

# THE ORIGINS OF FIRM HETEROGENEITY: A PRODUCTION NETWORK APPROACH\*

Andrew B. Bernard<sup>†</sup> Emmanuel Dhyne<sup>‡</sup> Glenn Magerman<sup>§</sup>

Kalina Manova<sup>¶</sup> Andreas Moxnes<sup>||</sup>

May 2021

## Abstract

This paper explores firm size heterogeneity in production networks. Using all buyer-supplier relationships in Belgium, the paper shows that firms with more customers have higher total sales but lower sales per customer and lower market shares among those customers. A decomposition of firm sales reveals that downstream factors, especially the number of customers, explain the large majority of firm size heterogeneity. These facts motivate a model of heterogeneous firms and network formation where firms search for and sell to downstream buyers and buy inputs from upstream suppliers. Firms vary in their productivity and relationship capability. Higher productivity results in more matches and larger market shares among customers. Higher relationship capability results in more customers and higher sales. Estimated model parameters suggest that productivity and relationship capability are strongly negatively correlated. A counterfactual exercise shows that the real wage gains from improving relationship capability are substantial, and much larger in our model compared to canonical models of one-dimensional heterogeneity.

---

\*This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreements No 715147 and No 724880). The views expressed here are those of the authors and do not necessarily reflect those of the National Bank of Belgium. We thank the editor, five anonymous referees, discussants Julian di Giovanni, Isabelle Mejean, Jasmine Xiao and Thomas Åstebro, and numerous seminar participants for helpful comments and valuable discussion.

<sup>†</sup>Tuck School of Business at Dartmouth, CEP, CEPR & NBER; andrew.b.bernard@tuck.dartmouth.edu

<sup>‡</sup>National Bank of Belgium & Université de Mons; Emmanuel.Dhyne@nbb.be

<sup>§</sup>ECARES - Université libre de Bruxelles, CEPR & National Bank of Belgium; glenn.magerman@ulb.be

<sup>¶</sup>University College London, CEPR & CEP; k.manova@ucl.ac.uk

<sup>||</sup>University of Oslo & CEPR; andreamo@econ.uio.no

JEL: D24, F10, F12, F16, L11, L25

Keywords: Firm size heterogeneity, productivity, relationship capability, production network, network formation, matching costs.

# 1 Introduction

Even within narrowly defined industries, there is massive dispersion in firm outcomes such as sales, employment and productivity. In Belgium, a firm at the 90th percentile of the size distribution has turnover 32 times greater than a firm at the 10th percentile within the same industry.<sup>1</sup> Understanding the origins of firm size heterogeneity is a fundamental question in economics and has important micro- and macroeconomic implications. At the micro level, bigger firms perform systematically better along many dimensions, including survival, innovation, and participation in international trade (e.g., Bernard et al., 2012). At the macro level, the skewness and granularity of the firm size distribution affect aggregate productivity, the welfare gains from trade, and the impact of idiosyncratic and systemic shocks (e.g., Pavcnik, 2002, Gabaix, 2011, di Giovanni et al., 2014, Melitz and Redding, 2015 and Gaubert and Itskhoki, 2021).

This paper examines the firm size distribution in a production network with firm heterogeneity and buyer-supplier connections.<sup>2</sup> The basic premise of the analysis is intuitive: firms can be large because (i) they have inherently attractive capabilities such as productivity or quality, (ii) they have low marginal costs from matches with more or better suppliers, and/or (iii) they have higher sales due to more or bigger buyers. Higher-order effects can also be important, as the customers of customers also ultimately affect firm economic outcomes.

While research has made progress in identifying underlying firm-specific supply- and demand-side factors driving firm size (e.g., Hottman et al., 2016), much less is known about the role of buyer-seller linkages in production networks. In particular, the focus on the supply side has been on heterogeneity in either firm productivity (e.g., Jovanovic, 1982, Hopenhayn, 1992, Melitz, 2003, Luttmer, 2007) or organizational capital (e.g., Prescott and Visscher, 1980, Luttmer, 2011), whereas work on the demand side has centered on final customer preferences (e.g., Fitzgerald et al., 2016) or firm-specific demand shocks (e.g. Foster et al., 2016). To the extent that the literature has considered firm-to-firm trade, it has typically remained anchored in one-sided heterogeneity by assuming that firms source inputs from anonymous upstream suppliers or sell to anonymous downstream buyers, without accounting for the heterogeneity of all trade partners in the production network.

Our contribution is threefold. First, we document new stylized facts about production networks using data on the universe of firm-to-firm domestic transactions in Belgium, and

---

<sup>1</sup> The p90/p10 ratio is averaged across all NACE 4-digit Belgian industries in 2014. Hottman et al. (2016) report even larger sales dispersion across firms using barcode data on US retail sales, and Autor et al. (2020) find that 52.6 percent of sales in the average US industry is accounted for by the 20 largest firms, representing less than half a percent of the total number of firms.

<sup>2</sup> Throughout the paper, firm size refers to sales or turnover, and we use both terms interchangeably.

present the first extensive analysis of how upstream, downstream and final demand heterogeneity translate into firm size heterogeneity. Second, we provide a theoretical framework for an endogenous production network with firm heterogeneity in both productivity/quality and relationship capability. Third, we estimate the parameters of the model using simulated method of moments to explore the relative importance of the two dimensions of firm heterogeneity and their interaction across firms.

We report three stylized facts from the production network data that motivate the subsequent analysis and model. First, the distributions of total sales and buyer-supplier connections exhibit high dispersion. The enormous dispersion of sales across firms is also found in the production network in terms of the number of links to buyers and suppliers. Second, firms with more customers have higher sales but lower average sales per customer and lower market shares (shares of input purchases) among their customers. Finally, there is negative degree assortativity between buyers and suppliers: sellers with more customers match with customers who have fewer suppliers on average.

Taken together, these facts both confirm intuition and challenge existing models of firm heterogeneity. The large variation in sales across firms within an industry is intuitively related to variation in the number of customers: large firms have more customers. However, firms with more customers have lower average sales per customer, connect with less well-connected customers, and account for a smaller share of those customers' input purchases. Models that emphasize heterogeneity in productivity across firms cannot explain all three facts simultaneously. In particular, such models imply that firms with more customers should also sell more to each of their customers: they should have larger, not smaller market shares.

A key advantage of the Belgian production network data is that sales from firm  $i$  to  $j$  can be decomposed into seller-, buyer- and match-specific components.<sup>3</sup> This allows us to understand how much of the value of pairwise sales is due to the seller, the buyer or the match itself. High dispersion in seller effects means that firms vary in how much they sell to their customers, controlling for demand by those customers, i.e. firms differ in their average market share across customers. Conversely, high dispersion in buyer effects means that some firms match with large customers while others do not, leading to larger sales even as the average market share remains the same. Given estimates of these fixed effects, the total sales of a firm can be decomposed into three distinct factors: (i) an upstream component that captures the firm's ability to obtain large market shares across its customers, (ii) a downstream component that captures the firm's ability to attract many and/or large customers, and (iii) a final demand component that captures the firm's ability to sell relatively more outside the

---

<sup>3</sup> Our approach is similar to the inter-temporal analysis of matched employer-employee data (e.g. Abowd et al., 1999).

domestic network, to final consumers at home or to foreign customers.

The results are striking: 81 percent of the variation in firm sales within narrowly defined (4-digit NACE) industries is associated with the downstream component, while the upstream component contributes only 18 percent. The variation in firm size is largely unrelated to the relative importance of sales to final demand (1 percent). These findings imply that trade in intermediate goods and the number of firm-to-firm connections are essential to understanding firm performance and, consequently, aggregate outcomes.

Motivated by these stylized facts and decomposition results, we develop a quantitative general equilibrium model of firm-to-firm trade. In the model, firms use a constant-elasticity-of-substitution production technology that combines labor and inputs from upstream suppliers. Firms sell their output to final consumers and to domestic producers. Firms differ in two dimensions – productivity and relationship capability – defined respectively as production efficiency and (the inverse of) the fixed cost of matching with a customer. The two dimensions are potentially correlated. Suppliers match with customers if the gross profits of the match exceed the supplier-specific fixed matching cost. Marginal costs, employment, prices, and sales are endogenous outcomes because they depend on the outcomes of all other firms in the economy. A link between two firms increases the total sales of both the seller and the buyer; for the seller this occurs mechanically because it gains a customer, while for the buyer this arises because a larger supplier base lowers the marginal cost of production.

The model equilibrium involves three nested fixed points. A backward fixed point determines the price of a firm as a function of its marginal cost, which in turn depends on the prices of its suppliers. A forward fixed point pins down the sales of a firm as a function of demand by its customers, which in turn depends on their sales to their customers. A link function fixed point relates the likelihood of a link to the profit from the match, which is itself a function of the network structure. Jointly, these determine the endogenous structure of the network in terms of connections, the value of bilateral sales for each link, and the total sales of the firm.

We estimate the model parameters using simulated method of moments. These parameters comprise the variance of productivity, the mean and variance of relationship capability, and the correlation between productivity and relationship capability across firms. The results reveal high dispersion in relationship capability across firms and a strong negative correlation between the two firm characteristics. Firms with higher productivity have lower relationship capability. This negative relationship is crucial for matching the stylized fact that firms with more customers have lower average sales per customer and lower market share in those customers. A canonical model without this negative relationship instead produces a strongly positive relationship between the number of customers and average sales (or average market

share). The model does well at matching untargeted moments, such as the variances of total sales and value added. In addition, it does well at matching moments on the upstream side of the production network, including the variances of the number of suppliers and total input purchases. Importantly, both dimensions of firm heterogeneity are necessary to match the data: shutting down one at a time results in poor model fit, including the inability to replicate the negative relationship between the number of customers and average sales per customer.

Why are the two dimensions of firm capabilities negatively related? While it is outside the scope of this paper to offer a fully specified explanation, a possible answer is imperfect reward and incentive systems. For example, Holmstrom and Milgrom (1991) offer a multitasking theory where agents will focus their effort on observable and rewarded tasks at the expense of other tasks. If the principal only rewards one dimension of performance, e.g. finding customers, this might reduce other dimensions of firm performance, such as product quality or productivity.<sup>4</sup>

Finally, we use the estimated model to quantify the role of firm heterogeneity in productivity and relationship capability for aggregate outcomes. Specifically, we cut relationship costs across all firms by 50 percent. We do so in two versions of our model; in the baseline estimated model and then in a restricted model with no correlation between productivity and relationship costs. The counterfactual reveals that the real wage gains from lowering relationship costs are substantial, and much larger in our baseline model compared to the model with no correlation structure. The reason is that the fall in relationship costs benefits the high-productivity firms relatively more in the baseline model, as they are more constrained by high relationship costs.

This paper contributes to several strands of literature. Most directly, the paper adds to the large literature on the extent, causes and consequences of firm size heterogeneity. The vast dispersion in firm size has long been documented, with recent emphasis on the skewness and granularity of firms at the top end of the size distribution (e.g., Gibrat, 1931, Syverson, 2011). This interest is motivated by the superior growth and profit performance of bigger firms at the micro level, as well as by the implications of firm heterogeneity and superstar firms for aggregate productivity, growth, international trade, and adjustment to various shocks (e.g., Gabaix, 2011, Bernard et al., 2012, Freund and Pierola, 2015, Gaubert and Itskhoki, 2021, Oberfield, 2018).

Traditionally, this literature has analyzed own-firm characteristics on the supply side as the driver of firm size heterogeneity. The evidence indicates an important role for firms'

---

<sup>4</sup> In a recent study, Hong et al. (2018) find that workers trade off quality for quantity under a bonus scheme that rewards quantity (but not quality), as predicted by multitasking theory.

production efficiency, management ability, and capacity for quality products (e.g., Jovanovic, 1982, Hopenhayn, 1992, Melitz, 2003, Sutton, 2007, Bender et al., 2018, Bloom et al., 2017). Recent work has built on this by also considering the role of either upstream suppliers or downstream demand heterogeneity, but not both. Results suggest that access to inputs from domestic and foreign suppliers matters for firms' marginal costs and product quality, and thereby performance (e.g., Goldberg et al., 2010, Bøler et al., 2015, Manova et al., 2015, Fieler et al., 2018, Antràs et al., 2017, Boehm and Oberfield, 2020 and Bernard et al., 2019), while final consumer preferences affect sales on the demand side (e.g., Foster et al., 2016, Fitzgerald et al., 2016).

By contrast, we provide a comprehensive treatment of both own firm characteristics and production network features, on both the upstream and the downstream sides. The paper is related to Hottman et al. (2016) who also find that demand-side factors such as variation in firm appeal and product scope rather than prices (marginal costs) drive firm size dispersion. However, as these authors do not observe the production network, they cannot distinguish between the impact of serving more customers, attracting better customers, and selling large amounts to (potentially few) customers. Since they have no information on the supplier margin, they also cannot separate own from network supply factors. On the other hand, while rich in network features, our data do not provide information on prices or products, and thus do not allow for a comparable decomposition into firm appeal and product scope.

The paper also adds to a growing literature on buyer-supplier production networks (see Bernard and Moxnes, 2018 for a recent survey). Bernard et al. (2019) study the impact of domestic supplier connections on firms' marginal costs and performance in Japan, whereas Bernard et al. (2018a), Eaton et al. (2016), and Eaton et al. (2018) explore the matching of exporters and importers using data on firm-to-firm trade transactions for Norway, US-Colombia, and France, respectively. Using the Belgian production network data, Magerman et al. (2016) analyze the contribution of the network structure of production to aggregate fluctuations, while Tintelnot et al. (2021) examine the effect of trade on the domestic production network. In recent work, Acemoglu et al. (2012), Baqaee and Farhi (2019), Baqaee and Farhi (2020) and Lim (2018) consider how microeconomic shocks shape macroeconomic outcomes in networked environments. Our work departs from this literature by focusing on the dispersion of firm outcomes and their relationship to upstream and downstream features of the network.

The rest of the paper is organized as follows. Section 2 introduces the data and presents stylized facts about the Belgian production network. Section 3 decomposes firm sales into upstream, downstream and final demand components. Section 4 develops a theoretical frame-

work with heterogeneous firms and endogenous matching in a production network. Section 5 estimates the parameters of the model and quantitatively assesses the two dimensions of firm heterogeneity. The last section concludes.

## 2 Data and Stylized Facts

### 2.1 Data sources and preparation

The empirical analysis draws from four micro-level datasets on Belgian firms and their sales relationships, administered at the National Bank of Belgium (NBB). These include (i) the universe of domestic firm-to-firm relationships within Belgium from the NBB B2B Transactions Dataset, (ii) standard firm characteristics from the annual accounts, (iii) additional information on sales and inputs from the VAT declarations, and (iv) the sector of main economic activity and postal code of the firm from the Crossroads Bank of Enterprises. These datasets cover firms and their transactions in all economic activities over the years 2002-2014. Firms are identified by a unique enterprise number, which allows unambiguous merging across all datasets. A detailed description of the construction of the datasets is also provided in Appendix A.

Central to this paper is the NBB B2B Transactions Dataset (Dhyne et al., 2015), which is used to construct the production network of Belgian firms. It is based on the VAT listings that all VAT liable firms have to submit to the VAT authorities at the end of each calendar year.<sup>5</sup> An observation in this dataset reports the yearly sales value in euro of firm  $i$  selling to firm  $j$  within Belgium (excluding VAT). Sales values are the sum of invoices from  $i$  to  $j$ , which implies we observe the value, but not the content of the flows. All yearly sales values of at least 250 euro have to be reported, and pecuniary sanctions by the tax authorities on late or erroneous reporting ensure a very high quality of the data.

The other datasets contain information on firm characteristics. From firms' annual accounts, we retain information on firm-level sales, input expenditures, employment and labor costs. Flow variables are annualized pro rata from fiscal years to calendar years, to match the reporting in calendar years in the NBB B2B dataset. All firms have to report employment and labor costs. Depending on size thresholds, small firms can submit abbreviated annual accounts, which omits information on turnover and inputs expenditures.<sup>6</sup> We fill in these values for small firms using the VAT declarations, which contains information on sales and inputs for all VAT liable firms.<sup>7</sup> The main economic activity of the firm is extracted at the

<sup>5</sup> For VAT listings templates, see here (Dutch) and here (French).

<sup>6</sup> See the size criteria here.

<sup>7</sup> For VAT declaration templates, see here (Dutch) and here (French).



NACE 4-digit level (harmonized over time to the NACE Rev. 2 (2008) version) from the Crossroads Bank of Enterprises. To control for geographical heterogeneity, we also extract the postal code of the firm from this dataset.

We include firms with at least one full-time equivalent, to avoid potential issues with shell or management companies. We only keep the set of firms that are active in the production network. The main analysis is within NACE 4-digit industries based on the seller’s industry. To avoid potential incidental parameter problems (estimates of means, fixed effects etc.), we drop sectors with fewer than 5 observations. Results are robust to changing this cutoff. We calculate final demand as a firm’s turnover minus its B2B sales to other enterprises in the domestic network. Consistent with national accounting standards, final demand thus contains both sales to final consumers at home and exports.

We use the 2014 cross-section for the main analysis, and provide additional results in the appendices. The main sample for 2014 ultimately contains 94,147 firms and all their production network connections.

## 2.2 Stylized Facts

We start by documenting three empirical facts about firm size, firm-to-firm relationships and their correlations in the Belgian production network. These facts point to inconsistencies with standard models of one-dimensional firm heterogeneity and will motivate the foundations of our model in Section 4.

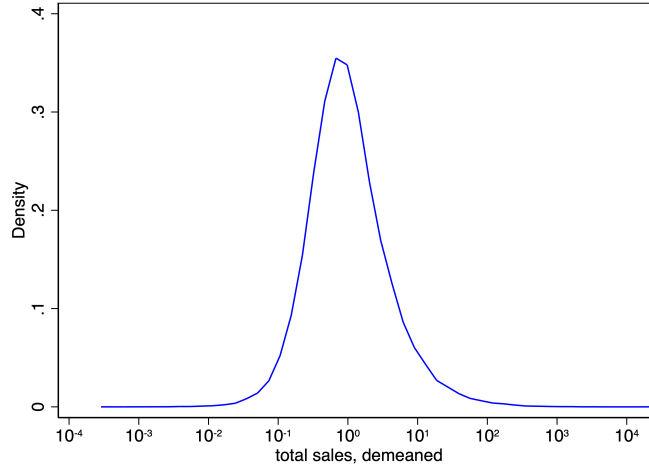
**Fact 1.** *The distributions of firm sales and supplier-buyer connections are highly dispersed.*

Even within narrowly defined industries, firms show significant heterogeneity along several dimensions.<sup>8</sup> Figure 1 documents the distributions of firm size and the number of customers and suppliers for Belgian firms. All variables are demeaned at the 4-digit NACE industry.<sup>9</sup> Panel (a) shows the firm size distribution, expressed in total sales value. As is well-known, the distribution spans several orders of magnitude: relative to the average firm in its industry, some firms are up to four orders of magnitude larger, and they co-exist with very small firms several orders of magnitude smaller than the average. Panel (b) reports the distribution of the number of customers of these firms. Here as well, firms can have over 1,000 times as many customers as their industry average, again co-existing with firms that have few customers. Similarly, in panel (c), the number of suppliers is shown. While less

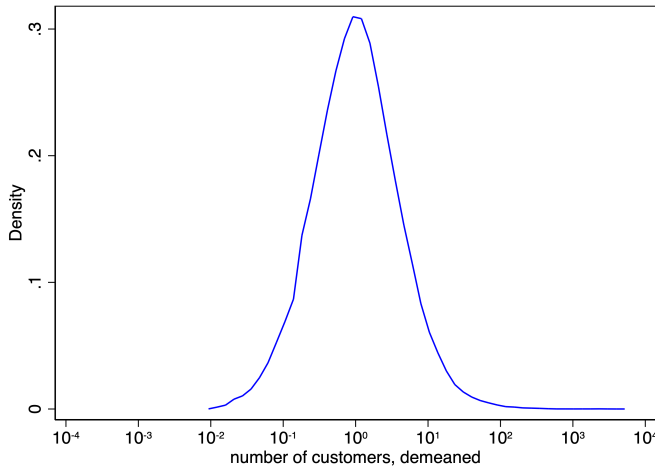
---

<sup>8</sup> These patterns mirror findings for firm-to-firm linkages in the domestic production network in Belgium (Dhyne et al., 2015) and Japan (Bernard et al., 2019) and for firm-to-firm export transactions in Norway (Bernard et al., 2018a).

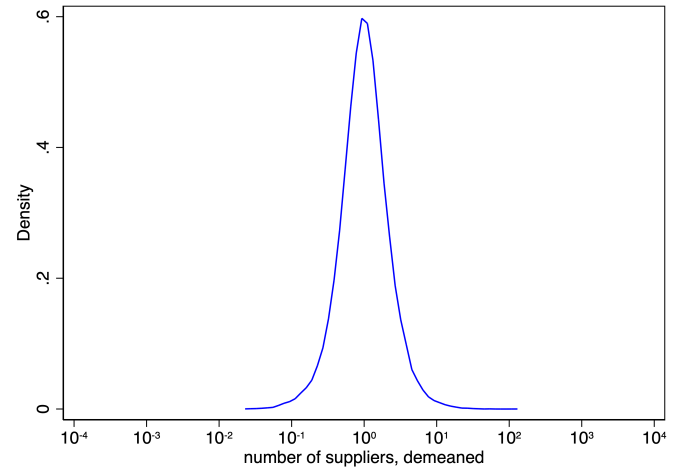
<sup>9</sup> We demean variables by regressing log variables on 4-digit sector fixed effects and retain the residuals as demeaned log variables.



(a)



(b)



(c)

Figure 1: Distribution of firm sales, number of customers and number of suppliers.

excessive, again this distribution spans several orders of magnitude. Appendix B reports additional moments on both demeaned and raw variables.

**Fact 2.** *Firms with more customers have higher sales but lower sales per customer.*

A sharp pattern in the data is that firms with more customers have higher sales but lower sales per customer. Figure 2a displays the binned scatterplot of firm sales to other producers in the network (y-axis) against the number of customers (x-axis), on a log-log scale. Both variables are demeaned by their 4-digit industry average, and observations are binned into 20 quantiles. The elasticity of sales with respect to the number of buyers is 0.77. Therefore, sales increases in the number of customers, but less than proportionally. This directly implies

that sales per customer decreases with the number of customers, as illustrated in Figure 2b, with an elasticity of  $-0.23$ .<sup>10</sup>

This pattern is not driven by composition effects among customers. Figure 3a demonstrates that sales per customer fall with the number of customers for both big and small customers. For each firm, we calculate the 10th, 50th and 90th percentiles of sales across its buyers, and plot these percentiles against the firm’s number of buyers. The slope coefficients are negative, and range between  $-.13$  and  $-.22$ . This implies that firms do not systematically tend to sell relatively more or less to their top customers at the expense of their bottom customers, when they add more buyers.

One may also wonder if the decline in average sales per customer is driven by selection. If sellers match with smaller buyers when they grow their customer base, they would record lower average and median bilateral sales. To address this concern, we leverage the network data and calculate a firm’s weighted average market share among its customers: the geometric mean of  $m_{ij}/M_j^{net}$ , where  $m_{ij}$  is sales from  $i$  to  $j$  and  $M_j^{net}$  is total network purchases by firm  $j$ , using sales shares  $m_{ij}/S_i^{net}$  as weights.<sup>11</sup> If selection were the main mechanism, this weighted average market share would be increasing in, or unrelated to, the number of customers. Figure 3b shows that this is not the case: Firms’ weighted average market share also declines with their number of customers, with an elasticity of  $-0.08$ .

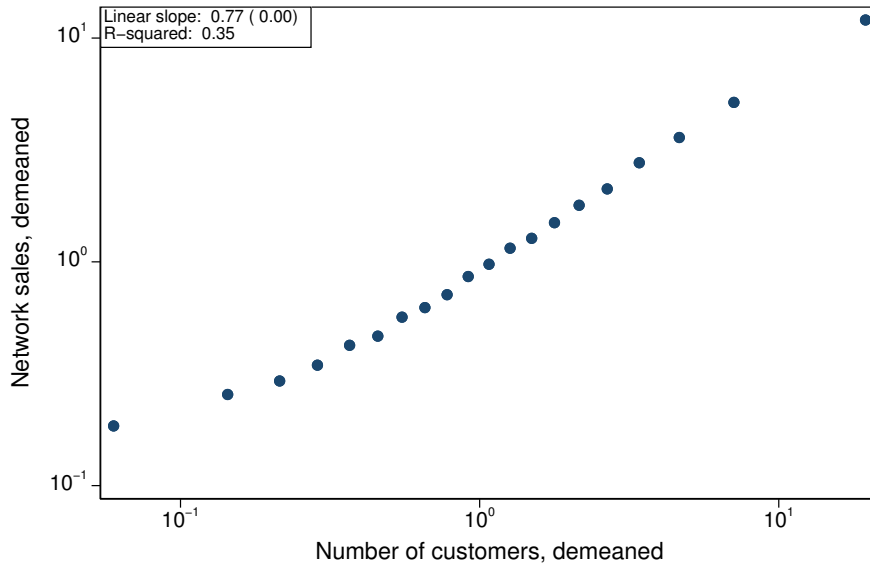
We explore the potential impact of additional dimensions of customer heterogeneity in Appendix B. In particular, we control for heterogeneity in input requirements across customers within the seller’s industry. We also consider the role of fringe buyers, i.e. relatively unimportant customers in terms of  $m_{ij}$ . In all cases, our empirical findings retain the same message. Taken together, these empirical regularities present a puzzle: big firms match with many buyers, but they are unable to gain a large market share among those buyers. By contrast, in canonical one-dimensional models of firm heterogeneity, or models with two or more, but independent, dimensions (e.g., Arkolakis, 2010, Bernard et al., 2018a, Eaton et al., 2018, Lim, 2018), highly productive firms would both attract many customers and have a high market share among those customers. The empirical evidence therefore calls for a model with an additional element of firm heterogeneity, where firm size is not only deter-

<sup>10</sup> We construct total domestic network sales as  $S_i^{net} = \sum_j m_{ij}$ . We thus obtain an identity between  $S_i^{net}$ , the number of customers  $n_i^c$  and the average sales per customer, as  $\frac{S_i^{net}}{n_i^c} = \frac{1}{n_i^c} \sum_j m_{ij}$ . Note that this identity implies that the elasticities in Figures 2a and 2b amount to  $0.77 - (-.23) = 1$ . As we observe total sales  $S_i$ , but not the number of customers in export destinations, this identity would no longer hold if using  $S_i$  instead of  $S_i^{net}$ . However, all results are very similar and qualitatively the same when using total sales instead of network sales.

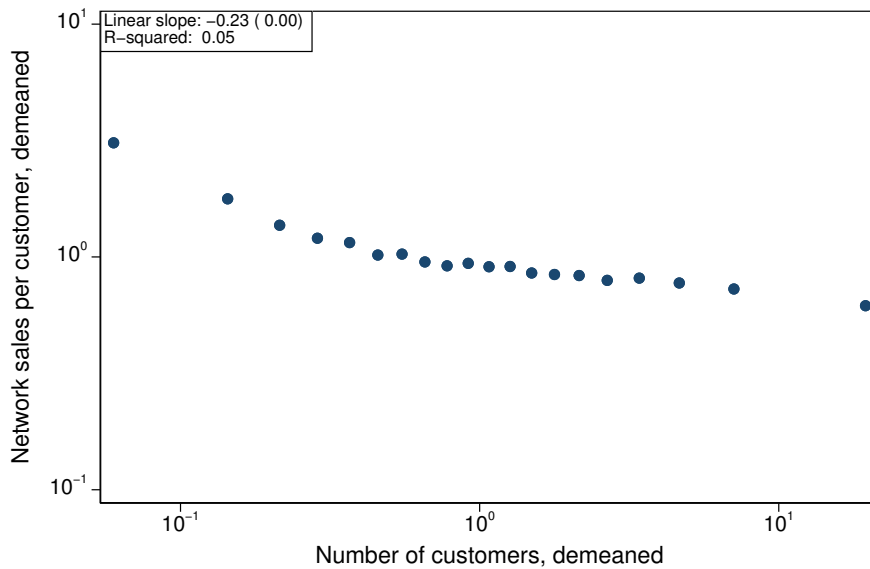
<sup>11</sup> I.e., the weighted average market share is  $\bar{\delta}_i = \prod_j \left( \frac{m_{ij}}{M_j^{net}} \right)^{\frac{m_{ij}}{S_i^{net}}}$ . The weighted average puts less emphasis on fringe customers. Using the unweighted average however, produces similar results.

Figure 2: Total Network Sales, Average Sales and Number of Customers

(a) Total Network Sales

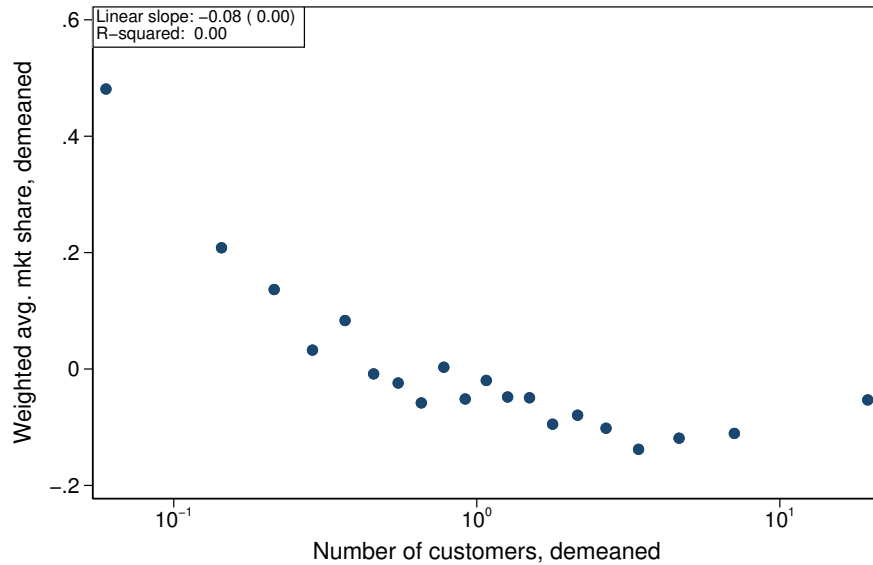
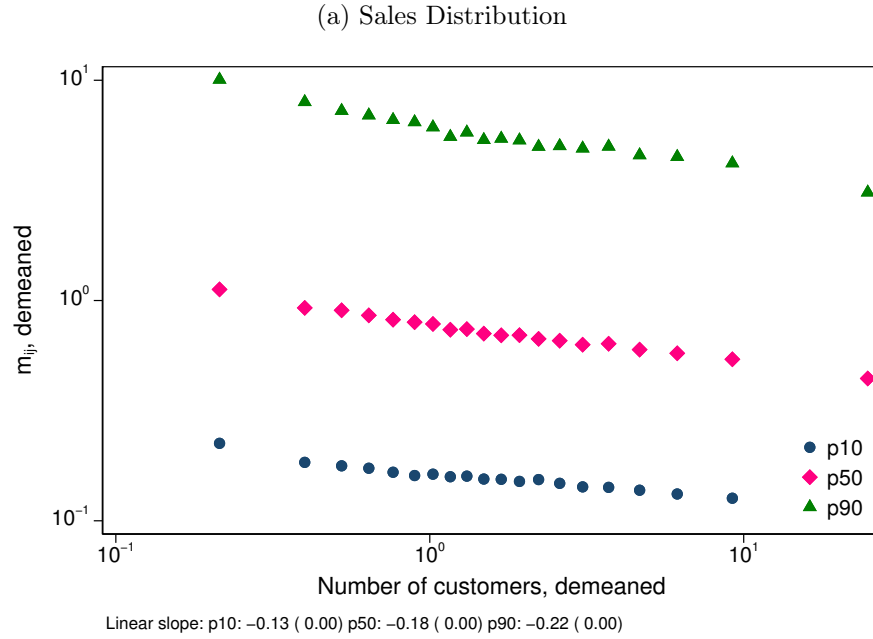


(b) Average Sales



Note: The binned scatterplots group firms into 20 equal-sized bins by log number of customers, and compute the mean of the variables on the x- and y-axes in each bin. Network sales refer to a firm's total sales to customers in the domestic production network. All variables are demeaned by NACE 4-digit industry averages. Implied elasticities and R-squared from OLS regressions with NACE 4-digit industry fixed effects are reported in the corner of each graph.

Figure 3: Sales Distribution, Market Share and Number of Customers



Note: The binned scatterplots group firms into 20 equal-sized bins by log number of customers, and compute the mean of the variables on the x- and y-axes in each bin. All variables are demeaned by NACE 4-digit industry averages. p90/p50/p10 refer to the 90th/50th/10th percentile of firm-to-firm sales  $m_{ij}$  across buyers  $j$  within firm  $i$ . Average market share refers to the weighted geometric mean of the market shares  $m_{ij}/M_j^{net}$  of firm  $i$  in the total network purchases of its buyers  $j$ , using sales shares  $m_{ij}/S_i^{net}$  as weights.

mined by productivity, but also by a second firm attribute that enables firms to match with more buyers.

**Fact 3.** *Sellers with more customers match with customers who have fewer suppliers on average.*

An important property of networks is the extent to which a well-connected node is linked to other well-connected nodes, so-called degree assortativity. The production network is characterized by negative degree assortativity. In other words, better connected firms match to less well-connected firms on average.<sup>12</sup> Figure 4 shows a binned scatterplot of the average number of suppliers to firm  $i$ 's customers on the y-axis against the number of  $i$ 's customers, on a log-log scale. The fitted regression line has slope -0.05, such that doubling the number of customers is associated with a 5 percent decline in the average customer's number of suppliers. We also find a robust, and more negative, relationship between a firm's number of suppliers and the average supplier's number of customers (see Table 5 in Section 5.2).

Negative degree assortativity motivates our choice of a parsimonious matching model, in which firm connections form whenever the gross profits of a match exceed the fixed cost of forming a relationship. In this class of models, the marginal (and average) customer of more capable firms is less capable, generating a pattern of negative degree assortativity.

### 3 An Exact Decomposition

In this section, we develop an exact variance decomposition of firm sales into upstream, downstream, and final demand margins. The downstream component reflects characteristics of a firm's customers (i.e., their number and size), while the upstream component captures firm characteristics that remain constant across customers (i.e., average sales to customers, controlling for their size). Final demand includes factors unrelated to the domestic production network (i.e., sales to final consumers or foreign customers). This method exploits the granularity of the firm-to-firm transaction data in a way that would not be feasible with standard firm-level datasets, and its results provide the rationale for the structural framework in Section 4.

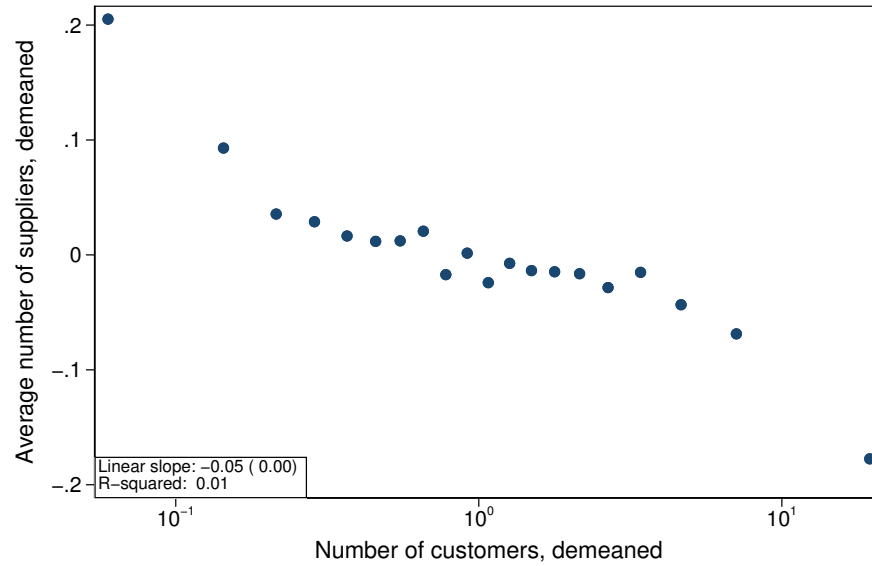
#### 3.1 Methodology

We start by estimating buyer, seller and buyer-seller match effects using data on sales between firms in the production network. We then use these estimates to decompose the variance of

---

<sup>12</sup> Negative degree assortativity has been documented in earlier research on production networks, e.g. Bernard et al. (2019), Bernard et al. (2018a) and Lim (2018).

Figure 4: Degree Assortativity



Note: The binned scatterplot groups firms into 20 equal-sized bins by log number of customers, and computes the mean of the variables on the x- and y-axes in each bin. Average number of suppliers refers to the geometric mean of the number of suppliers serving the customers of firm  $i$ . All variables are demeaned by NACE 4-digit industry averages. Implied elasticities and R-squared from OLS regressions with NACE 4-digit industry fixed effects are reported in the corner of the graph.

firm sales. The specification is a two-way fixed effects regression for firm-to-firm sales:

$$\ln m_{ij} = \ln G + \ln \psi_i + \ln \theta_j + \ln \omega_{ij}, \quad (1)$$

where  $\ln m_{ij}$  is log sales from  $i$  to  $j$ , and  $\ln G$  is the mean of  $\ln m_{ij}$  across all  $ij$  pairs. The seller effect  $\ln \psi_i$  reflects the amount of sales by  $i$  to its average customer  $j$ , controlling for total purchases by  $j$  via  $\theta_j$ . The seller effect is therefore related to the average market share of  $i$  among her customers.<sup>13</sup> Analogously, the buyer effect  $\ln \theta_j$  captures the value of input purchases by  $j$  from its average supplier  $i$ , controlling for total sales by  $i$  via  $\psi_i$ . Intuitively, attractive buyers (high  $\theta_j$ ) purchase a disproportionate share of suppliers' sales. Finally,  $\ln \omega_{ij}$  is the residual from the regression. A positive  $\ln \omega_{ij}$  reflects match-specific characteristics that induce a given firm pair to trade more with each other, even if they are not fundamentally attractive trade partners.

To illustrate the advantage of the bilateral sales data, consider first an extreme case in which the variation in  $\ln m_{ij}$  is only due to  $\psi_i$ . Seller  $i$  is then larger than seller  $i'$  because  $i$  sells more to every customer, while there is no variation in how much each of these customers buys from  $i$ . In this case, firm size heterogeneity is only driven by seller characteristics  $\psi_i$ ; who you are as a seller explains firm size. Consider next the opposite case in which the variance in  $\ln m_{ij}$  is only due to  $\theta_j$ . Seller  $i$  now dominates seller  $i'$  because  $i$  matches with bigger customers than  $i'$ , while sales to common customers  $j$  are identical. In this case, firm heterogeneity is only driven by differences in matching ability across sellers; who you meet as a seller explains firm size. In standard firm-level datasets, we cannot differentiate between these two scenarios because they are observationally equivalent. Estimating equation (1) using OLS poses some threats to identification. First, to obtain unbiased estimates, the assignment of suppliers to customers must be exogenous with respect to  $\omega_{ij}$ , so-called conditional exogenous mobility (Abowd et al., 1999). This assumption, as well as tests for exogenous mobility and functional form relevance, are discussed at length in Appendix C. Overall, we find strong support for the log-linear model and the conditional exogenous mobility assumption.

Second, to identify the fixed effects, firms must have multiple connections. Specifically, identifying a seller fixed effect requires a firm to have at least two customers, and identifying a buyer fixed effect requires a firm to have at least two suppliers. Therefore, single-customer and single-supplier links are dropped in the estimation procedure. Also, dropping customer  $A$  might result in supplier  $B$  having only one customer left. Supplier  $B$  is then also removed from the sample. This iterative process continues until a connected network component

---

<sup>13</sup> This is shown formally in Appendix D.2.



remains (i.e. a within-projection matrix of full rank), in which each seller has at least two customers and each buyer has at least two suppliers. This component is known as a mobility group in the labor literature on firm-employee matches.

Identification is obtained from cross-sectional variation. Compared to related work on firm-employee matches in the labor literature (e.g Abowd et al., 1999), this works to our advantage.<sup>14</sup> First, it attenuates an incidental parameter problem as the number of suppliers per customer and the number of customers per supplier is relatively large: the median number of customers and suppliers is 26 and 53, respectively (see Section 2). Second, we do not require the otherwise standard assumption that the fixed effects be constant over time, as identification comes from a single cross-section.

Once we have estimated parameters  $\Psi = \{\psi_i, \theta_j, \omega_{ij}\}$ , firm sales can be exactly decomposed into upstream, downstream, and final demand factors. Total sales of firm  $i$  are by construction  $S_i = \sum_{j \in \mathcal{C}_i} m_{ij} + \mathcal{F}_i$ , where  $\mathcal{C}_i$  is the set of firm  $i$ 's customers and  $\mathcal{F}_i$  is sales to final demand (i.e., sales outside of the domestic network). Therefore, total sales can be expressed as  $\ln S_i = \ln S_i^{net} + \ln \beta_i$ , where  $S_i^{net} \equiv S_i - \mathcal{F}_i$  is network sales and  $\beta_i$  is total sales relative to network sales,  $\beta_i \equiv S_i/S_i^{net} \geq 1$ , i.e. an inverse measure of (the share of) network sales.

As shown in Appendix D.1, total sales can be decomposed as

$$\ln S_i = \ln G + \ln \psi_i + \ln n_i^c + \ln \bar{\theta}_i + \ln \Omega_i^c + \ln \beta_i, \quad (2)$$

where  $n_i^c$  is the number of customers,  $\bar{\theta}_i \equiv \left(\prod_{j \in \mathcal{C}_i} \theta_j\right)^{1/n_i^c}$  is the average buyer fixed effect among customers, and  $\Omega_i^c \equiv \frac{1}{n_i^c} \sum_{j \in \mathcal{C}_i} \omega_{ij} \theta_j / \bar{\theta}_i$  is an interaction term between the buyer fixed effect and match quality.<sup>15</sup>

Each of these components has an intuitive economic interpretation. The  $\psi_i$  component represents *upstream* fundamentals that shape firm size: If sales dispersion is only due to variance in  $\psi_i$ , then large firms have larger market shares among their customers than small firms, while the number of customers is the same. The  $n_i^c$ ,  $\bar{\theta}_i$  and  $\Omega_i^c$  components represent *downstream* fundamentals that shape firm size: Firms face high network demand if (i) they are linked to many customers (high  $n_i^c$ ), (ii) their average customer has high input purchases (high  $\bar{\theta}_i$ ), and/or (iii) the interaction term  $\Omega_i^c$  is large, i.e. large customers (high  $\theta_j$ ) also happen to be good matches (high  $\omega_{ij}$ ). If sales dispersion is only due to variance in these downstream factors, large firms transact with more, bigger, and/or better-matched customers than small firms, while market shares are the same across customers. Finally,  $\beta_i$  represents

<sup>14</sup> In Abowd et al. (1999), identification comes from workers that move across firms over time.

<sup>15</sup> By the properties of ordinary least squares, the average term  $(1/n_i^c) \sum_{j \in \mathcal{C}_i} \ln \omega_{ij} = 0$  and is therefore omitted from the expression above.

Table 1: Buyer and Seller Effects

	$N$	$\frac{\text{var}(\ln\psi_i)}{\text{var}(\ln\psi_i+\ln\theta_j)}$	$\frac{\text{var}(\ln\theta_j)}{\text{var}(\ln\psi_i+\ln\theta_j)}$	$\frac{2\text{cov}(\ln\psi_i,\ln\theta_j)}{\text{var}(\ln\psi_i+\ln\theta_j)}$	$R^2$	Adjusted $R^2$
$\ln m_{ij}$	17,054,274	0.66	0.32	0.02	0.43	0.39

Note: The table reports the (co)variances of the estimated seller and buyer fixed effects from equation (1). The estimation is based on the high-dimensional fixed effects estimator from Correia (2016).

the share of sales that goes to final demand, capturing all downstream variation outside the production network.

Note that all elements in equation (2) are known:  $S_i$ ,  $\beta_i$ ,  $n_i^c$  and  $G$  come directly from the data, while  $\psi_i$ ,  $\bar{\theta}_i$  and  $\Omega_i^c$  are estimated from equation (1). In order to assess the role of each margin, we follow the literature (Eaton et al., 2004, Hottman et al., 2016) and perform a simple variance decomposition on equation (2). All observed and constructed variables are first demeaned by their NACE 4-digit industry average, to difference out systematic variation across industries. We then regress each component ( $\ln \beta_i$ ,  $\ln \psi_i$ ,  $\ln n_i^c$ ,  $\ln \bar{\theta}_i$  and  $\ln \Omega_i^c$ ) separately on log sales. By the properties of ordinary least squares, and from the exact nature of the decomposition, the coefficients from these regressions sum to unity and represent the share of the overall variation in firm size explained by each margin.

## 3.2 Results

The results from estimating equation (1) are reported in Table 1.<sup>16</sup> Three patterns stand out. First, the adjusted  $R^2$  from the regression is 0.39, indicating that the buyer and seller fixed effects explain a sizable share of the variation in firm-to-firm sales. Second, the variation in the seller effect  $\ln \psi_i$  is larger than that in the buyer effect  $\ln \theta_j$ . Third, the correlation between the fixed effects is close to zero.

The results in Table 1 shed light on the variation in transaction values,  $m_{ij}$ , but not on the variation in firm sales,  $S_i$ . Table 2 reports the results for the exact firm sales decomposition in equation (2). Relative differences in final demand across firms, as captured by the ratio of total to network sales,  $\ln \beta_i$ , account for an economically negligible 1% of the overall variation in firm size. Thus, large firms are not systematically selling relatively more (or less) to final demand than small firms. The upstream factor  $\ln \psi_i$  represents, roughly speaking, the average market share of  $i$  among its customers. Being an important supplier to one's customers is weakly related to overall firm success, contributing 18% of the variation in firm sales. The

<sup>16</sup> After removing firms with un-identified fixed effects, 99% of the links and 95% of the value of all transactions remain in the estimation sample.

Table 2: Firm Size Decomposition ( $\ln S_i$ ).

Relative final demand	$\ln \beta_i$	.01	(.00)
Upstream	$\ln \psi_i$	.18	(.00)
# Customers	$\ln n_i^c$	.51	(.00)
Avg Customer Capability	$\ln \bar{\theta}_i$	.05	(.00)
Customer Interaction	$\ln \Omega_i^c$	.25	(.00)

Note: The table reports coefficient estimates from separate OLS regressions of a firm size margin (as indicated in the row heading) on  $\ln S_i$ . All variables are first demeaned by their 4-digit NACE industry average. The number of firms in the core sample is 94,147. Standard errors in parentheses.

three final rows report the magnitude of the downstream margins  $\ln n_i^c$ ,  $\ln \bar{\theta}_i$  and  $\ln \Omega_i^c$ . In total, the downstream side accounts for 81% of the size dispersion across firms. Most of the variation in the downstream component across firms can be attributed to the extensive margin, i.e. the number of (domestic) buyers,  $\ln n_i^c$ . The average sourcing capability across a firm’s customers,  $\ln \bar{\theta}_i$ , and the customer interaction term,  $\ln \Omega_i^c$ , explain a more modest 5% and 25%, respectively.

Therefore, the single most important advantage of large firms is that they successfully match with many buyers, whereas the characteristics of these buyers play a smaller role. These findings suggest that a key to understanding the vast firm size heterogeneity observed in modern economies is how firms manage their sales activities, and specifically how they match and transact with buyers in the production network. Appendix E provides several robustness checks, and reports results for business groups, individual industries and different years, reinforcing our main conclusions.

### 3.3 Correlations

We conclude this section by documenting the correlations among various firm characteristics in the data. Column 1 in Table 3 shows that firm sales are strongly positively correlated with both the upstream,  $\ln \psi_i$ , and the various downstream components ( $\ln n_i^c$ ,  $\ln \bar{\theta}_i$ ,  $\ln \Omega_i^c$ ). The number of customers ( $\ln n_i^c$ ) and the upstream component ( $\ln \psi_i$ ) are negatively correlated. This mirrors the findings in Section 2, and implies that firms with many customers tend to have smaller average market shares among those customers. Our interpretation of this pattern is that firms are unlikely to succeed along both the extensive and the intensive margins: Some firms become large by accumulating a broad customer base, while other firms become large by being important suppliers to their clients, and few firms manage to do both.

Table 3: Correlation Matrix

Firm Size Component	$\ln S_i$	$\ln \psi_i$	$\ln n_i^c$	$\ln \bar{\theta}_i$	$\ln \Omega_i^c$	$\ln \beta_i$
$\ln S_i$	1					
$\ln \psi_i$	0.23	1				
$\ln n_i^c$	0.49	-0.33	1			
$\ln \bar{\theta}_i$	0.20	0.22	-0.18	1		
$\ln \Omega_i^c$	0.45	0.16	0.09	0.23	1	
$\ln \beta_i$	0.02	-0.36	-0.33	-0.16	-0.42	1

Note: All correlations are significant at 5%. All variables are demeaned at the NACE 4-digit level.

These results, coupled with the stylized facts and sales decomposition, are difficult to reconcile with canonical heterogeneous-firm models. They suggest that both upstream and downstream dimensions of firm activity underpin sales dispersion when firms interact in production networks. One interpretation of our findings is that firm attributes that matter for matching with customers and suppliers are orthogonal, or even negatively related, to firm attributes that determine sales conditional on a match.

## 4 Theoretical Framework

Motivated by the stylized facts, this section develops a theoretical framework of a buyer-supplier production network with two-sided firm heterogeneity and endogenous match formation. This framework admits a two-step analysis: We first present the model conditional on a set of firm-to-firm links, and subsequently introduce a parsimonious firm-to-firm matching model.

Our starting point is a framework in which firms are heterogeneous in two dimensions. First, firms within an industry have different productivities, which implies that they have different marginal costs and prices.<sup>17</sup> Second, firms have different relationship capabilities. These capabilities determine their ability to match with customers conditional on their (quality-adjusted) prices. We model relationship capability as a fixed cost that the firm must incur for each customer it chooses to serve. A firm with lower relationship fixed costs will endogenously be able to match with more customers, all else equal. In contrast to much of the earlier literature, productivity and relationship capability are potentially correlated.

<sup>17</sup> As is standard in this class of models, under the assumption of CES preferences and monopolistic competition, productivity and product quality enter equilibrium firm revenue in exactly the same way.

Firms operate in a production network, sourcing their inputs from other firms and selling their output to both other firms and final demand. In addition to productivity and relationship capability, a firm's size thus also depends on its input prices. Input prices are low and sales high if the firm has many low-price (or high-quality) suppliers.

## 4.1 Technology and Demand

The economy consists of a unit continuum of firms, each with the following production function:

$$y(i) = \kappa z(i) l(i)^\alpha v(i)^{1-\alpha},$$

where  $y(i)$  is output (in quantities) of firm  $i$ ,  $z(i)$  is productivity,  $l(i)$  is the amount of labor used by firm  $i$ ,  $\alpha$  is the labor share, and  $\kappa > 0$  is a normalization constant.<sup>18</sup>  $v(i)$  is a constant-elasticity-of-substitution (CES) input bundle:

$$v(i) = \left( \int_{\mathcal{S}(i)} \nu(k, i)^{(\sigma-1)/\sigma} dk \right)^{\sigma/(\sigma-1)},$$

where  $\nu(k, i)$  is the quantity purchased from firm  $k$ ,  $\mathcal{S}(i)$  is the set of suppliers to firm  $i$ , and  $\sigma > 1$  is the elasticity of substitution across suppliers. The corresponding input price index is  $P(i) = \left( \int_{\mathcal{S}(i)} p(k)^{1-\sigma} dk \right)^{1/(1-\sigma)}$ , where  $p(k)$  is the price charged by supplier  $k$ . Setting the wage  $w$  as the numéraire, the marginal cost of the firm is:

$$c(i) = \frac{P(i)^{1-\alpha}}{z(i)}. \quad (3)$$

*Final Demand.* Final consumers have a CES utility function with the same elasticity of substitution  $\sigma$  across output varieties. The representative consumer is the shareholder of all firms, so that aggregate profits  $\Pi$  become part of consumer income. Aggregate income  $X$  is therefore the sum of aggregate labor income and aggregate corporate profits,  $X = wL + \Pi$ , where  $L$  is inelastically supplied labor.

---

<sup>18</sup>  $\kappa \equiv \alpha^{-\alpha} (1 - \alpha)^{-(1-\alpha)}$ . This normalization maps the production function to the cost function, and simplifies the expression for the cost function, without any bearing on our results.

## 4.2 Firm-to-Firm Sales

Each firm faces demand from other firms, as well as from final consumers. Given the assumptions about technology, sales from firm  $i$  to firm  $j$  are:

$$m(i, j) = p(i)^{1-\sigma} P(j)^{\sigma-1} M(j), \quad (4)$$

where  $M(j)$  are total intermediate purchases by firm  $j$ ,  $M(j) = \int_{S(j)} m(i, j) di$ .

The market structure is monopolistic competition, such that firms charge a constant mark-up over marginal costs,  $p(i) = \mu c(i)$ , where  $\mu \equiv \sigma / (\sigma - 1)$ . After rearranging, sales from  $i$  to  $j$  can be expressed as:

$$m(i, j) = \left[ \frac{z(i)}{\mu P(i)^{1-\alpha}} P(j) \right]^{\sigma-1} M(j). \quad (5)$$

The model thus delivers a simple log linear expression for firm-to-firm sales, just as in the reduced-form equation (1) in Section 3.

## 4.3 Equilibrium Conditional on Network

We characterize the equilibrium in two separable steps. This section first describes properties of the partial equilibrium conditional on a fixed network structure. The next section then develops the firm-to-firm matching model, and specifies the general equilibrium with endogenous match formation.

To proceed, we introduce additional notation. A firm  $i$  is characterized by the tuple  $\lambda = (z, F)$ , where  $z$  is productivity and  $F$  is a relationship fixed cost, in units of labor.  $z$  and  $F$  are potentially correlated, and  $dG(\lambda)$  denotes the (multivariate) density of  $\lambda$ . We define the link function  $l(\lambda, \lambda')$  as the share of seller-buyer pairs  $(\lambda, \lambda')$  that match.<sup>19</sup>

*Backward fixed point.* For a given network structure, the equilibrium can be found by solving for two fixed points sequentially. Using the pricing rule  $p(\lambda) = \mu c(\lambda)$  and the equation for marginal costs (3), the input price index can be solved by iterating on a backward fixed point problem:

$$P(\lambda)^{1-\sigma} = \mu^{1-\sigma} \int P(\lambda')^{(1-\sigma)(1-\alpha)} z(\lambda')^{\sigma-1} l(\lambda', \lambda) dG(\lambda'). \quad (6)$$

<sup>19</sup> Due to idiosyncratic pairwise fixed cost shocks, the link function will take values between 0 and 1, see Section 4.4.

The input cost index of firm  $\lambda$ ,  $P(\lambda)$ , depends on the input cost index and productivity of all its suppliers  $\lambda'$ ,  $P(\lambda')$  and  $z(\lambda')$ .

*Forward fixed point.* Sales of a type- $\lambda$  firm are the sum of sales to final and intermediate demand:  $S(\lambda) = \mathcal{F}(\lambda) + \int m(\lambda, \lambda') l(\lambda, \lambda') dG(\lambda')$ , where  $m(\lambda, \lambda')$  now denotes sales by supplier  $\lambda$  to buyer  $\lambda'$ . Final demand is  $\mathcal{F}(\lambda) = p(\lambda)^{1-\sigma} \mathcal{P}^{\sigma-1} X$ , with the consumer price index equal to  $\mathcal{P}^{1-\sigma} = \int p(\lambda)^{1-\sigma} dG(\lambda) = \mu^{1-\sigma} \int P(\lambda)^{(1-\sigma)(1-\alpha)} z(\lambda)^{\sigma-1} dG(\lambda)$ . Also note that total input purchases are  $M(\lambda) = S(\lambda)(1-\alpha)/\mu$ . Using this together with equation (3) yields:

$$S(\lambda) = \mu^{1-\sigma} z(\lambda)^{\sigma-1} P(\lambda)^{(1-\sigma)(1-\alpha)} \left( \frac{X}{\mathcal{P}^{1-\sigma}} + \frac{1-\alpha}{\mu} \int \frac{S(\lambda')}{P(\lambda')^{1-\sigma}} l(\lambda, \lambda') dG(\lambda') \right). \quad (7)$$

Sales of a type- $\lambda$  firm depend on final demand,  $X$ , the productivity and input price index of the firm itself,  $z(\lambda)$  and  $P(\lambda)$ , and the sales and input prices of its customers,  $S(\lambda')$  and  $P(\lambda')$ . Appendix D.3 proves the existence and uniqueness of the equilibrium.

#### 4.4 Firm-to-Firm Matching

We now consider the general equilibrium when the production network is endogenous and sellers match with buyers if and only if the profits from doing so are positive. The seller incurs a relationship fixed cost  $F\epsilon$  for every buyer it chooses to sell to, where  $F$  varies across sellers, and  $\epsilon$  is an idiosyncratic component that varies across firm pairs. This matching model is similar to Bernard et al. (2018a) and Lim (2018), but in contrast to these papers,  $F$  is a firm-specific attribute that can be correlated with the productivity of the firm,  $z$ .<sup>20</sup>

The share of seller-buyer pairs  $(\lambda, \lambda')$  that match and trade with each other is then:

$$l(\lambda, \lambda') = \int I[\ln \epsilon < \ln \pi(\lambda, \lambda') - \ln F] dH(\epsilon), \quad (8)$$

where  $I[\cdot]$  is the indicator function,  $dH(\epsilon)$  denotes the density of  $\epsilon$ , and the gross profits from the potential match are:

$$\pi(\lambda, \lambda') = \frac{m(\lambda, \lambda')}{\sigma}.$$

The introduction of idiosyncratic match costs  $\epsilon$  is not needed to solve the model or to rationalize the stylized facts presented earlier in the paper. However,  $\epsilon$  will play a role in the structural estimation in Section 5. Formally, dispersion in  $\epsilon$  ensures that the link function is continuous in the parameters of the model, such that standard gradient-based numerical

<sup>20</sup> On the other hand, sales to final demand incur no fixed costs and vary across firms only due to differences in output prices.

methods can be used to minimize the objective function. Intuitively,  $\epsilon$  can be justified with seller-buyer specific costs that affect the profitability of the relationship, such as the fixed cost of adapting the seller’s output to the buyer’s production needs.

This link function is also a fixed point problem. The gross profits from a potential match,  $\pi()$ , determine link probabilities according to equation (8), and the link probabilities determine gross profits via the backward and forward fixed points in equations (6) and (7).

The general equilibrium of the model can be solved by a simple nested fixed point algorithm. (i) Start with a guess for the link function  $l()$ . (ii) Solve for  $P(\lambda)$  and  $S(\lambda)$  using the backward and forward fixed points in equations (6) and (7) sequentially. (iii) Calculate gross profits for all potential matches using equation (5). (iv) Calculate the share of seller-buyer pairs  $(\lambda, \lambda')$  that match according to equation (8). (v) Go back to step (ii) until the link function converges. We do not have a formal proof of existence and uniqueness. In practice, however, the nested fixed point problem is numerically well-behaved and always converges to the same solution irrespective of the chosen starting values.

## 4.5 Discussion

We conclude the exposition of the model by discussing some key implications and features. We start by considering the role of each dimension of firm heterogeneity in determining equilibrium outcomes on its own. Conditioning on relationship capability, firms with higher productivity have lower marginal costs, lower prices and higher profits from a match with any given buyer, see equation (3). As a result, higher productivity firms match with more buyers, see equation (8), and have greater sales (market share) conditional on a match, see equation (5). Larger total sales and input purchases make higher productivity firms more attractive partners for upstream firms, see equation (5). The increased number of upstream suppliers contributes to an additional reduction in marginal cost through the firm’s input price index, see equation (6).

Conditioning on productivity, firms with better relationship capability, lower  $F$ , are able to match with more buyers, see equation (8), and as a result have greater sales and greater input purchases. As with higher productivity, the greater input demand makes these firms relatively attractive to upstream suppliers, and the greater number of suppliers lowers their marginal cost of production through the input price index. The lower marginal cost results in greater sales (market share) to any given buyer.

Thus considered by itself, either higher productivity or better relationship capability leads to higher sales through both the extensive margin of more downstream buyers and the intensive margin of greater sales per buyer.



Several features of the model grant it analytical and quantitative tractability, as well as transparency in illustrating the main mechanisms. First, we consider a unit continuum of firms in the economy. This implies that individual sellers take other sellers’ prices and all buyers’ input price indices as given when deciding whether to match with a particular buyer and how much to sell to that buyer.

Separately, we focus on the costs that sellers incur to match with buyers, and assume that buyers do not face corresponding costs of matching with suppliers. Even with this assumption, in equilibrium the number of both suppliers and buyers varies across firms. This choice lends tractability because firms make separable sales decisions with respect to different buyers and do not internalize the effect of their match decisions on buyers’ input demand. It also avoids the well-known problem of interdependence of sourcing decisions in frameworks where buyers choose suppliers, see Antràs et al. (2017).

Finally, the model focuses on the domestic production network and does not directly consider the role of exports and imports, both of which are important in the Belgian context. Exports are implicitly included in final demand, even though this almost surely understates the importance of firm-to-firm sales, as almost all export sales are to firms rather than consumers. Similarly, while we do not model imports, they can be added to production without changing the implications for firm outcomes.<sup>21</sup>

## 5 Estimation and Results

This section provides a model-based assessment of the origins of firm heterogeneity. Specifically, we exploit the Belgian production network data to parameterize the model above, allowing for heterogeneity in both productivity and relationship capability across firms. We then estimate the model under alternative scenarios to evaluate the quantitative importance of each firm attribute.

### 5.1 Simulated Method of Moments

The general-equilibrium model is estimated by simulated method of moments (SMM). We assume that firm productivity  $z$  and relationship capability  $F$  are distributed joint log-normal with expectations  $\mu_{\ln z} = 0$  and  $\mu_{\ln F}$ , standard deviations  $\sigma_{\ln z}$  and  $\sigma_{\ln F}$ , and correlation coefficient  $\rho$ .<sup>22</sup> In sum, there are four unknown parameters to be estimated,

<sup>21</sup> See Bernard et al. (2018b) for a static model of a domestic production network with imports in the production function and idiosyncratic match-specific shocks.

<sup>22</sup> The mean of  $\ln z$  is not identified, and it is therefore normalized to zero. This normalization is appealing on conceptual grounds. Consider a shift in the productivity distribution, such that productivity increases

$$\Upsilon = \{\sigma_{\ln z}, \mu_{\ln F}, \sigma_{\ln F}, \rho\}.$$

In addition to the unknown parameters, information is needed on  $\alpha$  (labor cost share),  $\mu$  (markup), and  $X$  (aggregate income).  $\alpha$  is constructed by dividing labor costs by total costs for each firm, and then taking the simple average across firms.  $\mu$  is computed by dividing sales by total costs for each firm, and then taking the simple average across firms.  $X$  is inferred from sales going out of the network, i.e.  $X = \sum_i S_i - \sum_i \sum_{j \in \mathcal{C}_i} m_{ij}$ . The idiosyncratic matching cost  $\epsilon$  is assumed log-normal, with mean  $\mu_{\ln \epsilon} = 0$  and standard deviation  $\sigma_{\ln \epsilon}$ . The standard deviation is chosen so that the objective function is smooth in the parameters of the model.<sup>23</sup> Table 4 summarizes the parameters of the model, their definitions, and the values assigned to them.

We choose seven moments in the data to estimate  $\Upsilon$ . While all moments jointly pin down all unknown parameters in general equilibrium, there is an intuitive mapping between them. First, the mean log number of customers across firms,  $mean(\ln n_i^c)$ , helps identify the mean of the relationship costs. Second, the variance of log number of customers,  $var(\ln n_i^c)$ , and the variance of network sales,  $var(\ln S_i^{net})$ , together identify the variances of productivity and relationship costs. Third, the slope coefficient from the regression of average market share on the number of buyers,  $\ln \bar{\delta}_i = \alpha + \beta \ln n_i^c + \varepsilon_i$ , helps identify the correlation coefficient  $\rho$  (see Figure 3): implicitly, a smaller (or more negative) slope coefficient suggests that firms with low relationship costs and therefore high  $\ln n_i^c$  are relatively less productive and thus have lower  $\ln \bar{\delta}_i$ . Finally, we include the contribution of key decomposition margins in Table 2: the number of customers ( $\ln n_i^c$ ), average customer capability ( $\ln \bar{\theta}_i$ ), and the customer interaction term ( $\ln \Omega_i^c$ ). Intuitively, the contribution of these margins to firm size dispersion inform the role of relationship costs versus productivity in determining firm size.<sup>24</sup> Collecting the targeted empirical moments in vector  $x$  and the corresponding simulated moments in vector  $x^s(\Upsilon)$ , the SMM estimates for  $\Upsilon$  solve:

$$\arg \min_{\Upsilon} (x - x^s(\Upsilon))' (x - x^s(\Upsilon)).$$

We obtain standard errors by bootstrapping these estimates, see Appendix F for details.<sup>25</sup>

---

for all firms. While this would lower prices and increase welfare, it would not change firms' market shares or the network structure of the economy that are of interest to us.

<sup>23</sup> If the dispersion in  $\epsilon$  is small relative to the dispersion in  $z$  and  $F$ , the share of links for some  $(\lambda, \lambda')$  pairs will be close to zero. This complicates the SMM estimation using standard gradient-based methods, as the objective function is no longer smooth in the parameters of the model. In practice, we set  $\sigma_{\ln \epsilon} = 4$ , which makes the problem sufficiently smooth, similarly to the scale factor in the logit-smoothed AR simulator (McFadden, 1989). Other choices of  $\sigma_{\ln \epsilon}$  do not significantly improve the fit of the model.

<sup>24</sup> There are six margins in total, which by construction sum to one. The final demand margin is omitted as a targeted moment because final demand is measured as the difference between total sales and network

Table 4: Model Parameters

Parameter	Definition	Value	Source
$\alpha$	Labor cost share	.24	Mean of $\frac{w_i L_i}{w_i L_i + M_i}$
$\mu$	Markup	1.24	Mean of $\frac{S_i}{w_i L_i + M_i}$
$X$	Aggregate final demand	€470bln	$\sum_i S_i - \sum_i \sum_{j \in C_i} m_{ij}$
$\sigma_{\ln \epsilon}$	Pair matching cost dispersion	4	

## 5.2 Results

The estimated parameters are summarized in the first four rows of column (2) in Table 5. A striking result is the large positive correlation between productivity and relationship costs ( $\rho$ ). In other words, firms that are more efficient at converting inputs into outputs on the production side have lower relationship capability in matching with buyers on the sales side. In addition, the standard deviation of log relationship costs is an order of magnitude larger than the standard deviation of log productivity. To put matching frictions into perspective, we calculate the ratio of relationship costs ( $F\epsilon$ ) to firm-to-firm sales ( $m$ ) for successful matches in the economy. The mean of this ratio is 0.07, i.e. average relationship costs account for 7 percent of relationship sales. The mean of this ratio across all potential matches is orders of magnitude higher (222,760), consistent with the observation that the production network is sparse because many possible seller-buyer matches are not profitable.

Columns (1) and (2) of Table 5 report targeted and untargeted moments in the data and in the simulated model with the estimated parameters. The second panel demonstrates that the model hits targeted moments well as expected, with the exception of the contribution of the customer interaction term to the firm size decomposition. In particular, it replicates the negative relationship between the number of customers ( $n_i^c$ ) and the average market share per customer ( $\bar{\delta}_i$ ) across sellers (see Figure 3). The model does not only hit the moments of the distribution of network sales, it also hits the full distribution relatively well. Figure 5 shows the (within-industry) distribution of network sales according to the estimated model and the data.

The rest of Table 5 presents non-targeted moments. On the downstream side, the third panel shows that the model slightly overpredicts the variance of total sales. It also matches relatively well the negative relationship between the 90th/50th/10th sales percentiles and

---

sales, and it may therefore include sales to firms outside the observed network (e.g., foreign firms).

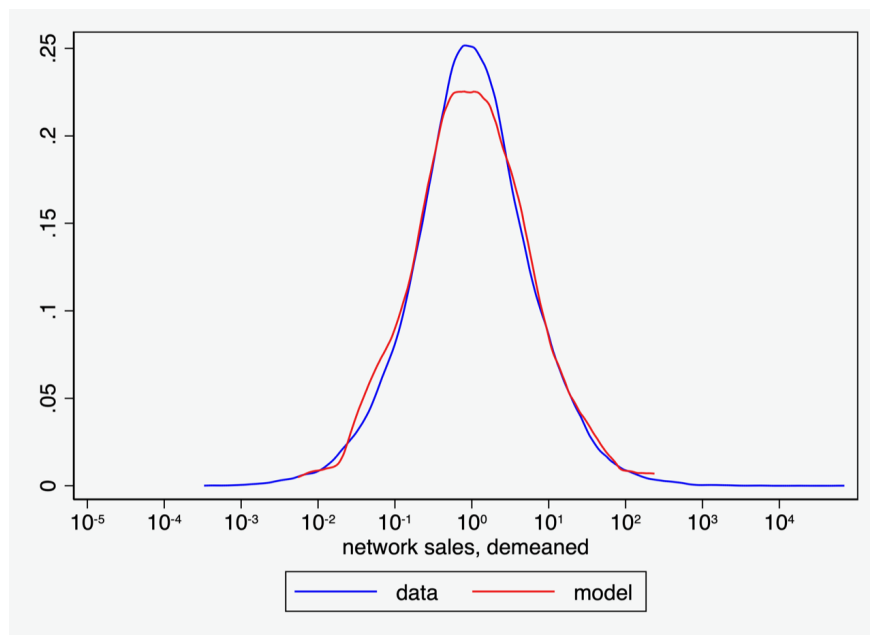
<sup>25</sup> We use the equally weighted minimum distance estimator. We have also estimated the model weighting the moments by the inverse of the variance-covariance matrix of the moments, which yielded similar results compared to the baseline.

Table 5: SMM Model Fit

	Data	Estimated models			
	(1)	(2) Baseline	(3) No F	(4) No Z	(4) No rho
<i>Estimated parameters:</i>					
$\mu_{\ln F}$		18.11 (.02)	19.18 (.02)	19.65 (.02)	18.21 (.02)
$\sigma_{\ln z}$		.24 (.00)	.13 (.00)		.07 (.00)
$\sigma_{\ln F}$		2.23 (.01)		1.48 (.01)	1.29 (.01)
$\rho$		.86 (.00)			.01
<i>Targeted moments:</i>					
$mean(\ln n_i^c)$	-8.12	-8.12	-8.27	-8.14	-8.12
$var(\ln n_i^c)$	1.87	1.86	0.81	2.20	1.92
$var(\ln S_i^{net})$	3.12	3.12	3.37	2.78	3.08
$\beta$ from $\ln \bar{\delta}_i = \alpha + \beta \ln n_i^c + \epsilon_i$	-.11	-.10	1.10	.15	.28
Decomp.: # cust.	.51	.52	.49	.89	.76
Decomp.: Avg cust. capability	.05	.01	.01	.03	.02
Decomp.: Customer interaction	.25	-.04	-.04	-.05	-.06
<i>Non-targeted moments:</i>					
<u>Downstream</u>					
$var(\ln S_i)$	1.73	2.08	1.61	.62	.90
$var(\ln \text{Value Added}/\text{Worker}_i)$	.62	.71	.15	.54	.42
$\beta$ from $\ln m_i^k = \alpha + \beta \ln n_i^c + \epsilon_i$	-.25/-.28/-.29	-.14/-.15/-.17	1.07/1.06/1.02	.15/.13/.11	.27/.25/.19
Downstream degree assort.	-.05	-.04	-.07	-.02	-.03
<u>Upstream</u>					
$var(\ln M_i^{net})$	2.12	2.08	1.61	.62	.90
$var(\ln n_i^s)$	.60	.41	.38	.12	.18
Upstream degree assort.	-.18	-.18	-.07	-.15	-.15

Notes: The number of customers in column (1),  $n_i^c$ , is normalized relative to the number of firms in the final sample.  $\bar{\delta}_i$  is the geometric mean of the market share  $\delta_{ij} = m_{ij}/M_j$  for seller  $i$  across its buyers  $j$ . Downstream degree assortativity refers to  $\beta$  from the regression for  $i$ 's customers  $\ln \text{Mean}_i n_j^s = \alpha + \beta \ln n_i^c + \epsilon_i$ . Upstream degree assortativity refers to  $\beta$  from the regression for  $j$ 's suppliers  $\ln \text{Mean}_j n_i^c = \alpha + \beta \ln n_j^s + \epsilon_j$ .  $\ln m_i^k$  is the  $k$ th (10th/50th/90th) percentile of log bilateral sales,  $\ln m_{ij}$ , for seller  $i$  across its customers  $j$ . The three decomposition moments refer to the contribution of different margins to firm size dispersion from Section 3. All variables in column (1) except  $mean(\ln n_i^c)$  are demeaned by NACE 4-digit industry averages. Bootstrapped standard errors in parentheses.

Figure 5: Density of network sales.



Note: The figure shows the density of network sales across firms in the data and in the model. The variable is demeaned by NACE 4-digit industry averages.

out-degree documented in Figure 3, as well as the pattern of negative degree assortativity downstream in Figure 4.

The bottom panel of Table 5 reports untargeted upstream moments. These moments are interesting because the model emphasizes sellers’ choice of buyers downstream but remains silent about firms’ choice of partners upstream. The model does a good job matching the variance of input purchases from the network, as well as the variance of the number of suppliers. There is also close correspondence between the negative upstream assortativity in the model and in the data: Buyers with more suppliers have suppliers who on average have fewer customers in the network.

We also evaluate to what extent the model fits the observed dispersion in labor productivity. In this class of models, value added per production worker,  $(S - M)/l$ , is constant across all firms within an industry (Hsieh and Klenow, 2009). In our model, however, a firm’s employment is the sum of production workers and “marketing” workers, i.e. workers allocated to relationship building, such that total employment is  $L = l + n^c F$ . Value added per worker therefore varies across firms and is increasing in value added per marketing worker.<sup>26</sup>

<sup>26</sup> Value added per worker is  $\frac{S-M}{L} = \left(\frac{l+n^c F}{S-M}\right)^{-1} = \left(\frac{l}{S-M} + \frac{n^c F}{S-M}\right)^{-1}$ , where the first term is constant across firms and the second term is the inverse of value added per marketing worker.

In equilibrium, firms with high productivity and/or high relationship capability have higher value added per marketing worker, and therefore also greater labor productivity. Table 5 confirms that the estimated model produces significant variance in log labor productivity, although dispersion in the model is slightly higher (.71 versus .62).

### 5.3 Restricted Models

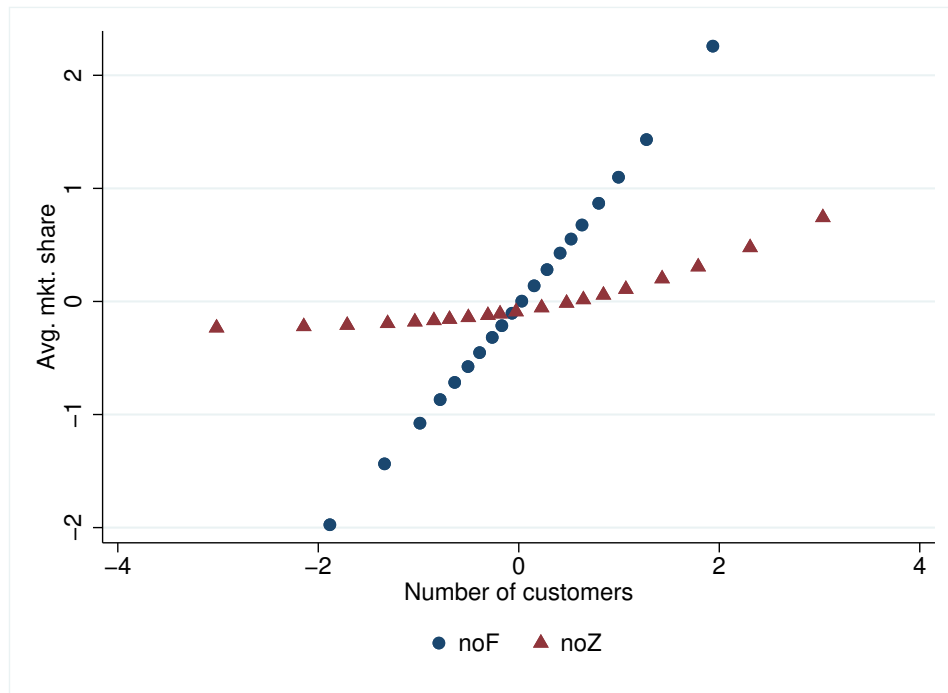
We next illustrate the need for two firm attributes in order to rationalize observed empirical patterns, by estimating a model with heterogeneity in either (i) productivity (“no F”) or (ii) relationship capability (“no Z”) but not both. Under (i), there are two parameters to estimate,  $\Upsilon = \{\sigma_{\ln Z}, \mu_{\ln F}\}$ , and we use the same moments to identify  $\Upsilon$ . Under (ii), the parameters to estimate are  $\Upsilon = \{\mu_{\ln F}, \sigma_{\ln F}\}$ .

The estimated parameters and fit of these two restricted models are summarized in columns (3) and (4) of Table 5. Both restricted models are unable to generate the negative correlation between the average market share ( $\bar{\delta}_i$ ) and the number of customers across sellers (the  $\beta$  coefficient from the regression  $\ln \bar{\delta}_i = \alpha + \beta \ln n_i^c + \epsilon_i$ ). Figure 6 plots this relationship according to the estimated restricted models. In both cases, the model generates the opposite pattern to the empirical regularity in Figure 3.

In addition, the single factor models do a relatively poor job in other dimensions. In the “noF” case, we more or less match dispersion in network sales, but this comes at the expense of not matching the variance in the number of customers or in value added per worker. In the “noZ” case, we match dispersion in the number of customers, but dispersion in total sales is too small, and the contribution of the number of customers in the firm size decomposition is too high. Both restricted models underestimate the variances of input purchases and of the number of suppliers, and counterfactually imply that bilateral sales at different customer percentiles increase rather than decrease with the number of customers.

Finally, we estimate a model with heterogeneity in both productivity and relationship capability, but where the correlation between them set to zero,  $\rho = 0$  (“no rho”). The last column of Table 5 shows that the restricted model produces a positive correlation between average market share ( $\bar{\delta}_i$ ) and the number of customers across sellers (the slope coefficient is .28), far from the slightly negative coefficient in the data. Furthermore, it does poorly for many non-targeted moments: the restricted model generates significantly less heterogeneity in both log sales, log input purchases and log number of suppliers. This result highlights that a data-generating process with unrelated  $Z$  and  $F$  is inconsistent with our data.

Figure 6: Restricted Models



Notes: The figure shows the binned scatterplot of the number of customers and the average market share in buyers' input purchases across sellers in the restricted models, where  $\text{var}(F) = 0$  ("noF") or  $\text{var}(z) = 0$  ("noZ").

## 5.4 Sensitivity

Next, we evaluate the sensitivity of the estimates to the vector of estimation moments. We use the methodology from Andrews et al. (2017). Specifically, we ask how sensitive the parameter estimate  $\rho$  (the correlation between productivity and relationship costs) is to perturbations of the various moments of the data. We consider perturbations that are additive shifts of the moment functions due to either misspecification of  $x^s$  ( $\Upsilon$ ) or measurement error in the empirical moments  $x$ . Andrews et al. (2017) show that sensitivity can be summarized by the matrix  $\Lambda = (S'WS)^{-1}S'W$ , where  $S$  is the matrix of partial derivatives of  $x^s$  ( $\Upsilon$ ) evaluated at the true value  $\Upsilon_0$  (see their Proposition 2).  $W$  is the method of moments weighting matrix, which in our case is the identity matrix.

Figure 7 plots the column of the estimated  $\Lambda$  corresponding to the parameter estimate  $\rho$ . The plot shows one-hundredth of the value of sensitivity of the parameter  $\rho$  with respect to the vector of estimation moments. The values are scaled by the standard deviation of the moments, so the values can be interpreted as the effect of a one standard deviation change in the moment on the parameter  $\rho$ . The plot largely confirms our intuition about identification. First, the slope coefficient  $\beta$  (from the regression of average market share on the number of customers) is the moment that matters the most, while the other moments are less relevant for this particular parameter. Furthermore, the direction of sensitivity is also in line with our expectations: A steeper slope coefficient has a negative impact on  $\rho$ , i.e. we get a lower positive correlation between  $Z$  and  $F$  as the slope becomes steeper.

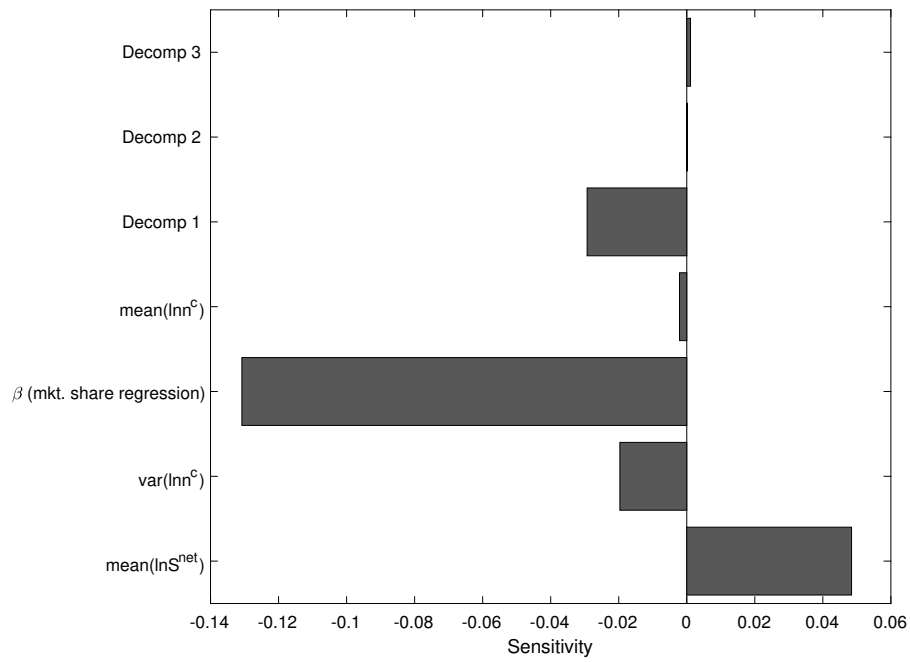
## 5.5 A Counterfactual

We end this section by quantifying the role of firm heterogeneity in productivity and relationship costs for aggregate outcomes. We do so by performing a simple counterfactual experiment: a common 50 percent reduction in relationship costs across all firms in the economy (i.e. a reduction in  $\mu_{\ln F}$ ). We do so both in the baseline estimated model and in a restricted model with no correlation between productivity and relationship costs, in order to illustrate the importance of the latter.

The simulation shows that real wages increase by 17% in the baseline and 12% in the model with no correlation, which implies that the welfare gains are 42% higher with correlation than without. Figure 8 shows a binned scatterplot of the counterfactual change in the log number of customers on the vertical axis against log productivity ( $\ln Z$ ) on the horizontal axis. The solid circles refer to the baseline counterfactual, while the triangles denote the no-correlation counterfactual. In both versions of the model, lower relationship costs generate many new customers per firm. In the baseline model, the increase is relatively similar across

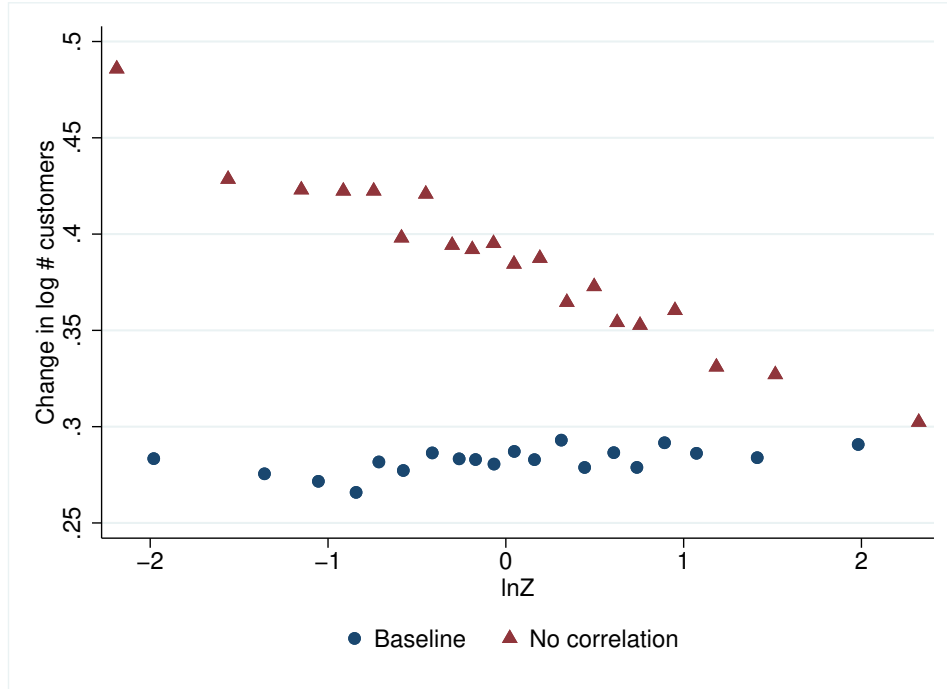


Figure 7: Sensitivity of the parameter estimate  $\rho$ .



Note: The plot shows one-hundredth of the value of sensitivity of the parameter  $\rho$  with respect to the vector of estimation moments. The values are scaled by the standard deviation of the moments, so the values can be interpreted as the effect of a one standard deviation change in the moment on the parameter  $\rho$ .

Figure 8: Counterfactual: 50% Reduction in Relationship Costs



Notes: The figure shows the binned scatterplots of the change in the log number of customers in the baseline counterfactual (circle markers) and in the counterfactual with no correlation between  $Z$  and  $F$  (triangle markers), against  $\ln Z$ .

firms with different productivity levels. In the no-correlation model, however, low productivity firms gain many more customers relative to high productivity firms (50% increase versus 30% in the tails of the distribution).<sup>27</sup> Recall that in the baseline, low productivity firms are also low relationship costs firms, such that the drop in  $\mu_{\ln F}$  will not have a large impact on their connections. In the no-correlation model, by contrast, low productivity firms are constrained because they face similar relationship costs as other firms. This explains the difference in slopes between the two models.

## 6 Conclusion

This paper quantifies the origins of firm size heterogeneity when firms are interconnected in a production network. We report three stylized facts from the production network data that motivate the subsequent analysis and model. First, the enormous dispersion in sales across firms is also observed in the production network in terms of the number of firm-to-

<sup>27</sup> The increase in the total number of firm-to-firm connections in the network is relatively similar in the two models: 26% in the baseline and 28% in the no-correlation model, respectively.

firm connections and the value of pairwise sales. Second, firms with higher sales have more customers but lower average sales per customer and lower market shares (of input purchases) among their customers. Finally, there is negative degree assortativity between buyers and suppliers, i.e. sellers with more customers match with customers who have fewer suppliers on average.

Taken together, these facts present challenges to many existing models of firm heterogeneity. The large variation in sales across firms within an industry is intuitively related to variation in the number of customers: larger firms have more customers. However, larger firms also sell less to their customers. Models that emphasize heterogeneity in productivity across firms cannot explain these facts simultaneously. In particular, such models imply that firms with more customers should also sell more to each of their customers and have higher rather than lower market shares.

We confirm the importance of the production network in a decomposition of the variance of firm sales within narrowly defined industries. 81% of the variation in firm sales is associated with the downstream component, and most of that is due to variation in the number of customers. The upstream component contributes 18%, and variation in the share of sales outside the domestic production network plays a minor role at 1%. These findings imply that trade in intermediate goods and the number of firm-to-firm connections are essential to understanding firm performance and, consequently, aggregate outcomes.

Motivated by the stylized facts and decomposition results, we develop a quantitative general-equilibrium model of firm-to-firm trade. In the model, firms differ along two dimensions – productivity and relationship capability – defined respectively as production efficiency and (the inverse of) the fixed cost of matching with a customer. Suppliers match with customers if the gross profits from the match exceed the supplier-specific fixed matching cost. Marginal costs, employment, prices, and sales are endogenous outcomes because they depend on the outcomes of all other firms in the economy. A link between two firms increases the total sales of both the seller and the buyer; for the seller this occurs mechanically because it gains a customer, while for the buyer this arises because a larger supplier base lowers the marginal cost of production.

We estimate parameters of the model using simulated method of moments. The results reveal a strong negative correlation between the two firm characteristics: Firms with higher productivity have lower relationship capability. Importantly, both dimensions of firm heterogeneity are necessary to match the data. Shutting down one at a time results in poor model fit, including the inability to replicate the negative relationship between the number of customers and average sales per customer.

Our results challenge current understanding of the sources of firm size heterogeneity,

and point to important areas for future research on the negative relationship between firm productivity and relationship capability. While we make progress in matching the relative importance of upstream and downstream factors in firm success, there is room for new models to better fit these features of the production network. In addition, research is needed to examine the factors that lead to a negative relationship between productivity and relationship capability across firms. One promising avenue for further work is examining span of control issues inside the firm and the allocation of resources to improving productivity versus acquiring more customers.

## References

- Abowd, J.M., F. Kramarz, and D.N. Margolis**, “High Wage Workers and High Wage Firms,” *Econometrica*, 1999, *67* (2), 251–333. 3, 3.1, 14, C.2
- Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi**, “The Network Origins of Aggregate Fluctuations,” *Econometrica*, 2012, *80* (5), 1977–2016. 1
- Allen, Treb, Costas Arkolakis, and Xiangliang Li**, “On the Existence and Uniqueness of Trade Equilibria,” Technical Report, Working paper 2016. D.3
- Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro**, “Measuring the Sensitivity of Parameter Estimates to Estimation Moments,” *The Quarterly Journal of Economics*, 2017, *132* (4), 1553–1592. 5.4
- Antràs, P., T.C. Fort, and F. Tintelnot**, “The Margins of Global Sourcing: Theory and Evidence from US Firms,” *American Economic Review*, September 2017, *107* (9), 2514–64. 1, 4.5
- Arkolakis, Costas**, “Market Penetration Costs and the New Consumers Margin in International Trade,” *Journal of Political Economy*, 2010, *118* (6). 2.2
- Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen**, “The Fall of the Labor Share and the Rise of Superstar Firms,” *Quarterly Journal of Economics*, 2020, *135* (2), 645–709. 1
- Baqae, David Rezza and Emmanuel Farhi**, “The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten’s Theorem,” *Econometrica*, 2019, *87* (4), 1155–1203. 1
- **and –**, “Productivity and Misallocation in General Equilibrium,” *Quarterly Journal of Economics*, 2020, *135* (1), 105–163. 1
- Bender, S., D. Card, N. Bloom, J. Van Reenen, and S. Wolter**, “Management Practices, Workforce Selection and Productivity,” *Journal of Labor Economics*, January 2018, *36* (1), S371–S409. 1
- Bernard, A., A. Moxnes, and Y.U. Saito**, “Production Networks, Geography and Firm Performance,” *Journal of Political Economy*, 2019. 1, 8, 12

- Bernard, A.B., J.B. Jensen, S.J. Redding, and P.K. Schott**, “The Empirics of Firm Heterogeneity and International Trade,” *Annual Review of Economics*, 2012, 4 (1), 283–313. 1
- Bernard, Andrew B. and Andreas Moxnes**, “Networks and Trade,” *Annual Review of Economics*, August 2018, 10, 65–85. 1
- , – , and **Karen Helene Ulltveit-Moe**, “Two-sided Heterogeneity and Trade,” *Review of Economics and Statistics*, 2018, 100 (3). 1, 8, 2.2, 12, 4.4
- , **Emmanuel Dhyne, Glenn Magerman, Kalina Manova, and Andreas Moxnes**, “The Origins of Firm Heterogeneity: A Production Network Approach,” Technical Report 25441, NBER Working paper 25441 2018. 21
- Bloom, Nick, Kalina Manova, John Van Reenen, Stephen Sun, and Zhihong Yu**, “Managing Trade: Evidence from China and the US,” *NBER Working Paper 24718.*, 2017. 1
- Boehm, Johannes and Ezra Oberfield**, “Misallocation in the Market for Inputs: Enforcement and the Organization of Production,” *Quarterly Journal of Economics*, 2020, 135 (4), 2007–2058. 1
- Bøler, Esther Ann, Andreas Moxnes, and Karen Helene Ulltveit-Moe**, “R&D, International Sourcing, and the Joint Impact on Firm Performance,” *American Economic Review*, 2015, 105 (12). 1
- Bonhomme, Stéphane**, “Econometric Analysis of Bipartite Networks,” in B. Graham and A. de Paula, eds., *The Econometric Analysis of Network Data*, Academic Press, 2020. 32
- Card, D., J. Heining, and P. Kline**, “Workplace Heterogeneity and the Rise of West German Wage Inequality\*,” *The Quarterly Journal of Economics*, 2013, 128 (3), 967–1015. C.2, 34
- Correia, Sergio**, “Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator,” Technical Report, Working Paper 2016. 1
- Dhyne, E., G. Magerman, and S. Rubinova**, “The Belgian Production Network 2002–2012,” *National Bank of Belgium Working Paper*, 2015, 288. 2.1, 8, A.1
- di Giovanni, J., A.A. Levchenko, and I. Mejean**, “Firms, Destinations, and Aggregate Fluctuations,” *Econometrica*, 07 2014, 82 (4), 1303–1340. 1

- Eaton, J., D. Jinkins, J. Tybout, and D.Y. Xu**, “Two-sided Search in International Markets,” 2016. NBER. 1
- , **S. Kortum, and F. Kramarz**, “Dissecting Trade: Firms, Industries, and Export Destinations,” *American Economic Review*, May 2004, *94* (2), 150–154. 3.1
- Eaton, Jonathan, Samuel Kortum, and Francis Kramarz**, “Firm-to-Firm Trade: Imports, Exports, and the Labor Market1,” Technical Report, Working paper 2018. 1, 2.2
- Fieler, Ana Cecília, Marcela Eslava, and Daniel Yi Xu**, “Trade, Quality Upgrading, and Input Linkages: A Theory with Evidence from Colombia,” *American Economic Review*, 2018, *108*. 1
- Fitzgerald, D., S. Haller, and Y. Yedid-Levi**, “How Exporters Grow,” Staff Report 524, Federal Reserve Bank of Minneapolis January 2016. 1
- Foster, L., J. Haltiwanger, and C. Syverson**, “The Slow Growth of New Plants: Learning about Demand?,” *Economica*, 2016, *83* (329), 91–129. 1
- Freund, C. and M.D. Pierola**, “Export Superstars,” *The Review of Economics and Statistics*, 2015, *97* (5), 1023–1032. 1
- Gabaix, Xavier**, “The Granular Origins of Aggregate Fluctuations,” *Econometrica*, 2011, *79*, 733–772. 1
- Gaubert, Cecile and Oleg Itskhoki**, “Granular Comparative Advantage,” *Journal of Political Economy*, March 2021, *129* (3), 871–939. 1
- Gibrat, Robert**, *Les inégalités économiques*, Librairie du Recueil Sirey, Paris, 1931. 1
- Goldberg, P., A.K. Khandelwal, N. Pavcnik, and P. Topalova**, “Imported Intermediate Inputs and Domestic Product Growth: Evidence from India,” *The Quarterly Journal of Economics*, 2010, *125* (4), 1727–1767. 1
- Holmstrom, Bengt and Paul Milgrom**, “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design,” *Journal of Law, Economics, and Organization*, 1991, *7*, 24–52. 1
- Hong, Fuhai, Tanjim Hossain, John A. List, and Migiwa Tanaka**, “Testing the Theory of Multitasking: Evidence From a Natural Field Experiment in Chinese Factories,” *International Economic Review*, 2018, *59* (2), 511–536. 4

- Hopenhayn, H.A.**, “Entry, Exit and Firm Dynamics in Long Run Equilibrium,” *Econometrica*, 1992, *60*, 1127–1150. 1
- Hottman, C. J., S.J. Redding, and D. E. Weinstein**, “Quantifying the Sources of Firm Heterogeneity,” *Quarterly Journal of Economics*, 2016, *131* (3), 1291–1364. 1, 3.1
- Hsieh, Chang-Tai and Peter J. Klenow**, “Misallocation and Manufacturing TFP in China and India,” *The Quarterly Journal of Economics*, 2009, *124* (4). 5.2
- Jovanovic, Boyan**, “Selection and the Evolution of Industry,” *Econometrica*, May 1982, *50* (3), 649–70. 1
- Lim, Kevin**, “Endogenous Production Networks and the Business Cycle,” Technical Report 2018. 1, 2.2, 12, 4.4
- Luttmer, Erzo G. J.**, “Selection, growth and the size distribution of firms,” *The Quarterly Journal of Economics*, 2007, *78* (3), 1042–1068. 1
- , “On The Mechanics of Firm Growth,” *The Review of Economic Studies*, 2011, *122* (3), 1103–1144. 1
- Magerman, Glenn, Karolien De Bruyne, Emmanuel Dhyne, and Jan Van Hove**, “Heterogeneous Firms and the Micro Origins of Aggregate Fluctuations,” Technical Report, KU Leuven 2016. 1
- Manova, Kalina, Shang-Jin Wei, and Zhiwei Zhang**, “Firm Exports and Multinational Activity Under Credit Constraints,” *The Review of Economics and Statistics*, July 2015, *97* (3), 574–588. 1
- McFadden, Daniel**, “A method of simulated moments for estimation of discrete response models without numerical integration,” *Econometrica*, 1989, *57*, 995–1026. 23
- Melitz, Marc.**, “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry,” *Econometrica*, 2003, *71* (6), 1695–1725. 1
- Melitz, Marc and Stephen Redding**, “New Trade Models, New Welfare Implications,” *American Economic Review*, 2015, *105* (3), 1105–1146. 1
- Oberfield, Ezra**, “A Theory of Input-Output Architecture,” *Econometrica*, 2018, *86* (2), 559–589. 1
- Pavcnik, Nina**, “Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants,” *The Review of Economic Studies*, 2002, *69*, 245–276. 1



**Prescott, Edward C. and Michael Visscher**, “Organizational Capital,” *The Journal of Political Economy*, June 1980, 88 (3), 446–461. 1

**Sutton, John**, “Market Share Dynamics and the "Persistence of Leadership" Debate,” *American Economic Review*, March 2007, 97 (1), 222–241. 1

**Syverson, C.**, “What Determines Productivity?,” *Journal of Economic Literature*, 2011, 49 (2), 326–365. 1

**Tintelnot, Felix, Ken Kikkawa, Magne Mogstad, and Emmanuel Dhyne**, “Trade and Domestic Production Networks,” *Review of Economic Studies*, 2021, 88 (2), 643–668. 1, E.6

# Appendix

## A Data Sources

### A.1 Data sources

The empirical analysis draws on four micro-level datasets at the National Bank of Belgium (NBB), covering the period 2002-2014: (i) the NBB B2B Transactions Dataset, (ii) annual accounts from the Central Balance Sheet Office supplemented by (iii) VAT declarations, and (iv) the Crossroads Bank. Firms are identified by a unique enterprise number, which is common across all databases and allows for unambiguous merging.

**Firm-to-firm relationships** The confidential NBB B2B Transactions Dataset contains the value of yearly sales among all VAT-liable Belgian enterprises for the years 2002 to 2014, and is based on the VAT listings collected by the tax authorities. All firms that deliver goods or services as an economic activity, on a regular and independent basis, are VAT liable. These enterprises have to charge an ad valorem tax on their sales and can recover VAT paid on their purchases. This includes foreign companies with a branch in Belgium and firms whose securities are officially listed in Belgium. Enterprises that only perform financial transactions, medical or socio-cultural activities such as education are exempt. Firms with sales less than €15,000 can choose to be exempt from the VAT liabilities. The standard VAT rate in Belgium is 21%, but for some goods a reduced rate of 12% or 6% applies.<sup>28</sup>

At the end of every calendar year, all VAT-liable enterprises have to file a complete listing of their Belgian VAT-liable customers over that year. An observation in this dataset refers to the value of sales in euro by enterprise  $i$  to enterprise  $j$  within Belgium, excluding the VAT due on these sales. The reported value is the sum of invoices from  $i$  to  $j$  in a given calendar year. Whenever this aggregated value is €250 or greater, the relationship has to be reported. Together with the VAT declarations, these listings are the basis for the VAT amount due per tax period. Fines for late or incomplete reporting ensure a very high quality of the data. Note that each relationship is directed, as the observation from  $i$  to  $j$  is different from an observation from  $j$  to  $i$ . The dataset thus covers both the extensive and the intensive margins of the Belgian production network. A detailed description of the collection and cleaning of this dataset is given in Dhyne et al. (2015). We refer to that paper for further details.

**Firm characteristics** We extract information on enterprises' annual accounts from the

---

<sup>28</sup> See [ec.europa.eu/taxation\\_customs](http://ec.europa.eu/taxation_customs) for a complete list of rates. These rates did not change over our sample period.

Central Balance Sheet Office at the NBB for the years 2002 to 2014. Enterprises above a certain size threshold have to file annual accounts at the end of their fiscal year. We retain information on the enterprise identifier (VAT ID), turnover (total sales in euro, code 70 in the annual accounts), input purchases (total material and services inputs in euro and net changes in input stocks, codes 60+61), labor cost (total cost of wages, social securities and pensions in euro, code 62), and employment (average number of full-time equivalent (FTE) employees, code 9087). We annualize all flow variables from fiscal years to calendar years by pro-rating the variables on a monthly basis.<sup>29</sup>

Enterprises below a size threshold can report abbreviated annual accounts. These firms still report labor cost and employment, but are not required to report turnover or input purchases. For these small enterprises, we supplement information on turnover and input expenditures from their VAT declarations. All VAT-liable enterprises have to file periodic VAT declarations with the tax administration. The VAT declaration contains the total sales value (including domestic sales and exports), the VAT amount charged on those sales (both to other enterprises and to final consumers), the total amount paid for inputs sourced (including domestic and imported inputs), and the VAT paid on those input purchases. This declaration is due monthly or quarterly depending on firm size, and it is the basis for the VAT due to the tax authorities every period. We aggregate the VAT declarations to the annual frequency to match the annual accounts.

We obtain information on the main economic activity of each enterprise at the NACE 4-digit level from the Crossroads Bank of Belgium for the years 2002 to 2014. We concord NACE codes over time to the NACE Rev. 2 version to deal with changes in the NACE classification over our panel from Rev. 1.1 to Rev. 2. Table 6 lists industry groups at the NACE 2-digit level. Finally, we also retrieve the postal code of the main establishment of the firm from the Crossroads Bank of Belgium. There are 589 municipalities in Belgium, which correspond to the Local Administrative Units level 2.

Since total sales and network sales come from different datasets, we enforce:

$$S_i = Snet_i + fd_i$$

where final demand  $fd_i \geq 0$ . For observations where  $Snet_i$  from the NBB B2B transactions dataset is larger than  $S_i$ , we replace  $S_i = \max(S_i, Snet_i)$ . This amounts to 6% of observations in our estimation sample. The median value of  $Snet_i$  that overshoots  $S_i$  is 17% of the value reported in the annual accounts for these observations, before we correct them.

---

<sup>29</sup> In our data, 78% of firms have annual accounts that coincide with calendar years, while 98% of firms have fiscal years of 12 months.

Table 6: NACE classification of industry groups (NACE Rev 2)

NACE Section	NACE Division	Description	Industry
A	NACE 01-03	Agriculture, forestry and fishing	Primary and Extraction
B	NACE 05-09	Mining and quarrying	Primary and Extraction
C	NACE 10-33	Manufacturing	Manufacturing
D	NACE 35	Electricity, gas, steam and air conditioning supply	Utilities
E	NACE 36-39	Water supply; sewage, waste management and remediation activities	Utilities
F	NACE 41-43	Construction	Construction
G	NACE 45-47	Wholesale and retail trade; repair of motor vehicles and motorcycles	Market Services
H	NACE 49-53	Transportation and storage	Market Services
I	NACE 55-56	Accommodation and food service activities	Market Services
J	NACE 58-63	Information and communication	Market Services
K	NACE 64-66	Financial and insurance activities	Market Services
L	NACE 68	Real estate activities	Market Services
M	NACE 69-75	Professional, scientific and technical activities	Market Services
N	NACE 77-82	Administrative and support service activities	Market Services
O	NACE 84	Public administration and defence; compulsory social security	Non-Market Services
P	NACE 85	Education	Non-Market Services
Q	NACE 86-88	Human health and social work activities	Non-Market Services
R	NACE 90-93	Arts, entertainment and recreation	Non-Market Services
S	NACE 94-96	Other service activities	Non-Market Services
T	NACE 97-98	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	-
U	NACE 99	Activities of extraterritorial organizations and bodies	-

	$\ln S_i$	$\ln n_i^c$	$\ln n_i^s$
variance	1.73	1.87	0.60
skewness	0.57	0.13	-0.02
kurtosis	4.54	3.45	4.17
p1	-2.88	-3.18	-2.04
p5	-1.93	-2.22	-1.26
p10	-1.50	-1.71	-0.93
p25	-0.84	-0.89	-0.46
p50	-0.11	-0.01	-0.00
p75	0.72	0.87	0.46
p90	1.66	1.72	0.93
p95	2.34	2.25	1.26
p99	3.82	3.41	1.97

Table 7: Moments of distributions.

Notes: all variables are demeaned by their NACE 4-digit industries by regressing log variables on sector fixed effects and retaining the residuals.

## B Additional results: Stylized facts

### B.1 Distributions of firm size and number of connections

We provide additional moments on the distribution of firm sizes and the number of customers and suppliers in Table 7. Again, all variables are in logs, and demeaned by their NACE 4-digit industries.

We also report information on the raw (non-demeaned) distributions of these variables, confirming that massive heterogeneity in the production network is not driven by particular sectors, but instead is natural across all types of economic activity. Table 8 reports the distribution of firm size in the full sample, as well as separately for six aggregated industries for the year 2014: primary and extraction, manufacturing, utilities, construction, market services and non-market services. Tables 9 and 10 report summary statistics for the number of firm customers and the number of firm suppliers, respectively. While there is natural variation in average values across different types of economic activity, sizable dispersion is prevalent in all industries.

### B.2 Market share and number of customers: seller-buyer sector-pair demeaning

A potential concern in interpreting the stylized facts in Section 2.2, in particular the relationship between average market shares and number of customers, is that firms with many

Table 8: Total sales.

Industry	N	Mean	SD	Percentiles						
				p10	p25	p50	p75	p90	p95	p99
Primary and extraction (NACE 01-09)	2,818	12.9	451	.2	.4	.8	2.0	5.1	10.0	53.6
Manufacturing (NACE 10-33)	16,827	15.0	259	.3	.5	1.2	4.0	14.6	36.1	218.0
Utilities (NACE 35-39)	840	36.1	444	.4	.8	1.9	6.8	24.8	62.4	492.6
Construction (NACE 41-43)	19,008	2.4	14	.2	.3	.6	1.5	3.7	7.1	27.6
Market services (NACE 45-82)	53,532	6.3	86	.2	.4	1.0	2.5	7.4	15.6	73.8
Non-market services (NACE 84-96)	1,122	2.0	8	.2	.3	.5	1.1	2.8	6.8	32.0
All	94,147	7.5	155	.2	.4	.9	2.4	7.4	16.2	88.8

Note: values are reported in millions of euro.

Table 9: Number of customers.

Industry	N	Mean	SD	Percentiles						
				p10	p25	p50	p75	p90	p95	p99
Primary and extraction (NACE 01-09)	2,818	74	187	3	7	23	79	187	276	655
Manufacturing (NACE 10-33)	16,827	130	513	5	13	39	119	280	464	1,192
Utilities (NACE 35-39)	840	363	3,067	5	13	51	178	481	895	3,202
Construction (NACE 41-43)	19,008	50	246	4	7	16	37	94	174	533
Market services (NACE 45-82)	53,532	147	926	3	7	29	112	290	504	1,661
Non-market services (NACE 84-96)	1,122	66	433	2	4	10	40	105	187	591
All	94,147	123	797	3	8	26	92	245	429	1,299

Table 10: Number of suppliers.

Industry	N	Mean	SD	Percentiles						
				p10	p25	p50	p75	p90	p95	p99
Primary and extraction (NACE 01-09)	2,818	75	83	22	34	54	86	138	197	411
Manufacturing (NACE 10-33)	16,827	110	154	25	42	70	126	225	313	670
Utilities (NACE 35-39)	840	152	280	25	45	80	153	286	456	1,508
Construction (NACE 41-43)	19,008	76	97	23	34	52	85	140	204	466
Market services (NACE 45-82)	53,532	69	87	17	30	50	80	132	184	363
Non-market services (NACE 84-96)	1,122	63	71	19	29	47	71	116	156	335
All	94,147	79	108	20	32	53	88	152	218	464

customers might be selling different types of products compared to firms with few customers, even within the same 4-digit industry of the seller. In that case, average market shares might depend on seller-buyer pair characteristics, rather than characteristics of the seller only. For example, there may be both niche and broad-market sellers within the same industry, where the broad seller has more customers than the niche seller. Presumably, buyers of niche and broad-market products to some extent also belong to different industries: e.g. broad-market coffee roasters may sell to grocery stores, while niche coffee roasters may sell to coffee shops.

To take this into account, we reproduce the correlation between customer market shares and the number of customers in Fact 2, using fixed effects for the seller-buyer industry pair. By also demeaning by the customer sector, we account for heterogeneity in input requirements across sectors *within* the sector of the supplier. Specifically, for every seller  $i$ , we calculate the average market share and the number of customers in each 4-digit buyer industry  $k$ ,  $\bar{\delta}_{ik}$  and  $n_{ik}$ , respectively. The average market share is calculated as in the main text: the weighted geometric mean of  $m_{ij}/M_j^{net}$  across customers  $j$  belonging to buyer industry  $k$ . We then regress  $\bar{\delta}_{ik}$  on  $n_{ik}$  (in logs), including seller-buyer industry pair fixed effects. The OLS slope coefficient remains negative and significant at -.03. While part of the variation is indeed absorbed by sector-pair-specific components, this shows that, even when looking at sales within a very detailed cell of the input-output table, e.g. coffee roasters selling to coffee shops, sellers with many customers obtain systematically lower market shares of their buyers' input purchases.

### B.3 Fringe buyers

Another potential concern is that sellers with many connections have relatively more fringe buyers than firms with fewer connections, i.e. less important customers in terms of bilateral sales  $m_{ij}$ . This pattern would imply that firms with more customers have lower average sales per customer and lower market shares in these customers, but higher sales to each customer at each customer rank, rejecting the model in the main text.

We investigate this possibility by looking at the relationship between firm sales and the number of customers after dropping potential fringe buyers. We identify and drop fringe buyers in two ways. First, we drop transactions  $m_{ij}$  that are below the first quartile of a firm's transactions  $m_{ij}$  (i.e., the quartile is firm-specific).<sup>30</sup> Alternatively, we drop transactions that are below the global 1st quartile across all  $m_{ij}$  in the sample. In each case we recalculate the adjusted number of customers  $n_i^c$  and network sales  $S_i^{net}$ . Table 11 repeats the baseline analysis in this modified sample. The first column reproduces the baseline results for the

---

<sup>30</sup> This means that the analysis is restricted to firms with 5 or more customers.

Dep. var: $\ln S_i^{net}$	(1) Baseline	(2) No fringe buyers (firm-level)	(2) No fringe buyers (global)
$\ln n_i^c$	.77*** (.00)	.81*** (.01)	.91*** (.00)
Seller industry fixed effects	Yes	Yes	Yes
$N$	94,147	80,224	80,156

Table 11: Fringe buyers, 2014.

Note: Regressions comparing the elasticity of sales with respect to the number of customers across (1) baseline, (2) after dropping fringe buyers (bottom quartile) within each firm, and (3) after dropping fringe buyers (bottom quartile) across all firms. Significance: \* < 5%, \*\* < 1%, \*\*\* < 0.1%.

regression of  $\ln S_i^{net}$  on  $\ln n_i^c$ , while columns 2 and 3 show the results after removing fringe buyers. The slope coefficient increases slightly, but the main message of Fact 2 remains, namely that average sales (and market shares) are smaller for firms with many customers.

## C Exogenous mobility

In this section, we first discuss the empirical relevance of buyer and seller effects in equation (1). We then examine the necessary assumptions on the assignment process of buyers and sellers for OLS to identify the underlying parameters of interest, and develop a simple evaluation of conditional exogenous mobility in the context of a production network.

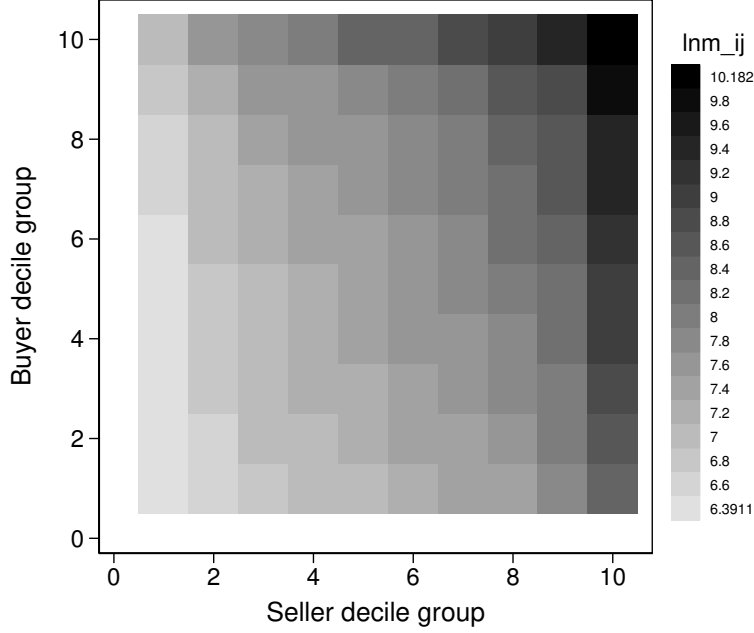
### C.1 Buyer and seller effects

The log-linear relationship in equation (1) predicts the following: (i) expected sales from seller  $i$  to customer  $j$  are increasing in the average sales of  $i$  to other customers  $k$ , and (ii) expected purchases by buyer  $j$  from seller  $i$  are increasing in the average purchases by  $j$  from other suppliers  $k$ .

Both properties can be tested non-parametrically as follows. For each seller  $i$  and buyer  $j$ , calculate the leave-out mean of log sales ( $\bar{s}_i^{-l}$ ) and purchases ( $\bar{m}_j^{-l}$ ) across its buyers and



Figure 9: Average log sales across seller and buyer decile groups



Note: The figure shows the average of  $\ln m_{ij}$  in all decile group pairs  $(q_{\bar{s}}, q_{\bar{m}})$ .

suppliers, excluding customer/supplier  $l$ , respectively:<sup>31</sup>

$$\bar{s}_i^{-l} = \frac{\sum_{j \in \mathcal{C}_i \setminus l} \ln m_{ij}}{n_i^c - 1}$$

$$\bar{m}_j^{-l} = \frac{\sum_{k \in \mathcal{S}_j \setminus l} \ln m_{kj}}{n_j^s - 1}.$$

Then sort firms into decile groups based on  $\bar{s}_i^{-l}$  and  $\bar{m}_j^{-l}$ , denoting the decile group the firm belongs to as  $q_{\bar{s}} = 1, \dots, 10$  and  $q_{\bar{m}} = 1, \dots, 10$ , respectively. Finally, calculate the mean of  $\ln m_{ij}$  for every decile group pair,  $\overline{\ln m_{q_{\bar{s}}, q_{\bar{m}}}}$ , e.g. the average  $\ln m_{ij}$  for the seller-buyer pairs in  $(q_{\bar{s}}, q_{\bar{m}}) = (1, 1)$ , and so on.

Figure 9 illustrates the results using a heatmap. The decile groups  $q_{\bar{s}}$  and  $q_{\bar{m}}$  are plotted on the horizontal and vertical axes, respectively.  $\ln m_{ij}$  is increasing in the average sales from  $i$  to other customers  $k$  (moving from left to right in the diagram), and  $\ln m_{ij}$  is increasing in the average purchases of  $j$  from other suppliers  $k$  (moving from bottom to top in the

<sup>31</sup> Using the overall mean generates a mechanical relationship between e.g. seller size and sales between  $i$  and  $j$ . We calculate  $\bar{s}_i^{-l}$  and  $\bar{m}_j^{-l}$  for all  $(i, l)$  and  $(j, l)$  pairs respectively. Firms with only one customer or supplier are by construction omitted from the sample.

diagram).

## C.2 Assumptions on the Assignment Process

Equation (1) is a two-way fixed effects model similar to the models that are used in the employer-employee literature (Abowd et al., 1999; Card et al., 2013).<sup>32</sup> OLS estimates of  $\ln \psi_i$  and  $\ln \theta_j$  will identify the effect of seller and buyer characteristics if the following moment conditions are satisfied:

$$\begin{cases} E[s_i' r] = 0 & \forall i \\ E[b_j' r] = 0 & \forall j. \end{cases} \quad (9)$$

Here  $S = [s_1, \dots, s_N]$  is the  $N^* \times N_s$  seller fixed effects design matrix,  $B = [b_1, \dots, b_N]$  is the  $N^* \times N_b$  buyer fixed effects design matrix,  $r$  is the  $N^* \times 1$  vector of residual match effects, and  $N^*$ ,  $N_s$  and  $N_b$  are the number of matches, sellers and buyers, respectively. The first condition states that for each seller  $i$ , the average  $\ln \omega_{ij}$  across buyers  $j$  is zero, while the second condition states that for each buyer  $j$ , the average  $\ln \omega_{ij}$  across sellers  $i$  is zero. Intuitively, a high  $\ln \omega_{ij}$  that is common across customers  $j$  of  $i$  will be automatically loaded onto  $i$ 's seller effect (and similarly for suppliers  $i$  of  $j$ ). In other words, these moment conditions require that the assignment of suppliers to customers is exogenous with respect to  $\omega_{ij}$ , so-called conditional exogenous mobility in the labor literature.

Exogenous mobility holds more generally than perhaps considered *prima facie*. It is instructive to review two important cases when these moment conditions hold. First, they hold if firms match based on their seller and buyer effects, e.g., highly productive firms match with more and/or different customers/suppliers than less productive ones. Second, the assumption holds if firms match based on idiosyncratic pair-wise shocks that are unrelated to  $\ln \omega_{ij}$ . One example of this is idiosyncratic fixed costs, such as costs related to search and matching, which affect profits for a potential match but not the value of bilateral sales.<sup>33</sup>

Now consider the case of endogenous mobility. To fix ideas, assume that matching is based on the idiosyncratic match component of sales,  $\omega_{ij}$ , together with the seller effect  $\psi_i$ . In that case, only high  $\psi_i$  sellers would want to match with low  $\omega_{ij}$  buyers. OLS would then give a downward bias in the estimated  $\psi_i$ , because OLS imposes that the average  $\ln \omega_{ij}$  across customers is zero.

---

<sup>32</sup> The linear fixed-effects approach imposes no restrictions on the seller and buyer effects, unlike random or mixed effects models. With random effects, one also needs to model the network formation game to assess the plausibility of the required distributional assumptions for unobserved heterogeneity (see Bonhomme (2020)).

<sup>33</sup> E.g., the matching framework presented in Section 4.

### C.3 Exogenous mobility evaluation

To explore the possibility that matching shocks are correlated with sales shocks, we evaluate conditional exogenous mobility as follows. Consider firm  $i$  selling to customers 1 and 2. The expected difference in bilateral sales is

$$\Delta \ln m_i \equiv E [\ln m_{i2} - \ln m_{i1} \mid (i, 1), (i, 2)] = \ln \theta_2 - \ln \theta_1 + E [\ln \omega_{i2} - \ln \omega_{i1} \mid (i, 1), (i, 2)].$$

Consider the case  $\theta_2 > \theta_1$ . Under exogenous mobility, the last expectation term is zero, and  $\Delta \ln m_i$  is unrelated to firm  $i$  characteristics. Under endogenous mobility, the last expectation term is non-zero, and  $\Delta \ln m_i$  is potentially a function of firm  $i$  characteristics. Now seller  $i$  will only want to match with customer 1 if  $\omega_{i1}$  is sufficiently large. The expectation  $E [\ln \omega_{i2} - \ln \omega_{i1} \mid (i, 1), (i, 2)]$  is then negative. Moreover, for small sellers (low  $\psi_i$ ), the size of  $\omega_{i1}$  is important for whether a match occurs or not, while for large sellers (high  $\psi_i$ ), the size of  $\omega_{i1}$  is less important (since matching is determined by both  $\psi_i$  and  $\omega_{ij}$ ). Under endogenous mobility, the expectation is therefore less negative for high- $\psi_i$  than low- $\psi_i$  firms, so that  $\Delta \ln m_i$  is greater for high- $\psi_i$  than low- $\psi_i$  firms. Under exogenous mobility, by contrast,  $\Delta \ln m_j$  should be unrelated to  $\psi_i$ .<sup>34</sup>

Going back to the seller and buyer decile groups constructed above, these predictions can be tested by looking at  $\overline{\ln m}_{q_{\bar{s}}, q_{\bar{m}}}$  when moving from a small to a big customer, for different groups of sellers. Figure 10 shows the results. Each line represents the mean of log sales for a given seller decile group (1,...,10). Within a seller group, we calculate  $\overline{\ln m}_{q_{\bar{s}}, q_{\bar{m}}}$  to small customers (buyer decile group 1) and to big customers (buyer decile group 10). Under exogenous mobility, those lines should be parallel, i.e. for buyer bins  $q_{\bar{m}}$  and  $q'_{\bar{m}}$ ,  $\overline{\ln m}_{q_{\bar{s}}, q_{\bar{m}}} - \overline{\ln m}_{q_{\bar{s}}, q'_{\bar{m}}}$  does not depend on the seller decile group. The lines are, to a large degree, parallel, in particular for the seller decile groups 2 to 9. Parallel lines are a sufficient but not necessary condition for exogenous mobility: If the data generating process is not linear in logs, then one could find non-parallel lines even under exogenous mobility.

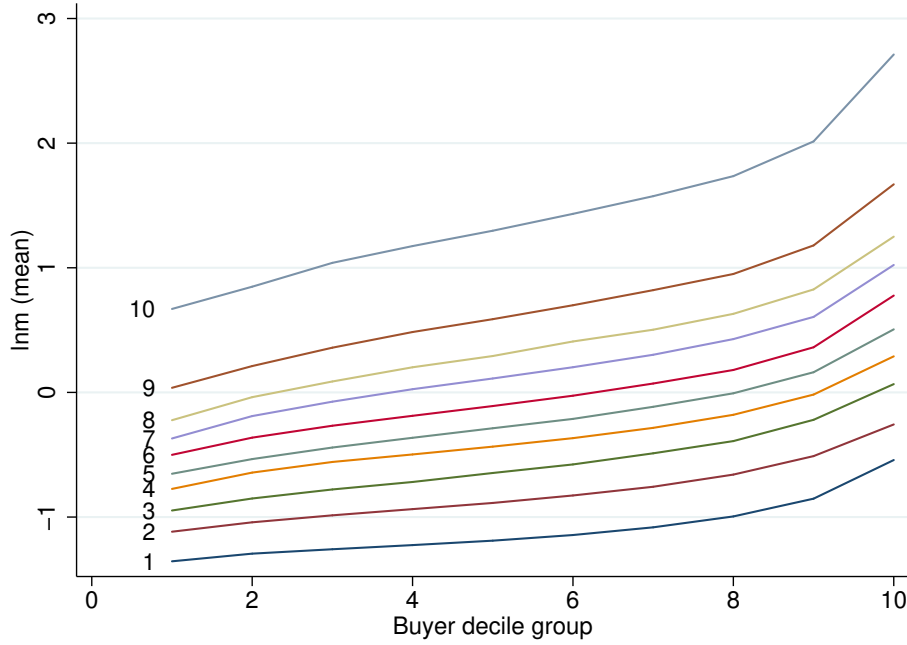
One can test for this non-parametrically as follows. Using the buyer and seller bins defined above, exogenous mobility implies that

$$\overline{\ln m}_{q'_{\bar{s}}, q'_{\bar{m}}} - \overline{\ln m}_{q_{\bar{s}}, q_{\bar{m}}} - (\overline{\ln m}_{q_{\bar{s}}, q'_{\bar{m}}} - \overline{\ln m}_{q_{\bar{s}}, q_{\bar{m}}}) = 0, \quad (10)$$

for any bins  $q_{\bar{s}}$ ,  $q'_{\bar{s}}$ ,  $q_{\bar{m}}$  and  $q'_{\bar{m}}$ . We form these averages for  $q'_{\bar{s}} = q_{\bar{s}} + 1$  and  $q'_{\bar{m}} = q_{\bar{m}} + 1$  and test the null hypothesis that the double difference equals zero. This yields 81 separate

<sup>34</sup> Card et al. (2013) test for endogenous mobility for employer-employee matches using a related, but different, test.

Figure 10: Average log sales across seller and buyer decile groups



Note: The figure shows  $\ln \bar{m}_{ij}$  across buyer decile groups  $q_{\bar{m}} = 1, \dots, 10$ . Each line represents a seller decile group,  $q_{\bar{s}} = 1, \dots, 10$ .

hypothesis tests across all buyer-seller pair bins.<sup>35</sup> Overall, the results mirror those in Figure 10: the double differences are not significantly different from zero in the middle of the distribution, whereas we find significant deviations in the tails. Significant deviations are typically relatively small: e.g. moving from a 6th to 7th decile buyer yields 12% more sales for seller decile 9 and 14% more sales for seller decile 10, a difference of 2%.

We report 81 separate hypothesis tests across all buyer-seller pair bins in Table 12. Each column refers to the change from buyer decile  $t$  to  $t + 1$ , and each row refers to the change from seller decile  $t$  to  $t + 1$ . For example, the cell (3-2,2-1) reports the difference  $\overline{\ln m_{3,2}} - \overline{\ln m_{3,1}} - (\overline{\ln m_{2,2}} - \overline{\ln m_{2,1}})$ .

<sup>35</sup> t-values are calculated using Welch's t-test.

Table 12: Exogenous mobility test

		Buyer decile								
		2-1	3-2	4-3	5-4	6-5	7-6	8-7	9-8	10-9
Seller decile	2-1	0.01*	0.02*	0.02*	0.03*	0.02*	0.01	0.01*	0.01	-0.04*
	3-2	0.04*	0.01*	0.01	0.00	0.00	0.01*	0.00	0.01	0.02*
	4-3	0.01	0.01*	0.02*	0.00	0.01	0.00	0.01	0.00	0.03*
	5-4	0.00	-0.01	0.00	0.01*	0.00	0.00	0.00	0.00	0.08*
	6-5	0.02*	0.02*	0.01	0.00	0.01	0.01	0.02*	0.00	0.02*
	7-6	0.04*	0.02*	0.02*	0.01*	0.01*	0.01	0.01	0.03*	0.00
	8-7	0.00	0.01*	0.00	0.00	0.03*	-0.01	0.00	-0.01	0.01
	9-8	-0.01	0.02*	0.02*	0.01	-0.01	0.02*	0.01	0.05*	0.08*
	10-9	0.03*	0.04*	0.01	0.02*	0.02*	0.02*	0.02*	0.04*	0.19*

Note: The table shows the double difference from equation (10) in the main text. Significance: \* < 5%, \*\* < 1%, \*\*\* < 0.1%. t values are based on Welsh's t-test.

## D Theory Appendix

### D.1 Variance decomposition

#### D.1.1 Identities

We decompose bilateral sales  $m_{ij}$  into  $i$ ,  $j$  and  $ij$  components:

$$m_{ij} \stackrel{def}{=} G\psi_i\theta_j\omega_{ij} \quad (11)$$

where  $G$  is a mean across all  $ij$ , and without loss of generality,  $\psi_i, \theta_j$  and  $\omega_{ij}$  are defined relative to that mean. Summing over all customers  $j$  of seller  $i$ ,  $j \in \mathcal{C}_i$ , on both sides:

$$\begin{aligned} S_i^{net} &\stackrel{def}{=} \sum_{j \in \mathcal{C}_i} m_{ij} = \sum_{j \in \mathcal{C}_i} G\psi_i\theta_j\omega_{ij} \\ &= G\psi_i \sum_{j \in \mathcal{C}_i} \theta_j\omega_{ij} \end{aligned} \quad (12)$$

where  $S_i^{net}$  is the total network sales of  $i$  to all its customers  $j$ . Total sales of  $i$  can be expressed as an identity:

$$\begin{aligned} S_i &= S_i^{net} \frac{S_i}{S_i^{net}} \\ \Rightarrow \ln S_i &= \ln S_i^{net} + \ln \beta_i \end{aligned}$$

where  $\beta_i \stackrel{def}{=} \frac{S_i}{S_i^{net}} \geq 1$ . Next, we can further decompose the  $j$ -specific term in eq.(12) into components that are observable in the data:

$$\sum_{j \in \mathcal{C}_i} \theta_j \omega_{ij} = n_i^c \bar{\theta}_i \bar{\omega}_i \frac{1}{n_i^c} \sum_j \frac{\theta_j \omega_{ij}}{\bar{\theta}_i \bar{\omega}_i}$$

where  $n_i^c$  is the number of business customers  $j$  of firm  $i$ , and we choose  $\bar{\theta}_i \stackrel{def}{=} \left( \prod_j \theta_j \right)^{\frac{1}{n_i^c}}$  and  $\bar{\omega}_i \stackrel{def}{=} \left( \prod_j \omega_{ij} \right)^{\frac{1}{n_i^c}}$  to be geometric means of  $\omega_{ij}$  and  $\theta_j$  respectively for each  $i$ . Taking logs on both sides:

$$\ln \left( \sum_{j \in \mathcal{C}_i} \theta_j \omega_{ij} \right) = \ln n_i^c + \ln \bar{\theta}_i + \ln \bar{\omega}_i + \ln \left( \frac{1}{n_i^c} \sum_{j \in \mathcal{C}_i} \frac{\theta_j \omega_{ij}}{\bar{\theta}_i \bar{\omega}_i} \right) \quad (13)$$

$$= \ln n_i^c + \frac{1}{n_i^c} \sum_{j \in \mathcal{C}_i} \ln \theta_j + \frac{1}{n_i^c} \sum_{j \in \mathcal{C}_i} \ln \omega_{ij} + \underbrace{\ln \left( \frac{1}{n_i^c} \sum_{j \in \mathcal{C}_i} \frac{\theta_j \omega_{ij}}{\bar{\theta}_i \bar{\omega}_i} \right)}_{\stackrel{def}{=} \ln \Omega_i^c} \quad (14)$$

Combining all the above leads to the decomposition of sales for every  $i$ :

$$\begin{aligned} \ln S_i &= \ln S_i^{net} + \ln \beta_i \\ &= \ln G + \ln \psi_i + \ln n_i^c + \ln \bar{\theta}_i + \ln \bar{\omega}_i + \ln \Omega_i^c + \ln \beta_i \end{aligned} \quad (15)$$

The residual term  $\ln \bar{\omega}_i$  is obtained from the 2-way FE OLS regression, and by the properties of OLS,  $\ln \bar{\omega}_i = 0$ .

### D.1.2 Properties of the Decomposition

Consider first a two component case:  $\ln S^{net} = \ln \psi + \ln \xi$ . Then  $Var(\ln S^{net}) = Var(\ln \psi) + Var(\ln \xi) + 2Cov(\ln \psi, \ln \xi)$ . When we run an “inverse” regression of components on sales ( $\ln \psi = \alpha + \beta_1 \ln S^{net} + \varepsilon$ ), OLS estimation implies  $\beta_1 = \frac{Cov(\ln S^{net}, \ln \psi)}{Var(\ln S^{net})}$ .

Note that

$$\begin{aligned}
Cov(\ln S^{net}, \ln \psi) &= Cov(\ln \psi + \ln \xi, \ln \psi) \\
&= Cov(\ln \psi, \ln \psi) + Cov(\ln \xi, \ln \psi) \\
&= Var(\ln \psi) + Cov(\ln \xi, \ln \psi)
\end{aligned} \tag{16}$$

where the penultimate equation follows from applying a linear combination of two random variables. Hence  $\beta_1 = \frac{Cov(\ln S^{net}, \ln \psi)}{Var(\ln S^{net})} = \frac{Var(\ln \psi)}{Var(\ln S^{net})} + \frac{Cov(\ln \xi, \ln \psi)}{Var(\ln S^{net})}$ . Similarly for  $\beta_2 = \frac{Cov(\ln S^{net}, \ln \xi)}{Var(\ln S^{net})} = \frac{Var(\ln \xi)}{Var(\ln S^{net})} + \frac{Cov(\ln \xi, \ln \psi)}{Var(\ln S^{net})}$ . Combining both:

$$\beta_1 + \beta_2 = \frac{1}{Var(\ln S^{net})} (Var(\ln \psi) + Var(\ln \xi) + 2Cov(\ln \xi, \ln \psi)) = 1 \tag{17}$$

From eq.(16), we see that these covariances are attributed equally to each component. I.e. the OLS coefficient thus captures the share of total variance in  $\ln S^{net}$  explained by the variance in  $\ln \psi$  plus any covariances across components.

The general case for the sum of  $K \geq 2$  general random variables  $X = \sum_{k=1}^K X_k$  is a straightforward extension, so that

$$\sum_{k=1}^K \beta_k = \frac{1}{Var(X)} \left( \sum_k Var(X_k) + \sum_{i \neq k} Cov(X_i, X_k) \right) = 1$$

Since we perform an exact decomposition (from the identities above), the sum of the variance shares sum to one, explaining all of the variance in  $X$ . Each OLS coefficient for  $X_k$  is equivalent to the share of the variance explained in  $X$ .

## D.2 Seller Fixed Effect And Average Market Share

This section shows the relationship between seller fixed effects and average market shares.

Sales from  $i$  to  $j$  is  $m_{ij} = e^G \psi_i \theta_j \omega_{ij}$ . Therefore network purchases can be written as  $M_j^{net} = e^G \theta_j \sum_{i \in \mathcal{S}_j} \psi_i \omega_{ij}$ . Rearranging, we get

$$\theta_j = \frac{M_j^{net}}{e^G \sum_{i \in \mathcal{S}_j} \psi_i \omega_{ij}}. \tag{18}$$

Seller  $i$ 's market share in  $j$ 's network purchases is

$$\frac{m_{ij}}{M_j^{net}} = \frac{e^G \psi_i \theta_j \omega_{ij}}{M_j^{net}}.$$

The average of log market shares (i.e., the log of the geometric mean) across  $i$ 's customers is

$$\begin{aligned} \ln MktSh_i &= \frac{1}{n_i^c} \sum_{j \in \mathcal{C}_i} \ln \frac{m_{ij}}{M_j^{net}} \\ &= \frac{1}{n_i^c} \sum_{j \in \mathcal{C}_i} \ln \frac{e^G \psi_i \theta_j \omega_{ij}}{M_j^{net}} \\ &= G + \ln \psi_i + \frac{1}{n_i^c} \left( \sum_{j \in \mathcal{C}_i} \ln \frac{\theta_j \omega_{ij}}{M_j^{net}} \right). \end{aligned}$$

Substituting  $\theta_j$  with equation (18) above, we obtain

$$\begin{aligned} \ln MktSh_i &= \ln \psi_i + \frac{1}{n_i^c} \left( \sum_{j \in \mathcal{C}_i} \ln \frac{\omega_{ij}}{\sum_{i' \in \mathcal{S}_j} \psi_{i'} \omega_{i'j}} \right) \\ &= \ln \psi_i - \frac{1}{n_i^c} \sum_{j \in \mathcal{C}_i} \ln \tau_j \end{aligned}$$

where  $\tau_j \equiv \sum_{i' \in \mathcal{S}_j} \psi_{i'} \omega_{i'j}$  and we used the fact that  $\sum_{j \in \mathcal{C}_i} \ln \omega_{ij} = 0$  by the properties of OLS.

### D.3 Existence and Uniqueness

We prove existence and uniqueness by showing that the fixed network equilibrium belongs to the class of models analyzed by Allen et al. (2016). They consider the following system of equations:

$$\prod_{h=1}^K (x_i^h)^{\gamma_{kh}} = c_i^k + \sum_{j=1}^N K_{ij}^k \prod_{h=1}^K (x_j^h)^{\beta_{kh}},$$

where  $i, j \in \{1, \dots, N\}$  are firms/sectors,  $x_i^h$  is the type  $h$  equilibrium variable,  $c_i^k$  is a constant and  $K_{ij}^k$  are exogenous linkages between  $i$  and  $j$ . With  $K = 1$  this reduces to

$$x_i^\gamma = c_i + \sum_{j=1}^N K_{ij} x_j^\beta. \quad (19)$$



The backward fixed point in equation (6) can be written in the form of equation (19) with  $\gamma = 1$ ,  $c_i = 0$ ,  $\beta = 1 - \alpha$  and  $K_{ij} = z(\lambda')^{\sigma-1}$ . Using their notation,  $\mathbf{A}$  is simply  $1 - \alpha$  and therefore the maximum eigenvalue of  $\mathbf{A}^p$  is also  $1 - \alpha < 1$ . According to their Theorem 2(i), there exists a unique and strictly positive solution to the backward fixed point.

The forward fixed point in equation (7) can be written in the form of equation (19) with  $\gamma = 1$ ,  $c_i = \mu^{1-\sigma} z(\lambda)^{\sigma-1} P(\lambda)^{(1-\sigma)(1-\alpha)} X/\mathcal{P}^{1-\sigma}$ ,  $\beta = 1$  and  $K_{ij} = (1 - \alpha) / (\mu P(\lambda')^{1-\sigma})$ . Using their notation,  $\mathbf{A}$  is 1 and therefore the maximum eigenvalue of  $\mathbf{A}^p$  is also 1. According to their Theorem 2(ii.a) there exists at most one strictly positive solution to the forward fixed point.

## E Additional results: Decomposition

### E.1 Results by sector

The main results pool the within-sector results across all sectors. To evaluate heterogeneity across sectors, we also perform the firm size decomposition separately by NACE 2-digit sector. To deal with possible incidental parameters, we drop sectors with fewer than 5 observations. Results are in Table 13. Table 14 shows the average value of the components across sectors, the standard deviation, and the coefficient of variation. All coefficients of variation are smaller than one, indicating relatively small differences in the firm size decomposition across sectors.

### E.2 Results by year

Table 15 performs the firm size decomposition separately for each year in the sample. The results are virtually identical across all years.

### E.3 Long differences

We also perform the decomposition in changes, or long differences from 2002 to 2014. First, we estimate equation (1) on two cross-sections, the baseline year 2014 ( $t = 1$ ) and year 2002 ( $t = 0$ ). We then calculate the change in every demeaned variable in the decomposition:

$$\Delta \ln S_i = \Delta \ln G + \Delta \ln \psi_i + \Delta \ln n_i^c + \Delta \ln \bar{\theta}_i + \Delta \ln \Omega_i^c + \Delta \ln \beta_i$$

where  $\Delta \ln S_i$  denotes the log difference in  $S_i$  from  $t = 0$  to  $t = 1$  etc. We then demean all variables at the NACE 4-digit level. Finally, we regress each log-differenced component on  $\Delta \ln S_i$ . Since this is a within-firm analysis, we can only estimate the variance components in changes on firms that are active in both 2002 and 2014. Results are in Table 16. The main finding, that the number of customers dominates the decomposition, remains.

NACE Sector	$N$	$\ln \psi_i$	$\ln n_i^c$	$\ln \bar{\theta}_i$	$\ln \Omega_i^c$	$\ln \beta_i$	NACE Sector	$N$	$\ln \psi_i$	$\ln n_i^c$	$\ln \bar{\theta}_i$	$\ln \Omega_i^c$	$\ln \beta_i$
1	2,292	0.26	0.49	0.02	0.29	-0.05	49	2,939	0.07	0.53	0.02	0.35	0.02
2	298	0.16	0.51	0.04	0.31	-0.02	50	101	0.09	0.28	0.07	0.23	0.32
3	48	0.71	-0.04	0.05	0.36	-0.08	51	32	0.29	0.25	0.04	0.30	0.11
7	5	-0.11	0.47	-0.14	0.51	0.27	52	884	0.07	0.47	0.04	0.30	0.12
8	174	0.16	0.42	0.06	0.32	0.04	53	102	-0.12	0.88	0.02	0.13	0.09
9	20	0.27	0.22	0.03	0.28	0.20	55	836	0.11	0.75	0.07	0.21	-0.15
10	2,720	0.41	0.49	0.07	0.39	-0.36	56	4,304	0.11	0.57	0.05	0.19	0.08
11	150	0.36	0.30	0.04	0.36	-0.05	58	430	0.19	0.48	0.04	0.27	0.01
12	18	0.07	0.20	-0.06	0.24	0.55	59	356	0.22	0.35	0.05	0.34	0.04
13	661	0.22	0.17	0.05	0.26	0.30	60	21	0.30	0.27	0.05	0.37	0.01
14	267	0.26	0.41	0.03	0.20	0.09	61	149	0.11	0.37	0.07	0.24	0.21
15	69	0.20	0.50	-0.02	0.21	0.11	62	1,962	0.24	0.41	0.10	0.18	0.08
16	1,081	0.12	0.73	0.04	0.23	-0.12	63	182	0.25	0.39	0.10	0.22	0.04
17	252	0.34	0.17	0.06	0.30	0.13	64	1,137	0.16	0.39	0.06	0.23	0.16
18	1,131	0.22	0.43	0.06	0.25	0.04	65	11	0.36	0.11	0.14	0.26	0.12
19	22	0.09	0.56	-0.03	0.33	0.05	66	336	0.19	0.37	0.11	0.22	0.12
20	590	0.29	0.17	0.10	0.29	0.15	68	1,305	0.20	0.43	0.09	0.19	0.10
21	156	0.27	0.37	0.05	0.33	-0.03	69	4,065	0.02	0.67	0.08	0.09	0.14
22	620	0.24	0.31	0.04	0.26	0.15	70	1,726	0.28	0.31	0.10	0.21	0.09
23	1,023	0.27	0.49	0.06	0.27	-0.08	71	1,382	0.22	0.39	0.12	0.25	0.03
24	359	0.25	0.17	0.06	0.29	0.22	72	58	0.13	0.01	0.10	0.24	0.52
25	3,294	0.17	0.44	0.06	0.26	0.07	73	650	0.29	0.31	0.09	0.31	0.01
26	462	0.17	0.30	0.07	0.25	0.21	74	177	0.11	0.43	0.08	0.33	0.05
27	458	0.20	0.43	0.08	0.26	0.03	75	101	0.19	0.96	0.01	0.15	-0.32
28	940	0.15	0.38	0.07	0.22	0.18	77	302	0.23	0.53	0.05	0.20	-0.02
29	249	0.30	0.07	0.09	0.37	0.16	78	255	0.14	0.75	-0.00	0.14	-0.02
30	81	0.25	0.17	0.06	0.33	0.19	79	236	0.17	0.52	0.00	0.25	0.06
31	899	0.13	0.66	0.04	0.20	-0.02	80	98	0.02	0.64	0.04	0.32	-0.02
32	648	0.43	0.39	0.05	0.20	-0.07	81	839	0.11	0.64	0.09	0.25	-0.09
33	755	0.14	0.54	0.06	0.29	-0.03	82	288	0.21	0.38	0.12	0.26	0.03
35	123	0.11	0.38	0.05	0.43	0.03	84	16	0.29	0.36	0.11	0.23	0.00
36	53	-0.03	0.50	0.03	0.32	0.18	85	139	0.07	0.94	0.06	0.15	-0.22
37	119	0.16	0.44	0.04	0.50	-0.15	86	39	0.10	0.08	0.13	0.15	0.54
38	519	0.20	0.35	0.05	0.32	0.07	87	13	0.13	0.31	0.05	0.21	0.31
39	38	0.18	0.50	0.05	0.28	-0.02	90	84	0.07	0.55	0.05	0.22	0.10
41	5,596	0.15	0.50	0.07	0.31	-0.04	91	5	-0.04	1.15	0.09	-0.07	-0.13
42	1,253	0.16	0.51	0.10	0.34	-0.11	92	25	-0.35	0.22	0.07	0.16	0.90
43	12,159	0.08	0.76	0.06	0.25	-0.15	93	265	0.18	0.62	0.03	0.26	-0.08
45	5,582	0.18	0.56	0.02	0.29	-0.05	94	5	0.42	-0.51	0.26	0.19	0.65
46	12,724	0.29	0.52	0.03	0.23	-0.06	95	29	-0.20	0.76	0.03	0.57	-0.16
47	10,012	0.04	0.48	0.00	0.18	0.29	96	526	0.16	0.65	0.07	0.26	-0.13

Table 13: Firm size decomposition by NACE 2-digit industry (2014)

NACE Sector	$\ln \beta_i$	$\ln \psi_i$	$\ln n_i^c$	$\ln \bar{\theta}_i$	$\ln \Omega_i^c$
mean	0.07	0.17	0.43	0.06	0.26
st. dev.	0.20	0.14	0.24	0.05	0.09
CV	2.70	0.81	0.55	0.82	0.34

Table 14: Variation in the firm size decomposition across NACE 2-digit industries (2014)

Year	$N$	$\ln \beta_i$	$\ln \psi_i$	$\ln n_i^c$	$\ln \bar{\theta}_i$	$\ln \Omega_i^c$
2002	81,254	0.05	0.17	0.49	0.04	0.25
2003	83,678	0.05	0.17	0.49	0.04	0.25
2004	85,030	0.04	0.18	0.49	0.04	0.25
2005	86,474	0.04	0.17	0.49	0.04	0.25
2006	88,581	0.04	0.17	0.50	0.04	0.25
2007	91,027	0.03	0.18	0.50	0.04	0.25
2008	92,280	0.03	0.18	0.50	0.04	0.25
2009	92,333	0.04	0.17	0.50	0.04	0.25
2010	92,713	0.03	0.17	0.50	0.04	0.25
2011	94,093	0.03	0.18	0.50	0.05	0.25
2012	95,375	0.03	0.18	0.50	0.05	0.25
2013	94,135	0.02	0.18	0.50	0.05	0.25
2014	94,147	0.01	0.18	0.51	0.05	0.25

Table 15: Firm size decomposition by year (2002-2014)

$N$	$\ln \beta_i$	$\ln \psi_i$	$\ln n_i^c$	$\ln \bar{\theta}_i$	$\ln \Omega_i^c$
40,974	0.10***	0.09***	0.57***	0.01***	0.23***

Note: Significance: \* < 5%, \*\* < 1%, \*\*\* < 0.1%.

Table 16: Firm size decomposition, long differences(2002-2014)

	$N$	$\ln \beta_i$	$\ln \psi_i$	$\ln n_i^c$	$\ln \bar{\theta}_i$	$\ln \Omega_i^c$
$NACE + NUTS3$	94,147	0.02***	0.17***	0.51***	0.05***	0.25***
$NACE \times NUTS3$	80,328	0.01***	0.16***	0.54***	0.05***	0.25***

Note: Significance: \* < 5%, \*\* < 1%, \*\*\* < 0.1%.

Table 17: Firm size decomposition by NACE 4-digit and NUTS3 (2014)

## E.4 Controlling for seller location

A potential concern is that firms sort into different locations based on size and productivity: Large firms might be located in large cities, where there are more potential customers but also more suppliers, simultaneously creating competition for market shares in each customer. Large firms might then have both more customers and lower market shares in each customer for reasons unrelated to our baseline model.

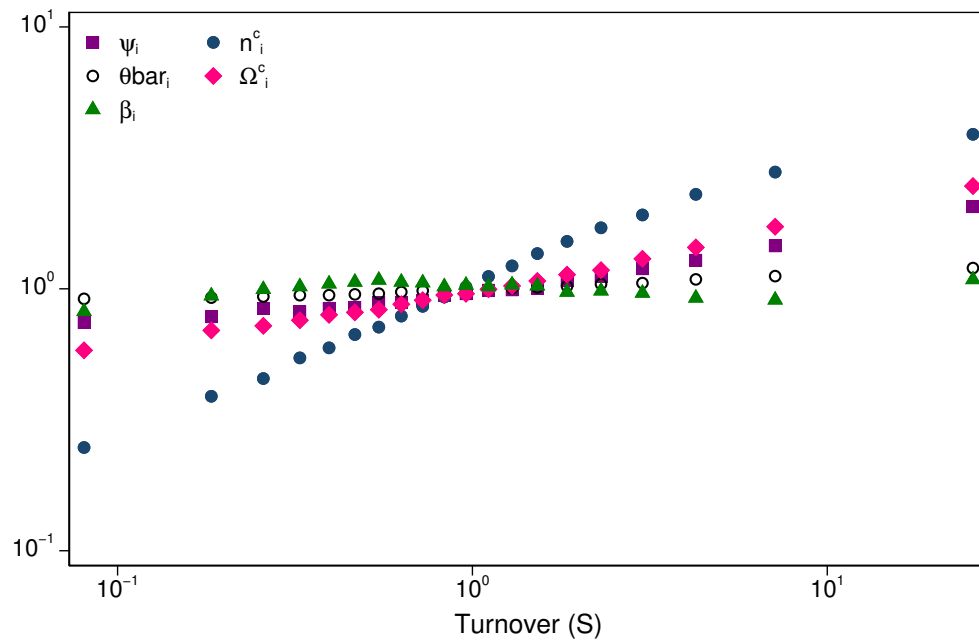
To assess this possibility, we perform the firm size decomposition controlling for the location of the seller.<sup>36</sup> In particular, we demean all variables by both the 4-digit NACE sector and the NUTS3 location of the seller. There are 44 NUTS3 regions in Belgium called *arrondissements*. We present two versions of the decomposition in Table 17. The first version demeans all variables by seller industry (NACE 4-digit) and seller location (NUTS3) fixed effects. The second version demeans all variables by seller industry-location pair fixed effects. As before, to account for potential incidental parameters, we drop cells with fewer than 5 observations. The results are almost identical to the baseline. This strongly suggests that the contribution of the different variance components is not significantly driven by agglomeration effects.

## E.5 Non-parametric results

We test whether the decomposition results are stable across the firm size distribution using a simple non-parametric approach. Figure 11 groups log sales into 20 equal-sized bins, computes the mean of log sales and the components  $\ln \psi_i$ ,  $\ln n_i^c$ ,  $\ln \bar{\theta}_i$ ,  $\ln \Omega_i^c$  and  $\ln \beta_i$  within each bin, and then creates a binned scatterplot of these data points. The result is a non-parametric visualization of the conditional expectation function, where the sum of the components on the vertical axis equals log sales on the horizontal axis. Overall, the slopes of each margin are close to linear, suggesting that the OLS decomposition results in the main text are applicable for both small and large firms.

<sup>36</sup> We thank an anonymous referee for this suggestion.

Figure 11: Firm size decomposition: Binned scatterplot.



Note: This binned scatterplot groups firms into 20 equal-sized bins by log sales, computes the mean of log sales and the components  $\ln \psi_i$ ,  $\ln n_i^c$ ,  $\ln \bar{\theta}_i$ ,  $\ln \Omega_i^c$  and  $\ln \beta_i$  within each bin, and graphs these data points. The result is a non-parametric visualization of the conditional expectation function.

Table 18: Business Group Size Decomposition ( $\ln S_i$ ).

Relative final demand	$\ln \beta_i$	.01	(.00)
Upstream	$\ln \psi_i$	.16	(.00)
# Customers	$\ln n_i^c$	.54	(.00)
Avg Customer Capability	$\ln \bar{\theta}_i$	.04	(.00)
Customer Interaction	$\ln \Omega_i^c$	.24	(.00)

Note: The table reports coefficient estimates from OLS regressions of a firm size margin (as indicated in the row heading) on total firm sales. All variables are first demeaned by their 4-digit NACE industry average. Standard errors in parentheses.

## E.6 Business Groups

The VAT identity number is the unique firm identifier we use across all datasets in this paper. This ID refers to the legal entity of the firm, and is the standard interpretation of a company in micro-level datasets. It is also the level at which a company reports annual accounts, firm-to-firm network data, and other typical firm characteristics. Still, there might be concerns that the legal entity does not constitute the economic entity of the company. In particular, multiple VAT identities may be owned by the same company. Intra-firm trade between these commonly owned entities might then exhibit non-market behavior in terms of the existence of particular buyer-supplier relationships and intra-firm pricing, affecting not only the value of sales relationships  $m_{ij}$  and the number of customers  $n_i^c$ , but potentially all components in the firm size decomposition.

We therefore perform the firm size decomposition on business groups instead of VAT identities, as in Tintelnot et al. (2021). In particular, VAT identities are grouped into a single firm if the same parent company owns at least 50% of their shares. Turnover, inputs, employment and labor costs are summed across subsidiaries to the group level, after subtracting within-group transactions from turnover and inputs to avoid double counting. The NACE code of the firm with the largest turnover is assigned to the group. We report results for the decomposition for 2014 in Table 18. The number of observations is now 86,245. The results are almost identical to the baseline.

## F Bootstrapped Standard Errors

The standard errors from the structural estimation in Section 5 are bootstrapped. The bootstrapping procedure is performed as follows. First, for each bootstrap sample, we randomly draw individual observations with replacement until we obtain the same sample size as the

main dataset. Then, for each bootstrap sample, we create the empirical moments used in the simulated method of moments estimation. We create 200 bootstrap moments in total. Finally, we estimate the parameters of the model 200 times using the 200 different bootstrap moments. The standard deviations of the estimated parameters are the standard errors reported in Table 5.