of the endogenous network literature.

Second, the planner internalizes congestion externalities, which arise through the matching function. This is the second term on the left in Equations 25 and 27. The planner understands that a (z, a) firm’s customer search effort affects matching probabilities for other firms searching for customers, and takes this into account when choosing search effort for the firm. Similarly, the planner understands that a (z, a) firm’s supplier search effort affects matching probabilities for other firms searching for suppliers, and takes this into account when choosing search effort for the firm.

In the case the two externalities exactly offset for all firms in the decentralized economy, the decentralized search efforts are efficient. However, under the decentralized surplus-splitting rule (i.e. constant markups), this is not possible and search effort is inefficient.