The Origins of Firm Heterogeneity: A Production Network Approach*

Andrew B. Bernard† Emmanuel Dhyne‡ Glenn C.G. Magerman§
Kalina Manova¶ Andreas Moxnes‖

PRELIMINARY AND INCOMPETE - DO NOT CITE OR CIRCULATE

Abstract

This paper quantifies the origins of firm size heterogeneity when firms are interconnected in a production network. We document new stylized facts about the universe of buyer-supplier relationships among all firms in Belgium during 2002-2014. These facts motivate a model in which firms buy inputs from upstream suppliers and sell to downstream buyers and final demand. Firms can be large not only because they have high production capability (i.e. productivity or product quality), but also because they interact with more, better and larger buyers and suppliers, and because they are better matched to their buyers and suppliers. The model delivers an exact decomposition of firm size into supply and demand margins with firm, buyer/supplier and match components. We establish three empirical results. First, demand factors explain 81% of firm size heterogeneity, while supply factors only 19%. Second, nearly all the variation on the demand side (99%) is driven by sales to other firms rather than final demand. By contrast, 88% of the variation on the supply side reflects heterogeneity in own production capability and 12% comes from input suppliers. Third, the extensive margin of the production network dominates the intensive margin. Most of the variance in the network demand and network supply components of firm size is determined by the number of customers and suppliers, rather than their size or capability. Second most important is that firms transact more with partners that are simultaneously more capable and better bilateral matches.

JEL: F10, F12, F16

Keywords: Production networks, productivity, firm size heterogeneity.

*This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 715147).

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

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

§ECARES - Université Libre de Bruxelles & National Bank of Belgium; glenn.magerman@ulb.ac.be

¶University of Oxford, CEPR & CEP; kalina.manova@economics.ox.ac.uk

‖University of Oslo & CEPR; andreamo@econ.uio.no

1 Introduction

Why are firms large or small? Even within narrowly defined industries, there is evidence of massive dispersion in firm outcomes such as revenue, employment, labor productivity or measured total factor productivity (see Syverson (2011) for a recent overview). In Belgium, a firm at the 90th percentile of the size distribution has turnover more than 35 times greater than a firm at the 10th percentile.[1] Understanding the origins of firm heterogeneity has important micro- and macro-economic implications. At the micro level, bigger firms perform systematically better along many dimensions, such as survival rate, innovation activity, 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., Melitz and Redding (2015), Pavcnik (2002), Gabaix (2011), di Giovanni et al. (2014), Gaubert and Itskhoki (2016)).

While the literature 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 firm-to-firm linkages in production networks. In particular, the focus has been on one-sided heterogeneity in either firm productivity on the supply side (e.g., Jovanovic (1982), Hopenhayn (1992), Melitz (2003), Luttmer (2007)) or final-consumer preferences on the demand side (e.g., Foster et al. (2015), Fitzgerald 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.

This paper examines how buyer-supplier connections in a complete production network shape the firm size distribution in the cross-section and its evolution over time.[2] The basic premise of the analysis is intuitive: firms can become large because they have inherently attractive capabilities such as productivity or product quality, because they interact with better and larger buyers and suppliers, and/or because they are particularly well matched to their buyers and suppliers. On the supply side, firms can improve their product quality or reduce their marginal costs if they enhance their own capabilities or if they buy more inputs from high-quality, efficient suppliers. On the demand side, firms can expand sales if they appeal to more final consumers or if they match with more and with bigger downstream producers. There may be higher-order effects in a production network as well, because the

[1] Estimates based on authors' calculations for the average NACE 4-digit industry in Belgium in 2012.

[2] Throughout the paper, firm size, sales, revenues and turnover are used interchangeably.

customers of the customers (and so on) of any one firm may ultimately also matter for that firm's economic performance.

The paper makes four main contributions. First, we document new stylized facts about a complete production network using 2002-2014 panel data on the universe of firm-to-firm domestic transactions in Belgium. Second, we provide a theoretical framework with minimal assumptions on production and demand that relates firm size to firm-specific characteristics, buyer and supplier characteristics, and buyer-supplier match characteristics. This allows the development of a new methodology for structurally estimating the primitives of the model from production network data. Third, we implement this methodology to decompose the Belgian firm size distribution into supply and demand components and to quantify the role of different firm-, buyer- and supplier characteristics. Finally, we simulate the model based on those estimated primitives and assess the welfare impact of policy-relevant shocks to the production network.

We first document five stylized facts about the incidence, magnitude and two-sided heterogeneity of firm-to-firm transactions in a complete domestic production network, using comprehensive value-added tax (VAT) records for Belgium during 2002-2014. First, the distributions of firms' total sales, number of buyer- and supplier connections, and value of buyer-supplier bilateral sales exhibit high dispersion and skewness. Second, bigger firms have more upstream suppliers and downstream buyers. Third, more connected firms match with less connected buyers and suppliers on average. Fourth, the distribution of a firm's sales across its buyers does not vary with its number of buyers, while the distribution of its purchases across its suppliers widens with the number of its suppliers. Finally, seller fixed effects and buyer fixed effects jointly account for only 38% of the variation in firm-to-firm sales. Together, these patterns suggest that the network of buyer-supplier links is key to understanding the firm size distribution. Specifically, they signal the important roles played by downstream input demand as distinct from final demand, by both seller- and buyer-specific firm characteristics, and by seller-buyer match quality.

Motivated by the stylized facts, we develop a theoretical framework that features two-sided firm heterogeneity in an input-output production network. This allows us to decompose firm sales into economically meaningful demand- and supply-side fundamentals. In the model, firms use a constant elasticity of substitution production technology that combines labor, imported inputs and inputs from upstream domestic suppliers. Firms sell their output to final consumers at home and abroad, as well as to downstream domestic producers. Since we want to examine how the network contributes to size dispersion, we take the observed production network as given and do not model the firm-to-firm matching decision. Note, however, that key firm metrics such as marginal costs, employment, prices, and sales are

nevertheless endogenous outcomes because they depend on the outcomes of all other firms in the economy.

In the framework, firms differ in production capability (a combination of efficiency and quality), as well as in sourcing capability (an input price aggregate that reflects the number and production capabilities of input suppliers). The value of a given firm-to-firm transaction depends on the production capability of the seller, the sourcing capability of the buyer, and the match quality of the specific seller-buyer pair. A new connection 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 implies greater opportunities to source cheaper or higher-quality inputs.

At the firm level, total firm sales can thus be decomposed into two overall margins: supply and demand. The supply margin can be further decomposed into own production capability and network supply (i.e. input costs), where the latter comprises the number of upstream suppliers, average production capability across suppliers, and the covariance of production capability and match quality across suppliers. Likewise, the demand margin can be further decomposed into final demand and network demand, where the latter comprises the number of downstream buyers, average sourcing capability across buyers, and the covariance of sourcing capability and match quality across buyers.[3]

We develop a three-step methodology to perform the exact model-based decomposition of firm size using the uniquely rich Belgian data. In the first step, we regress the value of bilateral firm-to-firm transactions on seller and buyer fixed effects, where the residual represents the bilateral match-specific component of the transactions. In the second step, we back out primitives of the model from these three terms. Intuitively, the seller and buyer fixed effects are related to firms' production and sourcing capability respectively, while the residual isolates firm-pair match quality. We also use balance sheet data to incorporate information on firms' activity outside the domestic production network (i.e. labor and imported inputs on the production side; exports and sales to final domestic consumers on the sales side). In the last step, we construct all components of the firm size decomposition, and regress each one on total firm sales. The coefficient estimates from this regression isolate the contribution of each supply- and demand-side margin to the overall variation in firm size.[4]

This methodology has several appealing features. It provides an agnostic decomposition of firm size as it imposes no restrictions on the relative magnitude of different margins. The decomposition is conceptually valid under alternative assumptions about market structure

[3]While the firm size decomposition explicitly accounts only for direct linkages to firms' immediate buyers and suppliers, the complete network of firm-to-firm links is implicitly captured by the sourcing capability of all participants in the network.

[4]This variance decomposition is similar in spirit to Redding and Weinstein (2017).

(e.g. with or without monopolistic competition; with or without constant mark-ups). Finally, although we treat the production network as pre-determined, the approach produces unbiased estimates even if firms endogenously match based on firm-specific attributes or firm-pair specific matching shocks, so long as these shocks are not correlated with a pairwise sales residual. We perform exogenous mobility tests to rule out the latter.

We establish three main empirical results about the sources of firm size heterogeneity. All of these results hold both in the cross-section of firms at a given point in time and in the evolution of firm size within firms over time. First, 81% of the variation in the firm size distribution comes from the demand side and 19% from the supply side. Second, almost the entire variation on the demand side (99%) is generated by demand from downstream firms rather than by final demand. On the supply side, 88% of the variation reflects heterogeneity in firm production capability, with the remaining 12% determined by upstream input suppliers. Thus firm-to-firm linkages in the production network account for fully 82.5% of the firm size dispersion in the data. Together, these two results imply that trade in intermediate goods and firm-to-firm connections are essential to understanding firm-level performance and consequently aggregate outcomes. Models that feature only supply-side factors such as firm productivity or that ignore the input-output structure of the economy would thus fail to capture the vast majority of firm size heterogeneity.

Third, the extensive margin of buyer-supplier matches is the most important aspect of the production network for firm sales. The number of customers drives 71% of the variation in the network demand margin, i.e. sales to downstream buyers, while the number of suppliers explains 49% of the variation in the network supply margin, i.e. input price index based on upstream purchases. In other words, the main reason why the production network enables firms to sell more is because they can sell to more buyers, rather than because their buyers tend to purchase more intermediates. And the main reason why the production network helps firms reduce marginal cost or improve quality is because they can match with many suppliers, rather than because their suppliers are necesarily much better on average. The second most important dimension of the production network for firm size is the covariance term on the intensive margin. What matters is not that firms' buyers and suppliers are better on average, but that firms trade more with partners that are simultaneously more capable and better bilateral matches. On the supply side, firms benefit more when their production needs are especially well served by highly capable suppliers. On the demand side, firms sell more when their products are especially well suited to the production needs of highly capable buyers. These covariances account for 38% and 12% of the network supply and network demand margins, respectively.

This paper contributes to several strands of literature. Most directly, the paper adds to

the vast literature on the extent, causes and consequences of firm size heterogeneity. The vast dispersion in firm size has long been documented, with a 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 larger 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., Bernard et al (2012), Gabaix (2011), Freund and Pierola (2015), Gaubert and Itskhoki (2016)).

Traditionally, this literature has looked to own firm characteristics on the supply side as the driver of firm size heterogeneity. The evidence indicates an important role for firms' production efficiency, management ability, and capacity for quality products (e.g., Jovanovic (1982), Hopenhayn (1992), Melitz (2003), Sutton (2007), Bloom et al (2016)). 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 thereby performance (e.g., Goldberg et al (2010), Fieler et al (2014), Bernard et al. (2015a), Antras et al (2015)), while final-consumer preferences affect sales on the demand side (e.g., Foster et al. (2015), Fitzgerald et al. (2016)).

By contrast, we provide a comprehensive treatment of both own firm characteristics and production network features, on both the supply and demand sides. The paper is thus related to Hottman et al. (2016) who also find that demand rather than supply is the primary factor driving firm size dispersion. However, as they 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 compare own vs. network supply factors.

The paper also adds to a growing literature on buyer-supplier production networks (see Bernard and Moxnes (2017) for a recent survey). Due to the limited availability of data on firm-to-firm linkages, this literature has examined the formation of production networks primarily on the theoretical side. It has focused on the propagation of shocks through the network and its effects on aggregate volatility (e.g., Chaney (2014), di Giovanni et al (2014), Acemoglu et al (2012)). On the empirical side, Bernard et al. (2015a) study the impact of domestic supplier connections on firms' marginal costs and performance in Japan, whereas Bernard et al. (2015b) and Eaton et al (2016) examine the matching of exporters and importers using data on firm-to-firm trade transactions for Norway and US-Colombia, respectively. While the findings in these papers are consistent with ours, we

examine transaction-level data on a complete domestic production network and focus on the implications of two-sided heterogeneity and production networks for the firm size distribution. Using the same data as us, De Bruyne et al (2016) and Dhyne et al (2017) instead examine shock propagation and the link between domestic and imported inputs in firms' production process.

Finally, the methodology in this paper is related to the econometrics of two-sided heterogeneity in other economic contexts (see Bonhomme (2017) for a review). In particular, we estimate seller fixed effects, buyer fixed effects, and residual seller-buyer match effects from data on seller-buyer sales. This is similar in spirit to gravity models of international trade flows by exporting country - importing country pair, where exporter, importer and bilateral characteristics play a role (e.g., Helpman et al (2008)). Another recent contribution is Kramarz et al (2016), who estimate buyer and seller effects in a bipartite trade network. It also builds on employer-employee matching models in the labor literature (e.g., Abowd et al (1999), Card et al (2013)). However, each economic agent plays a unique role in the labor market - either a firm or a worker - such that both panel data and worker transitions across firms are necessary to identify the employer, employee and match effects. By contrast, each firm can in principle be both a buyer and a supplier in a production network, such that cross-sectional data is sufficient to identify the effects of interest.

The rest of the paper is organized as follows. Section 2 introduces the data and presents novel stylized facts. Section 3 outlines the theoretical framework. Section 4 operationalizes the first two steps of the estimation strategy to construct all necessary firm size components from the data. Section 5 presents the results of the firm size decomposition. Section 6 provides a general equilibrium formulation of the model which enables counterfactual welfare exercises in Section 7. The last section concludes.

2 A Complete Production Network: Data and Stylized Facts

2.1 Data

We exploit two comprehensive data sources on firm operations in Belgium over the 2002-2014 period. We are able to match these data at the firm level based on firms' unique ID that is common across datasets. This allows us to examine the complete domestic network of buyers and suppliers in Belgium using information on firm sales and production activity.

The primary data source is the NBB B2B (Business-To-Business) Transactions Database.[5]

[5]See Dhyne, Magerman and Rubinova (2015) for details on the construction of the NBB B2B Transactions Database.

It reports the value of all sales relationships between any two registered, VAT liable enterprises in Belgium. Administrative sanctions guarantee very high data quality standards. The unit of observation is the annual sum m_{ij} of sales invoices (in euro) from enterprise i to enterprise j. While we can obtain the primary industry of activity for each firm from the matched balance-sheet data, we do not know the specific product content of each transaction.[6] Coverage is almost universal, as all sectors except financial services and all annual sales m_{ij} above €250 are included. We thus have information both on the extensive margin of buyer-supplier links and on the intensive margin of bilateral sales in the production network.[7] In recent work, Bernard et al. (2015a) analyze the former using data on most firm-to-firm links in Japan constructed from each firm's top 24 suppliers and customers.

We match B2B to firm-level production statistics from the Balance Sheet Office (BSO).[8] BSO provides the annual accounts of nearly all active legal entities in Belgium. We extract the data on firms' total sales, input purchases, employment, labor costs, and main industry of activity at the NACE 2-digit level.[9]

Our firm size decomposition distinguishes between firm sales to final consumers and downstream producers, as well as between firms' domestic and foreign inputs from upstream suppliers. Combining data from B2B and BSO, we construct firms' sales to final demand (including exports) as the difference between their reported turnover and the sum of all their sales to other enterprises in the domestic production network. In a similar manner, we construct firms' purchases from outside the network (including imports) as the difference between their reported intermediate inputs and the sum of all their purchases from other enterprises in the domestic production network.

We provide further details on B2B and BSO data coverage and cleaning in the Data Appendix.

[6]To bridge theory and data, the model correspondingly assumes that firms assemble multiple inputs potentially sourced from multiple suppliers into a single product that they sell to other firms and to final consumers.

[7]Henceforth, we refer to the Belgian production network as the set of firm-to-firm connections in B2B, i.e. subject to the limitations across sectors and transaction size.

[8]While B2B covers all firm-to-firm transactions, the vast majority of firms included in B2B do not have to report full BSO annual accounts with turnover data because they are very small. In 2014 for example, there were around 800,000 firms in the B2B database. Our B2B-BSO matched sample includes about 95,000 of these firms with combined turnover accounting for 85 percent of total turnover for all firms in those sectors in Belgium.

[9]The raw BSO data is collected by tax year and has been annualized to match the calendar year definition used in B2B. The industry classification for firms' main economic activity has been harmonized over time to NACE Rev.2 (2008). Total input purchases are the sum of material and service inputs, and include both new inputs and net changes in input stocks. Total labor costs include wages, social security and pension contributions.

2.2 Stylized Facts

We document five stylized facts about firm size and firm linkages in the Belgian domestic production network.[10] These facts provide evidence that buyer-supplier relationships are key to understanding the firm size dispersion in an economy, and they motivate the subsequent theoretical and empirical analysis. In particular, they signal an important role for (i) downstream input demand relative to final output demand, (ii) the numbers of supplier and buyer partners of the firm, (iii) seller and buyer firm characteristics, and (iv) seller-buyer match characteristics. Accommodating these features in a unified theoretical framework thus requires the combination of two-sided firm heterogeneity and an input-output production network. This is precisely the approach we take. By contrast, existing models of one-sided firm heterogeneity such as differentiated firms producing only for final consumers cannot account for (ii)-(iv), while exisiting models of two-sided firm heterogeneity have so far ignored (i) or (iv).

Fact 1. *The distributions of firms' total sales, buyer-supplier connections, and buyer-supplier bilateral sales exhibit high dispersion and skewness.*

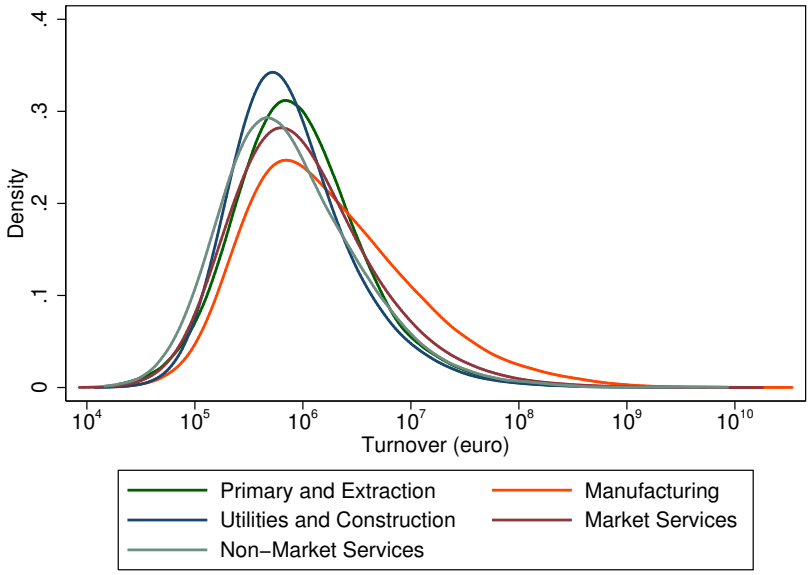
Firm size varies dramatically in Belgium, as in other countries. Table 1 provides summary statistics for firm sales in the cross-section for 2012, both overall and within five broad industries (primary and extraction, manufacturing, utilities, market services, non-market services). While average firm turnover was €8,6mil, firms at the 90th percentile generated turnover over 35 times higher than firms at the 10th percentile. The kernel density graphs in Figure 1 illustrate the full distribution of firm sales in the raw data by broad industry, as well as demeaned across firms within 4-digit NACE sectors.

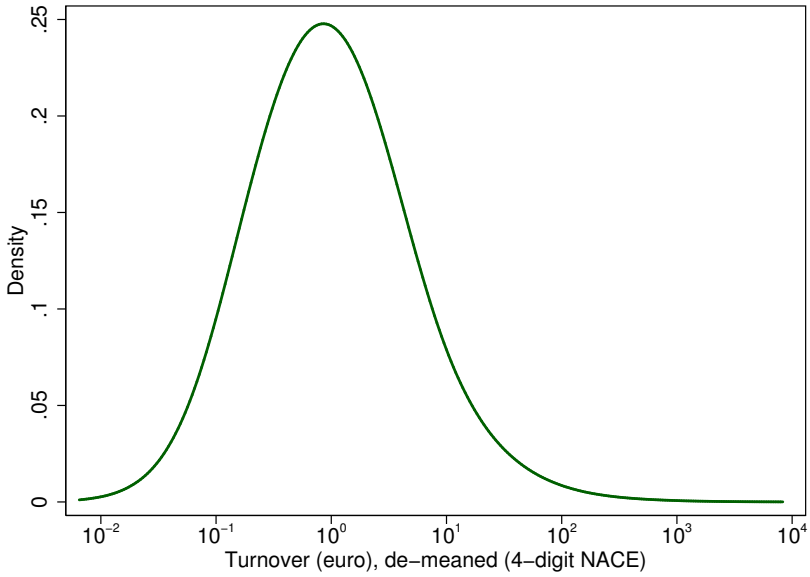
Turning to firm-to-firm connections in the production network, we also find that the number of customers per seller (*out-degree*) and the number of suppliers per buyer (*in-degree*) are very skewed. In 2012, we observe 16.8 million sales relationships between 812,543 enterprises, for an average of 20.68 connections per firm; see Table 2. Of these, 555,722 (68%) enterprises sell to other firms and 797,595 (98.2%) buy from other firms in the network. Thus approximately 32% of all enterprises sell only to final demand, and a very small minority of 1.8% do not purchase inputs from within the domestic network. Conditional on trading with others in the network, 89% of producers have more than one supplier and 75% have more than one buyer. The average connected firm has 30.23 customers and 21.06 suppliers. However, there is large dispersion around these means, with a high concentration of links among a few

[10]These stylized facts echo patterns established for the extensive margin of firm-to-firm linkages in the domestic production network in Japan (Bernard et al. (2015a)) and for both the extensive and the intensive margins of firm-to-firm export transactions in Norway (Bernard et al. (2015b)).

Figure 1: Firm sales distribution.

(a) Across sectors.





(b) Within NACE-4 sectors.

Table 1: Firm sales (€000s).

Industry	N	Mean	p10	p25	p50	p75	p90	p99
All	94,796	8,654	229	415	920	2,560	8,137	101,000
Primary and Extraction	1,481	3,032	217	403	829	2,069	5,033	42,900
Manufacturing	10,288	24,600	281	518	1,417	5,166	19,700	279,000
Utilities	18,708	4,386	216	354	669	1,566	4,270	41,100
Market Services	62,902	7,404	229	430	983	2,694	8,251	94,200
Non-Market Services	1,417	10,900	183	326	623	1,525	4,854	58,500

Note: Summary statistics for all firms in the matched B2B - CBO data in 2012.

Table 2: Number of firm buyers and suppliers.

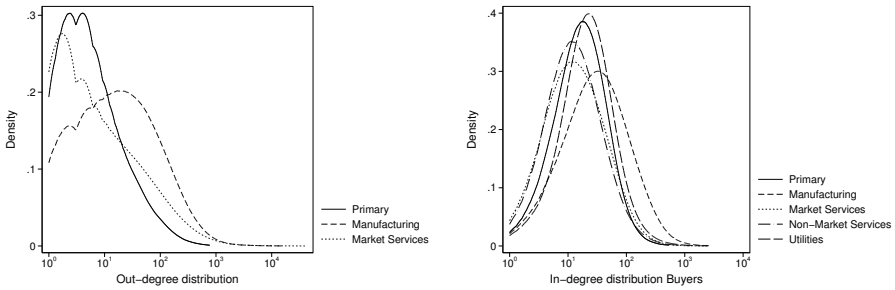
		N	Mean	Median
All	# Buyers	812,543	20.68	2
	# Suppliers	812,543	20.68	9
≥ 1	# Buyers	555,722	30.23	4
	# Suppliers	797,595	21.06	9

firms: The median number of customers and suppliers is only 4 and 9, respectively, while the top 1 percent of firms transact with more than 400 buyers and 177 sellers. This dispersion and skewness is also evident in the kernel density graphs in Figure 2. The distributions also differ across broad industries in line with priors, with both in-degree and out-degree highest for firms in manufacturing, intermediate in primary materials and extraction, and lowest in services. These patterns are consistent with those observed by Bernard et al. (2015a) and Bernard et al. (2015b) respectively for domestic firm-to-firm linkages in Japan and for Norwegian firms' export partners.

Table 3: Firm-to-firm transaction values (€s).

	N	Mean	Median
Transaction Value (€)	16,800,368	30,884	1,414

Figure 2: Distribution of firm buyer and supplier linkages.



(a) Out-degree (# buyers) distribution by industry.

(b) In-degree (# suppliers) distribution by industry.

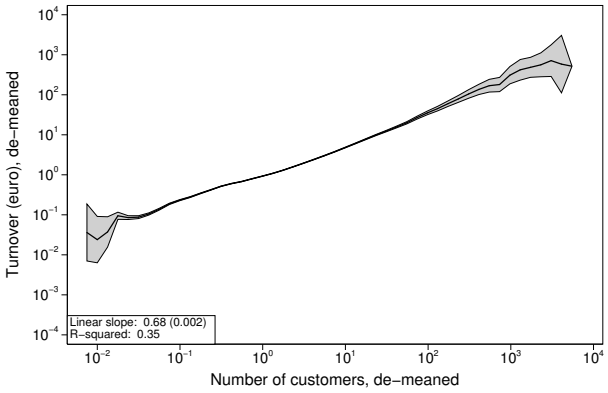
The intensive margin of firm-to-firm bilateral sales is again highly skewed, with the vast share of economic activity concentrated in a small number of buyer-supplier transactions. The mean transaction across the 16,800,368 buyer-supplier links in 2012 amounted to €30,884, see Table 3. At the same time, the median purchase totaled only €1,414. This extreme dispersion in transaction values in a buyer-supplier production network was first documented by Dhyne et al. (2015) using the same data for Belgium.

Fact 2. *Bigger firms have more buyers and suppliers.*

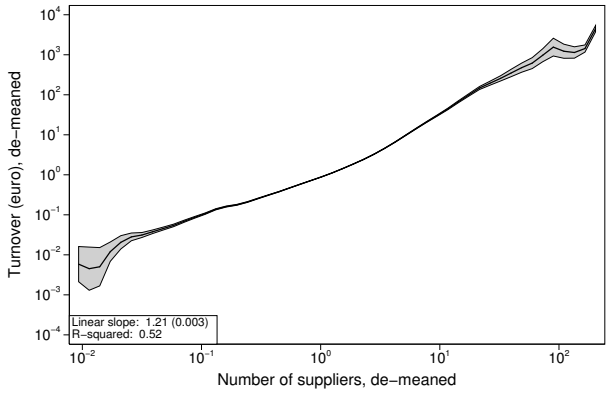
A sharp pattern in the data is that bigger firms transact with more buyers and suppliers in the production network. Figure 3a plots the fitted line and 90% confidence interval based on a local polynomial regression of firm turnover on the number of firm customers, on a log-log scale. Both variables have been first demeaned by their sector average, such that the latter corresponds to the point with coordinates (1,1) in the graph. The tightly estimated upward-sloping line implies that a firm with 10 times more customers than the sector mean has 7 times higher sales. Figure 3b repeats the exercise for the number of firms' input providers. A producer with 10 times more suppliers attains 12 times higher sales than the sector average.

Figure 3: Firm size and number of buyers and suppliers.

(a) Turnover and number of buyers (demeaned by industry).



(b) Turnover and number of suppliers (demeaned by industry).



The strong positive association between firm size and the extensive margin of firm connections is corroborated by the OLS regressions in Table 4. In Columns 1 and 2, we regress the log of firms' out-degree and in-degree on the log of firm turnover. Conditioning on a full set of NACE-5 industry fixed effects, we estimate elasticities of 0.517 and 0.427 that are significant at 0.1%.

Table 4: Firm size and number of firm connections.

	$\ln(out-degree_i)$	$\ln(in-degree_i)$	$\ln(out-degree_i)$
$\ln(size_i)$	0.517***	0.427***	
	(0.002)	(0.001)	
$\ln(in-degree_j)$			-0.067***
			(0.001)
R^2	0.519	0.625	0.448
N	192,635	277,150	7,447,816

Notes: Columns 1 and 2 include NACE-5 sector fixed effect for firm i. Column 3 includes NACE-5 sector fixed effect for both seller firm i and buyer firm j. Robust standard errors in parentheses. Significance: * p $<$ 5%, ** p $<$ 1%, *** p $<$ 0.1%.

Fact 3. *More connected firms match with less connected buyers and suppliers on average.*

The matching of buyers and suppliers in the production network is characterized by negative degree assortativity. This is evidenced by two patterns, one at the firm level and one at the level of firm-to-firm matches.

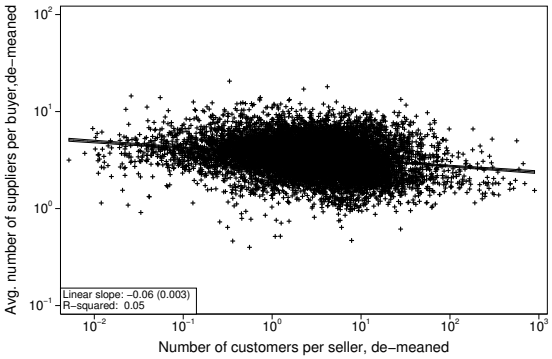
At the firm level, firms that sell to more customers have customers that on average buy inputs from fewer suppliers. This is illustrated in Figure 4a. For each firm i with $out-degree_i$,

we identify all of its customers j, obtain the number of suppliers that each of these customers buys from, $in-degree_j$, and finally calculate the average number of suppliers $\overline{in-degree_j}^i$ across i's customers j. Figure 4a plots $\overline{in-degree_j}^i$ against $out-degree_i$ on a log-log scale. The fitted line reveals a statistically significant negative slope of -0.12 in the raw data and -0.06 within NACE-2 industries.

Symmetrically, firms that source inputs from more suppliers have suppliers that on average sell to fewer buyers. This negative relationship between $in-degree_i$ and $\overline{out-degree_k}^i$ is displayed in Figure 4b, where $in-degree_i$ on the horizontal axis represents the number of firm i' suppliers and $\overline{out-degree_k}^i$ gives the average number of customers across i's suppliers k (TO BE ADDED).

Figure 4: Degree assortativity.

(a) Demeaned by NACE-2 industry.



Negative assortativity manifests also at the level of individual firm-to-firm matches. For each seller-buyer pair ij, we regress the number of i's customers, $out-degree_i$, on the number if j's suppliers, $in-degree_j$, controlling for i's and j's NACE-5 industry with two-way fixed effects. The correlation is negative, -0.067, and significant, see Column 3 of Table 4. Further analysis reveals that the negative assortative matching of buyers and suppliers reflects an imperfect pecking order based on firm size and connectedness: While large and well connected firms tend to buy from and sell to all types of firms, less connected firms transact only with large and well connected firms.

Fact 4. *The distribution of sales across buyers does not vary with the number of buyers, while the distribution of purchases across suppliers widens with the number of suppliers.*

Facts 2 and 3 pertain to the extensive margin of firm links in a buyer-supplier network, and reveal broadly symmetric patterns on the production side (i.e. firm suppliers) and the

sales side (i.e. firm buyers). In contrast, fact 4 characterizes the intensive margin of business-to-business transactions and uncovers asymmetry between purchases and sales. Specifically, the distribution of a firm's bilateral sales across customers does not vary systematically with the number of its customers. On the other hand, the distribution of a firm's input purchases across input suppliers widens monotonically with the number of its suppliers.

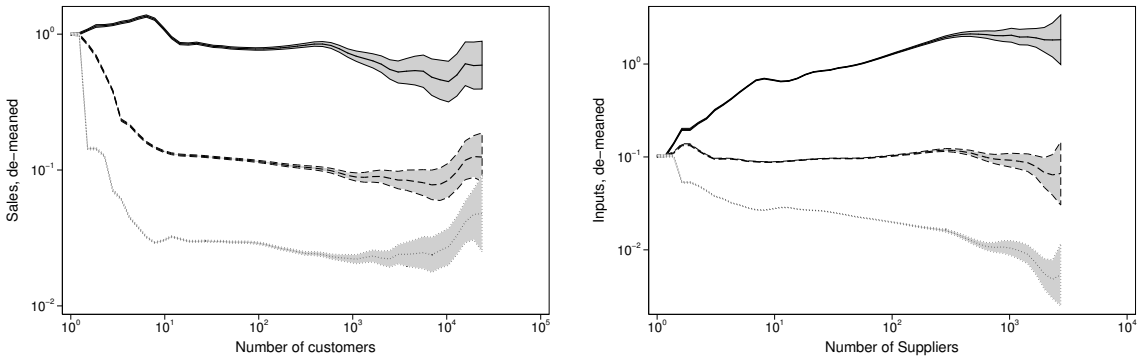
Figure 5a illustrates the sales side. For each firm i, we identify its customers j and order its bilateral sales by value m_{ij}. We capture the distribution of i's sales transactions with the 10th, 50th and 90th percentile of m_{ij} across buyers j, and demean by i's NACE-5 sector. Figure 5a shows the fitted lines from local polynomial regressions of these percentile values against i's out-degree, with bands for the 90% confidence interval. For firms with more than 10 customers, the three lines are almost flat, or slightly declining. Sales to the median (and 10th and 90th percentile) customer are essentially the same for firms with 10 customers and for firms wtih 1,000 customers.Together with fact 2, this invariance property implies that large firms have higher sales primarily because they serve more customers, but they neither sell more to their buyers nor vary their sales more across their buyers.

Figure 5b demonstrates firms' distribution of transaction values on the purchases side. For each firm i, we identify its input suppliers k and order its bilateral input purchases by value m_{ki}. We capture the distribution of i's input purchases with the 10th, 50th and 90th percentile of m_{ki} across suppliers k, and demean by i's NACE-5 sector. Figure 5b displays fitted lines and confidence bands based on local polynomial regressions of these percentile values against i's in-degree. Purchases from the median supplier are again invariant across firms with large and small supply bases. However, firms that source inputs from more suppliers purchase more from their largest suppliers and less from their smallest.

Figure 5: Transaction value distribution across firm buyers and suppliers.

(a) Out-degree and bilateral sales at 10th, 50th, 90th percentile.

(b) In-degree and bilateral purchases at 10th, 50th, 90th percentile.



Fact 5. *Firm-to-firm transaction value depends on seller characteristics, buyer characteristics, and seller-buyer match characteristics.*

Facts 2-4 isolate systematic patterns in the extensive and intensive margins of buyer-supplier linkages that point to an important role for two-sided heterogeneity in a production network. Fact 5 further suggests that not only heterogeneity on the buyer and seller side, but also heterogeneity in the buyer-seller match characteristics contributes significantly to the dispersion in firm-to-firm transactions and thereby in firm size. In Table **??**, we regress bilateral sales m_{ij} from seller i to buyer j on a full set of seller fixed effects in Column 1, buyer fixed effects in Column 2, and both seller and buyer fixed effects in Column 3. The R-squared from these regressions reaches A%, B%, and 38%, respectively. Since the fixed effects in the last specification absorb all relevant firm-specific characteristics on both sides of the transaction, any residual variation in transaction value must be attributed to the unique buyer-supplier match. The evidence therefore suggests that the heterogeneity in buyer-supplier match quality contributes fully 62% of the dispersion in bilateral transactions in a production network.

Summary. We have documented five stylized facts which suggest that buyer-supplier linkages in a production network are key to understanding the origins of the firm size distribution. Together they signal an important role for (i) downstream input demand (as distinct from final demand), (ii) the numbers of firms' buyers and suppliers,(iii) seller- and buyer-specific firm characteristics, and (iv) seller-buyer match characteristics. Motivated by these stylized facts, we next develop a unified theoretical framework with these features that introduces two-sided firm heterogeneity in an input-output production network and allows us to decompose the variation in firm size into economically meaningful components.

3 Theoretical Framework

This section develops a theoretical framework that serves several purposes. First, the model allows for various sources of firm heterogeneity both on the demand size (e.g., being connected to many or large customers) and the supply side (e.g., having access to cheap intermediate inputs). Second, the framework gives a clear mapping between model parameters and firm-level estimated coefficients from production network data. Section 4 below describes the identification and estimation of those coefficients. Third, the framework allows a decomposition of firm sales into various demand and supply side margins (a model-based decomposition). And finally, the model can be used for counterfactual analyses (Section 6).

Our starting point is a model where firms are heterogeneous in productivity or quality, as in Melitz (2003). Firms sell to other firms and to final demand, and how many and which

buyers they meet will affect firm size. In addition, firms source inputs from one or more suppliers, and those input prices will determine output prices and consequently also firm sales. Since the main aim of the paper is to understand the role of the network in generating heterogeneity, we take the observed production network as given, i.e. we do not model the firm-to-firm matching decision itself.

3.1 Technology

To implement our approach, we start with the following production function of firm i:

$$y_i = z_i l_i^{\alpha} v_i^{1-\alpha},$$

where y_i is output, z_i is productivity, l_i is labor and α is the labor share. v_i is a constant elasticity of substitution (CES) input bundle:

$$v_i = \left(\sum_{k \in \mathcal{S}_i} \left(\phi_{ki} \nu_{ki} \right)^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)},$$

where ν_{ki} 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. ϕ_{ki} is a demand shifter that captures the idea that firms (and industries) may have very different production technologies, and that their purchases from a given supplier may vary greatly. We also allow for heterogeneity in the labor share α across industries, however for ease of notation we drop industry subscripts for now. The corresponding input price index is $P_i^{1-\sigma} = \sum_{k \in \mathcal{S}_i} \left(p_{ki} / \phi_{ki} \right)^{1-\sigma}$, where p_{ki} is the price charged by supplier k to firm i. The marginal costs of the firm is then

$$c_i = \frac{\kappa w^{\alpha} P_i^{1-\alpha}}{z_i}, \tag{1}$$

where κ is a constant.[11]

3.2 Firm-to-Firm Sales, Total Sales and Purchases

Each firm is facing demand from other firms as well as from final demand. Given the assumptions about technology, sales from firm i to j are

$$m_{ij} = \left(\frac{\phi_{ij}}{p_{ij}} \right)^{\sigma-1} P_j^{\sigma-1} M_j, \tag{2}$$

where M_j is total intermediate purchases of firm j. In the baseline decomposition, final demand is directly observed as the difference between total sales and firm-to-firm sales, and

[11] $\kappa = \alpha^{-\alpha} \left(1-\alpha\right)^{\alpha-1}$.

as such it is unnecessary to model it explicitly. In this part of the paper, we therefore take final demand as given and denote it by $\mathcal{F}_i$, while Section 6 extends the model with endogenous final demand. Firm i's sales to other firms in the network are therefore

$$\beta_i^c S_i \equiv \sum_{j \in \mathcal{C}_i} m_{ij} \tag{3}$$

where S_i is total sales (including exports), $\beta_i^c \equiv (S_i - \mathcal{F}_i)/S_i$ is the share of sales going to other firms in the network and $\mathcal{C}_i$ is the set of customers of firm i. In our data, we only observe firm-to-firm links in the domestic economy. Hence, demand from foreign firms (exports) is classified as final demand.[12] In a similar manner, we write firm j's purchases from other firms in the network as

$$\beta_j^s M_j \equiv \sum_{i \in \mathcal{S}_j} m_{ij}, \tag{4}$$

where M_j is total purchases (including imports), β_j^s is the share of purchases from other firms in the network and $\mathcal{S}_j$ is the set of domestic suppliers of firm j.

In the following, it will be useful to collapse parameters that are related to either the buyer, the seller, or the match. We assume that the match quality term ϕ_{ij} can be written as $\phi_{ij} = \phi_i \tilde{\phi}_{ij}$, where ϕ_i captures the average quality of firm i and $\tilde{\phi}_{ij}$ is an idiosyncratic match term. In a similar fashion, we assume that the price p_{ij} can be written as $p_{ij} = \tau_i \tilde{\tau}_{ij} c_i$, where c_i is marginal costs, τ_i captures the average mark-up and trade cost of i, and $\tilde{\tau}_{ij}$ is the match-specific trade cost/mark-up term.[13] Note that τ_{ij} could reflect any type of price variation, e.g. heterogeneity in mark-ups across customers. Equation (2) can therefore be rewritten to

$$m_{ij} = \psi_i \theta_j \omega_{ij}, \tag{5}$$

where $\psi_i \equiv (\phi_i/(\tau_i c_i))^{\sigma-1}$ is a seller effect, $\theta_j \equiv P_j^{\sigma-1} M_j$ is a buyer effect and $\omega_{ij} \equiv \left(\tilde{\phi}_{ij}/\tilde{\tau}_{ij}\right)^{\sigma-1}$ is a match effect.

The exact decomposition in Section 3.4 only requires the assumptions described so far (the production function and the functional forms of p_{ij} and ϕ_{ij}). In particular, there is no need to assume anything about market structure, firms' pricing behaviour or the elasticity of substitution. However, a few additional elements are required to solve the general equilibrium and to perform counterfactuals. We therefore introduce those assumptions when needed in Section 6.

[12] Final demand is calculated as $\mathcal{F}_i = S_i - \Upsilon_i$, i.e. the difference between overall sales according to the balance sheet data and sales to other firms according to the transaction data, see Section **??**.

[13] In the empirical application, $\tilde{\phi}_{ij}$ and $\tilde{\tau}_{ij}$ will be normalized such that $(1/n_i^c)\sum_{j \in \mathcal{C}_i} \tilde{\phi}_{ij} = 1$ and $(1/n_i^c)\sum_{j \in \mathcal{C}_i} \tilde{\tau}_{ij} = 1$, where n_i^c is the number of customers of firm i.

3.3 Mapping Buyer and Seller Effects to Model Parameters

Section 4 describes how we can estimate the firm-level parameters ψ_i, θ_i and ω_{ij} from production network data. Given knowledge about those parameters and total input spending M_i, we can invert the expressions for ψ_i, θ_i and ω_{ij} and back out structural parameters. Using equation (1) and rearranging yields:

$$P_j^{\sigma-1} = \frac{\theta_j}{M_j} \tag{6}$$

$$\left(\frac{\phi_i z_i}{\kappa \tau_i w^\alpha}\right)^{\sigma-1} = \psi_i \left(\frac{\theta_i}{M_i}\right)^{1-\alpha} \tag{7}$$

$$\left(\frac{\tilde{\phi}_{ij}}{\tilde{\tau}_{ij}}\right)^{\sigma-1} = \omega_{ij}. \tag{8}$$

Loosely speaking, the buyer effect θ_j reflects the magnitude of average purchases controlling for the size of suppliers. We refer to θ_j as *sourcing capability*. Hence, if the buyer effect is small and total purchases M_j are large, it must mean that purchases are spread out over many suppliers. The input price index P_j in equation (6) must then be small.

Conversely, the seller effect ψ_i reflects the magnitude of average sales controlling for the size of customers. After adjusting for input costs $P_i^{1-\alpha}$, ψ_i thus identifies the productivity (z_i) or average attractiveness (ϕ_i/τ_i) of the seller in equation (7). In the following, it will become useful to define the left hand side of equation (7) as $\tilde{z}_i \equiv \left(\phi_i z_i / \left(\kappa \tau_i w^\alpha\right)\right)^{\sigma-1}$. We refer to $\tilde{z}_i$ as *production capability*. Note that according to our model, the seller and buyer effects for a firm i are negatively related. Rearranging equation (7), we get $\psi_i = \tilde{z}_i \left(M_i/\theta_i\right)^{1-\alpha}$. Hence, a marginal increase in the buyer effect reduces the seller effect by $1-\alpha$. This occurs because a higher buyer effect, all else constant, implies higher input costs. This translates into higher output prices and therefore lower sales. We will test this prediction in Section 3.3.[14] Finally the match-specific component ω_{ij} jointly captures the demand (taste) shifter $\tilde{\phi}_{ij}$ and the supply (mark-up, trade cost) shifter $\tilde{\tau}_{ij}$ in equation (8).

3.4 An Exact Firm Sales Decomposition

In this section, we develop an exact decomposition of firm sales into different margins relating to the demand and supply side network.

[14] While the model predicts a relationship between the buyer and seller effects, it has no prediction about the relationship between productivity/quality $\phi_i z_i$ and sourcing capability P_i. In frameworks with endogenous network formation, the correlation between the two will generally depend on the matching model and the market structure.

Combining equations (3) and (5) above, log total sales are

$$\ln S_i = \ln \psi_i + \ln \xi_i - \ln \beta_i^c, \tag{9}$$

where $\xi_i \equiv \sum_{j \in \mathcal{C}_i} \theta_j \omega_{ij}$.

ψ_i and ξ_i represent supply and demand side fundamentals in explaining firm size, respectively, while β_i^c represents the importance of final demand. As we show below, we can identify $\ln \psi_i$, $\ln \theta_i$ and $\ln \omega_{ij}$ from the production network data. Furthermore, $\ln S_i$ and $\ln \beta_i^c$ are directly observed in our data. Hence, all components of equation (9) are known.

In order to assess the importance of each margin, we follow the literature (Eaton, Kortum, Kramarz, 2004, Hottman et al., 2016) and regress each component ($\ln \beta_i^c$, $\ln \psi_i$ and $\ln \xi_i$) on log sales. By the properties of ordinary least squares, the sum of those three coefficients will sum to unity, and the coefficient magnitudes will represent the share of overall variation in firm size explained by each margin.

We can further decompose the supply and demand side margins into various sub-margins. Let us start with the demand side. The demand side parameter $\ln \xi_i$ can be rewritten as

$$\ln \xi_i = \ln n_i^c + \ln \bar{\theta}_i + \ln \Omega_i^c, \tag{10}$$

where n_i^c is the number of customers and $\bar{\theta}_i \equiv \left(\prod_{j \in \mathcal{C}_i} \theta_j \right)^{1/n_i^c}$.[15] The covariance term Ω_i^c is defined as

$$\Omega_i^c \equiv \frac{1}{n_i^c} \sum_{j \in \mathcal{C}_i} \omega_{ij} \frac{\theta_j}{\bar{\theta}_i}.$$

What is the economic interpretation of these components? First, firms face high demand if they are linked to many customers (high n_i^c). Second, they face high demand if the average customer has high sourcing capability (high $\bar{\theta}_i$). Third, they face high demand if the covariance term Ω_i^c is large. This would be the case if large customers (high θ_j) also happen to be a good match (high ω_{ij}). As with the overall decomposition, we will regress each component in equation (10) on $\ln \xi_i$.

Next, we turn to the supply side decomposition. A firm may be large because it has high production capability (high $\tilde{z}_i$), or because it benefits from cheap or high-quality inputs (low P_i). Just as above, the input price index can be decomposed into a component for the number of suppliers, average capability of a supplier and a covariance term. This can be shown in three steps. First, from the inversion in equation (7), the production capability of a firm, $\tilde{z}_i$, is a function of the estimated buyer and seller effects. Second, combining equations

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

(4) and (5) above, log total purchases are

$$\ln M_j = \ln \theta_j + \ln \sum_{i \in \mathcal{S}_j} \psi_i \omega_{ij} - \ln \beta_j^s. \tag{11}$$

Third, solving equation (7) for $\ln \psi_i$ and substituting $\ln (M_i/\theta_i)$ using equation (11) above yields

$$\ln \psi_i = \ln \tilde{z}_i + (1-\alpha) \left[\ln n_i^s + \ln \bar{\psi}_i + \ln \Omega_i^s - \ln \beta_i^s \right], \tag{12}$$

where n_i^s is the number of suppliers, $\bar{\psi}_i \equiv \left(\prod_{k \in \mathcal{S}_i} \psi_k \right)^{1/n_i^s}$ and the covariance term Ω_i^s is

$$\Omega_i^s \equiv \frac{1}{n_i^s} \sum_{k \in \mathcal{S}_i} \omega_{ki} \frac{\psi_k}{\bar{\psi}_i}.$$

Detailed derivations are found in Appendix A.1. Again, each component of this expression is either observed directly (α, β_i^s and n_i^s) or can be estimated from the production network data ($\tilde{z}_i$, ψ_i, $\bar{\psi}_i$ and Ω_i^s). The interpretation of each element is as follows. A firm has a large market share among customers (high ψ_i) because it is inherently productive or is high quality (high $\tilde{z}_i$), because it has many suppliers (large n_i^s), because those suppliers are on average attractive suppliers (high $\bar{\psi}_i$), or if attractive suppliers also happen to be a good match (high Ω_i^s). As with the overall decomposition, we regress each component in equation (12) on $\ln \psi_i$. The coefficient estimates will mechanically sum to one because the left and right hand side of equation (12) are by construction identical.[16]

Summing up, the exact decomposition consists of three steps. *Step One* is to estimate buyer and seller effects from the production network data. *Step Two* is to calculate the firm-level parameters from those estimates ($\ln \xi_i$, $\ln \tilde{z}_i$, $\ln \bar{\theta}_i$, $\ln \bar{\psi}_i$, $\ln \Omega_i^s$ and $\ln \Omega_i^c$). *Step Three* is the variance decomposition itself, regressing each component on $\ln S_i$, $\ln \xi_i$ and $\ln \psi_i$ using equations (9), (10) and (12), respectively.

We summarize the overall decomposition of firm size in the equation below to faciliate the economic interpretation of each component and for ease of reference in the empirical analysis. Firm size is determined by a supply factor and a demand factor. The supply factor comprises own production capability and network supply, where the latter constitutes the number of input suppliers, average production capability across suppliers, and a covariance term. The demand factor includes final demand and network demand, where the latter contains number of downstream customers, average sourcing capability across customers,

[16] This holds for any α. A change in α, e.g. due to measurement error, would lead to different coefficient estimates of each component, but the components would still sum to one.

and a covariance term.

$$\overset{Size}{\ln S_i} = \overbrace{\underbrace{\ln \tilde{z}_i}_{ProdCapability} + (1-\alpha)\overbrace{\left[\underbrace{\ln n_i^s}_{\#Suppliers} + \underbrace{\ln \bar{\psi}_i}_{AvgSupProdCapab} + \underbrace{\ln \Omega_i^s}_{Covariance} - \underbrace{\ln \beta_i^s}_{NetworkInputShare}\right]}^{NetworkSupply}}^{\overset{Supply}{\ln \psi_i}} + \overbrace{\underbrace{\ln \xi_i - \ln \beta_i}_{\overset{Demand}{\ln \xi_i}}}^{\overbrace{\underbrace{\ln n_i^c}_{\#Customers} + \underbrace{\ln \bar{\theta}_i}_{AvgCustSourceCapab} + \underbrace{\ln \Omega_i^c}_{Covariance}}^{NetworkDemand} - \overset{FinalDemand}{\ln \beta_i^c}} \quad (13)$$

The limited assumptions we have placed on the economic environment imply that this is an agnostic firm size decomposition that allows us to evaluate the contribution of different margins to the overall variation in firm size. Our approach imposes no restrictions on the absolute and relative contribution of these margins. In particular, we have not explicitly modeled the endogenous formation of the production network, and we do not aim to explain why some firms match with more or more capable buyers and suppliers. Instead, our goal is to understand how these implicit firm decisions account for the observed firm size distribution.

4 Estimation

The exact firm size decomposition consists of three steps. In *Step One*, we estimate buyer, seller and match effects from the production network data ($\ln \psi_i$, $\ln \theta_j$ and $\ln \omega_{ij}$). In *Step Two*, we calculate unobserved firm outcomes from the first-stage estimates and observed firm outcomes ($\ln \xi_i$, $\ln \tilde{z}_i$, $\ln \bar{\theta}_i$, $\ln \bar{\psi}_i$, $\ln \Omega_i^s$ and $\ln \Omega_i^c$). In *Step Three*, we perform the variance decomposition itself, regressing each component of firm size on total sales $\ln S_i$, the demand factor $\ln \xi_i$ and the supply factor $\ln \psi_i$, using equations (9), (10) and (12), respectively.

We discuss the first two steps of the econometric analysis in this section and present the firm size decomposition in Section 5. Our ultimate goal is to understand the cross-sectional variation in firm size at a given point in time, as well as the relative importance of different margins to changes in firm size over time. Since the production network continuously evolves, we therefore perform *Step One* and *Step Two* separately for each year in the 2002-2014 sample period. We report detailed results for these two steps for the most recent year in the data, 2014. The patterns for other years are not systematically different and are available upon request.

4.1 Step One: Buyer, Seller and Match Effects

Our first step is to estimate the buyer, seller and buyer-seller match effects from the B2B data on the Belgian domestic production network. This step exploits the granularity of firm-to-firm transanctions to inform the microfoundations of firm size in a way that would be impossible without such rich data.

We estimate a two-way fixed effects specification for firm-to-firm sales based on equation (5):

$$\ln m_{ij} = \ln \psi_i + \ln \theta_j + \ln \omega_{ij}. \tag{14}$$

In this OLS regression, the seller effect $\ln \psi_i$ is identified from the variation in input purchases across the suppliers of an average buyer. Intuitively, attractive suppliers account for a large share of input expenditures for all their downstream customers and receive a high $\ln \psi_i$. Analogously, the buyer effect $\ln \theta_j$ is identified from the variation in sales across the customers of an average producer. Intuitively, attractive buyers purchase a disproportionate share of upstream suppliers' sales and receive a high $\ln \theta_j$. The estimated residual $\ln \omega_{ij}$ is by construction orthogonal to the fixed effects. It thus 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. We let $\ln \omega_{ij} = \gamma \ln d_{ij} + e_{ij}$, where d_{ij} is observed match frictions and e_{ij} is unobserved ones. We proxy d_{ij} with distance in kilometers between the zip codes of the buyer and the seller.[17]

In order to estimate the two-way fixed effects model, firms must have multiple connections. Specifically, identification of a seller fixed effect requires a firm to have at least two customers, and identification of a buyer fixed effect requires a firm to have at least two suppliers. Furthermore, dropping customer A might result in supplier B having only one customer left. Supplier B would then also be dropped from the sample. This avalanching process reduces the sample even further.[18] In practice, the estimation sample retains a substantial portion of the production network. For the baseline year, 2014, we are able to use 5,589,506 firm-to-firm transactions which capture 32.3% of all links in the data and 50.6% of their total sales value. We thus obtain seller fixed effects for 96,983 firms and buyer fixed effects for 109,720 firms. We report the characteristics of the initial and estimation samples in Table 5 below.

[17] In the model, $\ln \omega_{ij}$ combines bilateral trade costs, demand shocks (e.g. how well the seller's product fits the production needs of the buyer), and variable mark-ups. To the extent that bilateral distance proxies trade costs, the residual e_{ij} captures the sum of match-specific demand and match-specific mark-up.

[18] We have to drop some small firms whose location is unobserved because we have no information on the bilateral distance to their trade partners. Our results are however robust to including them in the analysis by estimating equation (5) without including distance as a covariate.

Table 5: Full sample vs. first-stage estimation sample (2014).

	Full Sample			Estimation Sample		
	N	Mean	St Dev	N	Mean	St Dev
Firm Level						
Total Sales	109,908	6,740,855	1.45e+08	95,529	7,374,816	1.54e+08
# Suppliers	840,607	20.6	49.5	95,529	54.9	76.4
# Buyers	590,271	29.3	394.0	95,529	56.8	307.9
Transaction Level						
Transaction Value	17,304,408			5,589,506		

Note: Summary statistics for all firms and firm-to-firm transactions in the raw B2B data and in the estimation sample in *Step One*. Data on total sales is available only for firms in the matched B2B-CBO data.

Figure 6 summarizes the estimation results for 2014. Three patterns stand out. First, the variation in the buyer effect $\ln \theta_j$ is large compared to that in the seller effect $\ln \psi_i$. This anticipates the significantly bigger role we find for demand-side factors in explaining the firm size distribution compared to supply-side factors. Second, the R^2 from the regression is 0.38. Together with the large dispersion in the residual $\ln \omega_{ij}$, this signals the importance of buyer-supplier match quality to the value of firm-to-firm sales. Finally, while the estimation imposes no constraints on the relationship between the buyer and seller fixed effects for a given firm i, the theoretical framework implies that they should be negatively correlated, see Section 3.3. In the model, this occurs because firms that have higher input purchases and more suppliers can better allocate inputs towards more capable suppliers. By reducing their quality-adjusted production costs, this makes such firms more attractive suppliers to other firms in the network and thereby increases their sales. Our results confirm that this is borne out in practice: The correlation between $\ln \psi_i$ and $\ln \theta_j$ is -0.10 and significant at 1%.

4.2 Conditional Exogenous Mobility

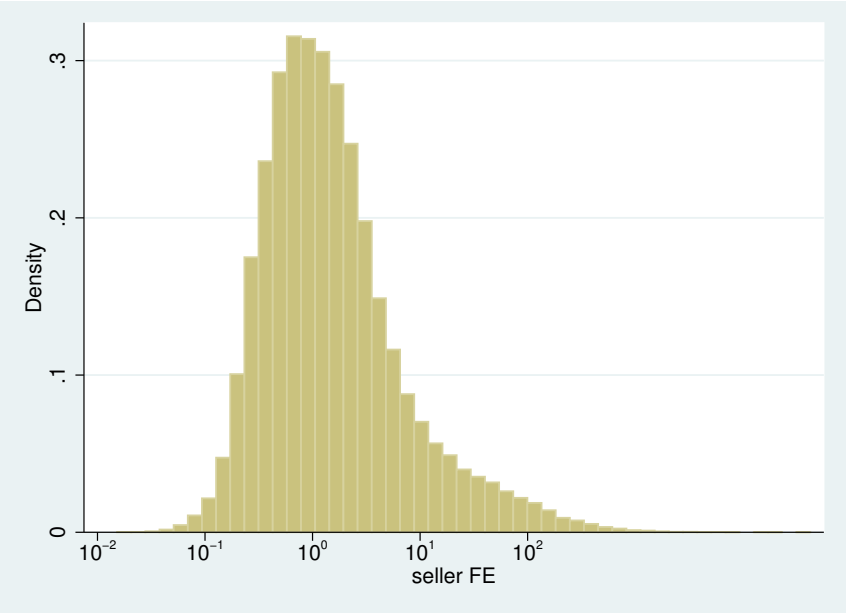
Equation (14) is a two-way fixed effects model similar to the models that are used in the employer-employee literature following Abowd et al (1999). The identifying assumptions needed for ordinary least squares to identify the parameters $\ln \psi_i$ and $\ln \theta_j$ are also similar:

$$E\left[\ln \psi_i e_{ij}\right] = E\left[\ln \theta_j e_{ij}\right] = E\left[\ln d_{ij} e_{ij}\right] = 0.$$

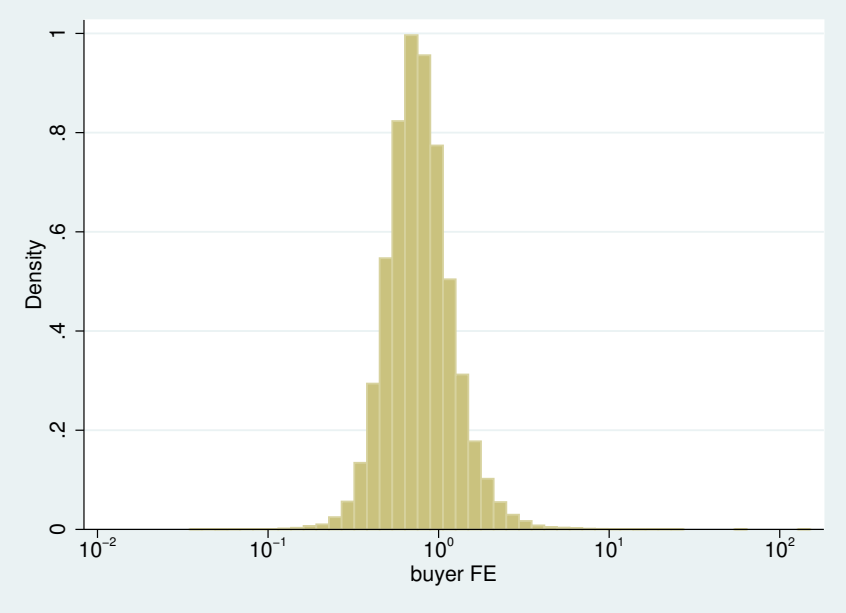
Hence, both seller and buyer fixed effects and match-specific observed variables must be orthogonal to the error term e_{ij}. As is well known from the employer-employee literature, the

Figure 6: Distribution of seller and buyer effects (demeaned by NACE-2 sector, 2014).

(a) Seller Fixed Effect, $\ln \psi_i$. N=96,983.



(b) Buyer Fixed Effect, $\ln \theta_j$. N=109,720.



key identification assumption is that the assignment of suppliers to customers is exogenous with respect to e_{ij}, so-called conditional exogenous mobility. Hence, the identification assumption is violated if a positive shock both increases the likelihood of matching and raises e_{ij}. It is instructive to review some cases where this assumption holds. First, the identifying assumption holds if firms match based on supplier and customer capability, i.e. they match based on their buyer and seller effect. Second, it holds if firms match based on observed trade costs d_{ij} (and those trade costs potentially matter for sales m_{ij}). It would also hold if firms match based on unobserved fixed costs (and those costs do not matter for sales). The models of Bernard et al (forthcoming) and Lim (2017) are examples of the first and second cases. Third, the assumption holds if firms match based on idiosyncratic pair-wise shocks and those shocks are unrelated to the error term e_{ij}. Eaton et al (2015) develop a quasi-random matching model which would be consistent with the third case.

However, we cannot completely rule out the possibility that matching shocks are also correlated with sales shocks. We therefore also test for conditional exogenous mobility, using a methodology inspired by Card et al (2015). The key idea is to check whether a switch from a small to a large customer increases sales, while a switch from a large to a small lowers sales, and that these changes are of equal magnitude in absolute value. Under the exogenous mobility assumption, the expected change in sales when moving from a customer k to j is identical to the change when moving from j to k (in absolute value):

$$E\left[\ln m_{ij} - \ln m_{ik}\right] = -E\left[\ln m_{ik} - \ln m_{ij}\right] = \ln \theta_j - \ln \theta_k.$$

Intuitively, if exogenous mobility fails, then a switch from large to small may not result in a large sales decline because both matching and sales are driven positive unobserved shocks. The methodology in Card et al (2015) cannot be adopted directly to our setting, because firms have many connections both upstream and downstream, while in employer-employee data a worker is typically only associated with one employee at the same time. We therefore proceed as follows. First, we estimate the fixed effects model from equation (14) for the 2005 cross-section ($t = 0$). We group firms into quartiles based on the magnitude of their estimated buyer effect. The quartiles are denoted by q_k, $k = 1, 2, 3, 4$. Second, we consider the set of firms that have a at least one q_1 buyer in $t = 0$ and add at least one q_4 buyer in $t = 1$, i.e. *upgraders*. For each upgrading firm, we calculate the change in sales when moving from a q_1 to a q_4 customer, $\ln m_{ij(q_4),t=1} - \ln m_{ij(q_1),t=0}$, where $j\,(q_k)$ denotes a customer in the q_k'th quartile. Since firms may add many q_4 buyers in $t = 1$ (and potentially have many q_1 buyers in $t = 0$), we form the average of all possible combinations, and denote the average by $\bar{\Delta^{Up}}_i$. Third, we take the average of $\bar{\Delta}_i^{Up}$ across all upgraders. In a similar way, we calculate the outcomes among firms that have a q_4 buyer in $t = 0$ and add a q_1 buyer

in $t = 1$, i.e. *downgraders*, and denote the firm-level change $\bar{\Delta}_i^{Down}$. The mean of $\bar{\Delta}_i^{Up}$ is 0.49 and the mean of $\bar{\Delta}_i^{Down}$ is -1.46. Hence, the results suggest that there is assymmetry in upgrading and downgrading. However, the assymmetry has the opposite sign of what would be expected from endogenous mobility. Specifically, endeogenous mobility would imply that downgrading leads to a smaller change in sales compared to upgrading (in absolute value).

4.3 Step Two: Firm Size Components

In the first step of the econometric analysis, we estimated equation (5). In the second step, we now interpret the buyer, seller and match effects from the first step through the lens of the model. The only assumption required for this interpretaion is on the production technology available to firms, and we do not need additional restrictions on consumer demand or the market structure. In particular, we combine the estimates from the first stage ($\ln \psi_i$, $\ln \theta_j$ and $\ln \omega_{ij}$) with observed measures of firm activity ($\ln S_i$, $\ln M_i$, $\ln \beta_i^c$, $\ln \beta_i^s$, $\ln n_i^c$, $\ln n_i^s$, $\mathcal{C}_i$ and $\mathcal{S}_i$) to back out model-consistent measures for unobserved firm attributes necessary for the firm size decomposition ($\ln \xi_i$, $\ln \tilde{z}_i$, $\ln \bar{\psi}_i$, $\ln \bar{\theta}_i$, $\ln \Omega_i^s$ and $\ln \Omega_i^c$).

To construct the observed firm metrics we combine Business-to-Business (B2B) network data with data from the Central Balance Sheet Office (CBSO) in Belgium. Firms below a certain size threshold submit abbreviated annual accounts to CBSO and do not have to report turnover. We therefore perform the decomposition for the 95,529 firms with complete data for the purposes. These firms account for 85% of gross output. Importantly, excluded firms with missing annual accounts are still part of the first step and thus included in the buyer and supplier margins of firms in the decomposition sample.

We measure firm sales S_i with total reported turnover from CBSO. The network sales share, β_i^c, is calculated as the sum of all sales to other firms in the domestic network from B2B divided by total sales. Since we observe firm-to-firm links only within Belgium, sales to foreign firms (exports) are classified as final demand.[19] We measure firm purchases M_i with total input expenditures from CBSO.[20] The network input share, β_j^s, is then the cost share of inputs from suppliers in the domestic network divided by total purchases. We obtain directly from B2B the number n_i^c and the set $\mathcal{C}_i$ of firms' domestic customers, as well as the number n_i^s and the set $\mathcal{S}_i$ of firms' suppliers.

[19]This assumption reduces the importance of firm-to-firm sales as almost all international trade is between firms. Using the BEC classification, around 2/3 of Belgian exports are in intermediate goods.

[20]We consider all import transactions to be purchases of foreign inputs to production. For the purposes, even if a firm imports final rather than intermediate goods and services to sell alongside its own-manufactured products, such imports are part of its overall expenses in serving downstream buyers and final consumers. Their contribution to the firm size decomposition will therefore be the same as that of other production inputs. See Bernard et al (2017) on the role of carry-along trade.

Using the first-stage estimates, the observed variables just described, and equations (1), (6), (7) and (8), we solve for firms' unobserved production capability $\ln \tilde{z}_i$, input price index $\ln P_i$ and marginal production costs $\ln c_i$. This requires three parameter values: the labor share α in the Cobb-Douglas production technology, the wage rate w, and the elasticity of substitution σ. To accommodate the variation in production technologies and factor costs across industries, we proxy α with the ratio of the total wage bill across all firms operating in a NACE-2 industry to the total production costs of all firms in that industry. We likewise measure w with the total wage bill in an industry, divided by total employment in that industry. We take a standard value for σ from the literature and set it equal to 4. This choice is however not consequential given the log-linear specification of the OLS regression in *Step One* and the fact that we demean all firm size components by industry after *Step Two*.

Finally, we back out each firm's network demand $\ln \xi_i$, the average production cability of its suppliers $\ln \bar{\psi}_i$, the average sourcing capability of its buyers $\ln \bar{\theta}_i$, its supply and demand covariance terms $\ln \Omega_i^s$ and $\ln \Omega_i^c$. This completes the second step of the econometric analysis, as we now have measures for all firm size components. Table 6 below provides summary statistics for these components.

5 Results

In the last step of the econometric analysis, we perform the firm size decomposition according to equation (13). In Section 5.1, we begin by analyzing the contribution of different supply- and demand-side margins to the cross-sectional variation in firm size in the baseline year, 2014. In Section 5.2, we repeat this cross-sectional decomposition separately for each year to examine time trends in the 2002-2014 panel. In Section 5.3, we turn to the evolution of firm size within firms over time and evaluate how changes along different supply and demand margins shape firm growth. Finally, in Section 5.4 we subject the results to sensitivity analysis and explore the variation in patterns across sectors and different segments of the size distribution.

A potential concern is that industries are inherently different, and those differences may be systematically related to supply or demand side characteristics. We therefore demean all observed and constructed variables by their NACE 2-digit industry average after the second step. For example, the overall decomposition from (9) becomes

$$\Delta_S \ln S_i = \Delta_S \ln \psi_i + \Delta_S \ln \xi_i - \Delta_S \ln \beta_i^c,$$

where Δ_S denotes the difference between the outcome of firm i and the average outcome in that sector. We then regress each component, e.g. $\Delta_S \ln \psi_i$, on $\Delta_S \ln S_i$. The baseline

Table 6: Firm size components (demeaned by NACE-2 sector, 2014).

Firm Size Component	Estimated?	N	Mean	Median	St Dev
Total Sales, $\ln S_i$		95,529	0.000	-0.131	1.385
Overall Decomposition:					
Supply Factor, $\ln \psi_i$	Y	95,529	0.000	-0.230	1.382
Demand Factor, $\ln \xi_i - \ln \beta_i^c$	Y	95,529	0.000		
Supply Decomposition:					
Production Capability, $\ln \tilde{z}_i$	Y	95,529	0.000	-0.259	1.681
Network Supply, $\ln c_i$	Y	95,529	0.000	-0.042	1.033
Domestic Network Input Share, $\ln \beta_i^s$		95,529	0.000	-0.055	0.246
# Suppliers, $\ln n_i^s$		95,529	0.000	-0.009	0.291
Avg Supplier Capability, $\ln \bar{\psi}_i$	Y	95,529	0.000	0.009	0.101
Supplier Covariance, $\ln \Omega_i^s$	Y	95,529	0.000	0.034	0.244
Demand Decomposition:					
1 - Final Demand Sales Share, $\ln \beta_i^c$		95,529	0.000	-0.372	1.612
Network Demand, $\ln \xi_i$	Y	95,529	0.000	0.050	1.944
# Customers, $\ln n_i^c$		95,529	0.000	-0.027	1.487
Avg Customer Capability, $\ln \bar{\theta}_i$	Y	95,529	0.000		
Customer Covariance, $\ln \Omega_i^c$	Y	95,529	0.000	0.165	1.487

Table 7: Overall firm size decomposition (2014).

	Supply Factor, $\ln \psi_i$		Demand Factor, $\ln \xi_i - \ln \beta_i^c$	
$\ln S_i$	0.144***		0.856***	
	Production Capability $\ln \tilde{z}_i$	Network Supply $\ln c_i$	Final Demand $\ln \beta_i^c$	Network Demand $\ln \xi_i$
$\ln S_i$	0.139***	0.006***	0.003	0.853***
	(96%)	(4%)	(0.4%)	(99.6%)

Note: Rows 1 and 2 report coefficient estimates from OLS regressions of a firm size margin (as indicated in the column heading) on total firm sales, across 95,529 firms in 2014. Row 3 reports the percent share of a supply (demand) sub-component in the total supply (demand) factor. All standard errors are below 0.004. Significance: * < 5%, ** < 1%, *** <0.1%.

variance decomposition therefore estimates the importance of each margin in explaining within-industry size heterogeneity.

5.1 Firm Size in the Cross-Section: 2014 Baseline

5.1.1 Top-tier decomposition

We first examine the origins of firm size heterogeneity in the cross-section for 2014, the most recent year in the data. We start with the top-tier decomposition of firm sales $\ln S_i$ into the supply factor $\ln \psi_i$ and the demand factor $\ln \xi_i - \ln \beta_i^c$, by regressing each factor on $\ln S_i$. Recall that by the properties of OLS, the coefficient estimates from these two regressions sum to 1 by construction, and indicate what fraction of the total variation in firm sales can be attributed to each factor. We report the results in the top row of Table 7. The demand side accounts for fully 85.6% of the size dispersion across firms, while supply fundamentals explain only 14.4%.

What is the interpretation of these results? The supply factor $\ln \psi_i$ represents, loosely speaking, the average market share of i among its customers. Hence, the relatively small role for supply fundamentals means that average market share is not strongly correlated with total firm sales. In other words, being an important supplier to your customers is only weakly related to overall firm success. This does not mean, however, that supply-side factors in general are unimportant in explaining firm size. Rather, the results suggest that supply-side factors that are othogonal to average market share might be important. Examples of such factors are efficiency in marketing or skills in finding and attracting a customer base.

We next unpack this top-tier decomposition by distinguishing between the network-

dependent and the firm-specific, network-independent components of the supply and demand factors. In the second row of Table 7, we present the coefficient estimates from OLS regressions of each of these four margins on total firm sales. The third row expresses the percent contribution of each supply (demand) sub-component to the workings of the overall supply (demand) factor in the first row. On the supply side, the seller-specific production capability $\ln \tilde{z}_i$ drives almost the entire supply factor (96%) and is responsible for 13.9% of the overall variation in firm size. On the other hand, the role of network supply is extremely limited: The cross-firm variation in marginal production costs $\ln c_i$ due to differential sourcing of inputs from the network of suppliers adds a trivial 0.6% to the variation in firm sales and only 4% to the top-tier supply margin.

The patterns look dramatically different on the demand side. Differences in final demand across firms, as captured by the fraction of firm sales to final consumers $(1 - \ln \beta_i^c)$, account for an economically negligible and statistically insignificant 0.3% of the overall variation in firm size. Practically the entire (99.6%) demand factor is governed by demand from other firms in the production network $\ln \xi_i$. Network demand thus generates a whopping 85.3% of the overall firm size dispersion.

These findings suggest that the key to understanding the vast firm size heterogeneity observed in modern economies is in how firms manage their sales activities, and specifically how they match and transact with buyers in the production network. This does not imply that the production side is irrelevant: models of the production process within firms inform various important aspects of firm operations beyond firm sales, such as value added and profits. In addition, the evidence lends support to the large class of models that focus on a single firm attribute on the production side (e.g. productivity), such as Melitz (2003).

Finally, this top-tier firm size decomposition speaks to the stylized facts we presented in Section 2. At a basic level, the evidence here suggests that there is an intimate relationship between the skewed distributions of firms' total sales and various aspects of their production network activity, as summarized in *Fact 1*. In turn, the important role we uncover for the network supply and network demand factors of the firm size suggests that the other *Facts 2-5* also implicitly reflect how the production network shapes the firm size distribution.

5.1.2 Network supply decomposition

We next decompose the network supply and network demand margins into their constituent parts to assess the specific channels through which the production network shapes firm sales. Table 8 reports the results from regressing each component of network supply against the full network supply factor $\ln c_i$. As before, the coefficient estimates across these components sum to 1 by construction, while their absolute and relative values are not mechanically

Table 8: Network supply decomposition (2014).

	Domestic Network Input Share $\ln \beta_i^s$	# Suppliers $\ln n_i^s$	Avg Supplier Capability $\ln \bar{\psi}_i$	Supplier Covarianc $\ln \Omega_i^s$
Network Supply, $\ln c_i$	0.090***	0.507***	0.046***	0.360***

Note: Coefficient estimates from OLS regressions of a firm size margin on the supply side (as indicated in the column heading) on the network supply component of firm sales (i.e. input cost), across 95,529 firms in 2014. All standard errors are below 0.002. Significance: * < 5%, ** < 1%, *** <0.1%.

constrained and thus quantify the sources of heterogeneity in $\ln c_i$. Just over half (50.7%) of the variation in marginal production costs across firms can be attributed to the extensive margin, i.e. the number of domestic suppliers $\ln n_i^s$ that producers buy inputs from. Second most important is the supplier covariance term $\ln \Omega_i^s$, which accounts for another 36%. At the same time, the average production capability of a firm's input suppliers $\ln \bar{\psi}_i$ is much less consequential at 4.6%. Imported inputs acquired outside the domestic production network, 1-$\ln \beta_i^s$, account for 9%.

These results shed light on how successful firms are able to reduce their marginal costs by accessing the production network. Since we do not model the endogenous match formation in the network, however, the analysis reveals the ex-post contribution of different channels, rather than the ex-ante firm decisions that trigger them. We can nevertheless draw conclusions about the latter from the former. On the extensive margin, firms benefit primarily by sourcing inputs from more suppliers, rather than by matching with more capable suppliers. On the intensive margin, conditional on matching with a certain number of suppliers, firms' marginal cost decreases more if they happen to be particularly good matches to the more capable among their suppliers.

These patterns would be consistent with the combination of search frictions and asymmetric information in the production network. In particular, producers may have to pay fixed search costs in order to meet input suppliers, while also facing ex-ante uncertainty about the primitive production capability of these suppliers and their buyer-supplier, pairwise match quality (e.g. how well suited the widget produced by a given suppliers is to my own production process). In such an environment, exogenously more capable firms would be able to invest in meeting more suppliers on the extensive margin (but with a similar average supplier capability) than less capable firms. On the intensive margin, more capable firms could also more effectively allocate their input purchases towards suppliers with both higher production capability and match quality.

Table 9: Network demand decomposition (2014).

	# Customers $\ln n_i^c$	Avg Customer Capability $\ln \bar{\theta}_i$	Customer Covariance $\ln \Omega_i^c$
Network Demand, $\ln \xi_i$	0.690***	0.140***	0.174***

Note: Coefficient estimates from OLS regressions of a firm size margin on the demand side (as indicated in the column heading) on the network demand component of firm sales (i.e. sales to other firms), across 95,529 firms in 2014. All standard errors are below 0.003. Significance: * < 5%, ** < 1%, *** <0.1%.

The network supply decomposition elucidates the relationship between the firm size dispersion and the stylized facts in Section 2. The conclusions here are consistent with *Fact 2* that bigger firms have more upstream suppliers. The results for the covariance term echo *Fact 4* that the distribution of a firm's input purchases across suppliers widens with the number of its suppliers, even though its median barely changes. They also illustrate the important role for buyer-supplier match characteristics implied by *Fact 5*, i.e. that one-sided seller and buyer fixed effects jointly account for only 38% of the variation in firm-to-firm sales.

5.1.3 Network demand decomposition

We complete the firm size decomposition by unbundling the network demand margin of firm sales, $\ln \xi_i$. Table 9 reports the results from regressing each component of network demand against $\ln \xi_i$, such that the coefficient estimates quantify the relative importance of each component. An overwhelming 69% of the variation in network sales across firms can be attributed to the extensive margin, i.e. the number of (domestic) buyers $\ln n_i^c$ that producers sell to. On the other hand, the average sourcing capability across a firm's customers $\ln \bar{\theta}_i$ and the customer covariance term $\ln \Omega_i^c$ contribute a much more modest 14% and 17.4%, respectively. It is instructive to compare these three margins to their counterparts above on the production side of network supply, which stand at 50.7%, 4.6% and 36%.

These patterns have several implications for the opportunities that buyer-supplier production networks create for firm sales expansion. As with network supply, the network demand decomposition is an ex-post description, but it implicitly informs the ex-ante firm maximization behavior that generated it. We conclude that on the sales side, the single most important advantage of large firms is that they successfully match with a big number of downstream buyers. They also match with slightly (but not dramatically) more capable

buyers on average, especially when compared to their ability to match with better suppliers. While bigger firms do tend to be especially well-suited suppliers of capable buyers, the covariance term is only half as important to network sales as it is to network input purchases.

Our findings for the network demand and network supply decompositions are thus complementary. They both point to search frictions and asymmetric information in production networks with two-sided heterogeneity in firm characteristics and variation in buyer-supplier match quality. At the same time, the contrasting breakdown of network demand and network supply suggest important differences in the process of meeting and transacting with buyers vs. suppliers. The evidence would be consistent with firms incurring lower costs of searching and identifying good buyers than suppliers on the extensive margin. This is compensated by lower sales gains from optimally allocating downstream sales across buyers with different capability and match quality, compared to the cost savings from optimally allocating input purchases across suppliers.

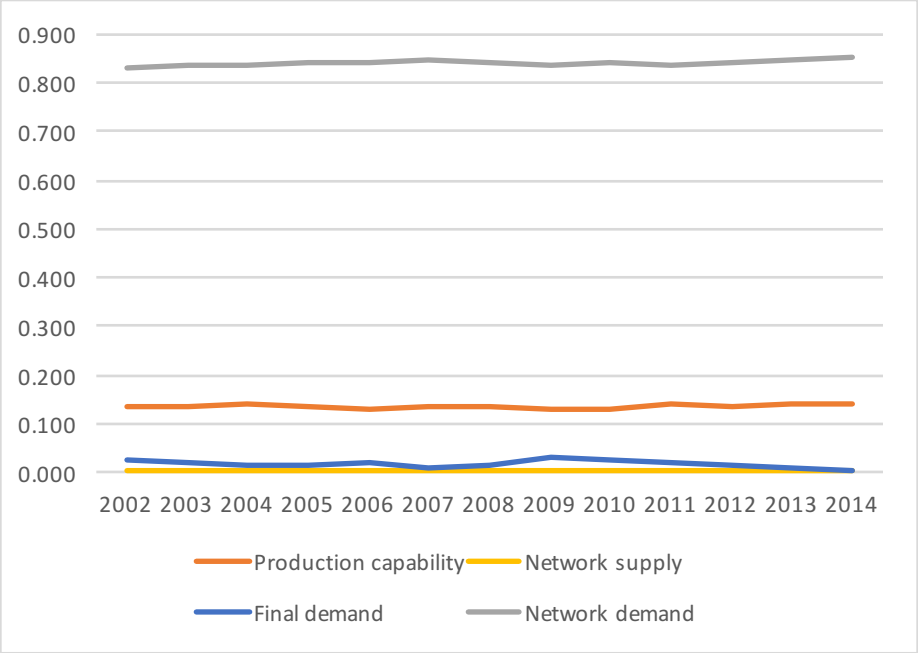
Finally, the network demand decomposition also sheds light on several stylized facts in Section 2. It powerfully illustrates *Fact 2* that bigger firms have more downstream buyers. The limited role of the covariance term reinforces *Fact 4* that the distribution of a firm's sales across customers is generally invariant with the number of its customers.

5.2 Firm Size in the Cross-Section: 2002-2014 Time Trends

In this subsection, we next explore the evolution of the firm size distribution in Belgium over the 2002-2014 period in the sample. We find that despite the increase in the number of firms and in their sales dispersion over time, the sources of firm size heterogeneity have remained remarkably stable.

We perform the three-step firm size analysis separately for each year in the data, and plot the results for the top-tier decomposition in Figure 7. On the supply side, the roles of production capability and network supply have hovered around 13-14% and 0.4-0.7% respectively. On the demand side, the importance of network demand has gradually risen from 83.1% to 85.3%, closely following a decline in that of final demand from 2.7% to 0.3%.

Figure 7: Top-tier firm size decomposition, 2002-2014.

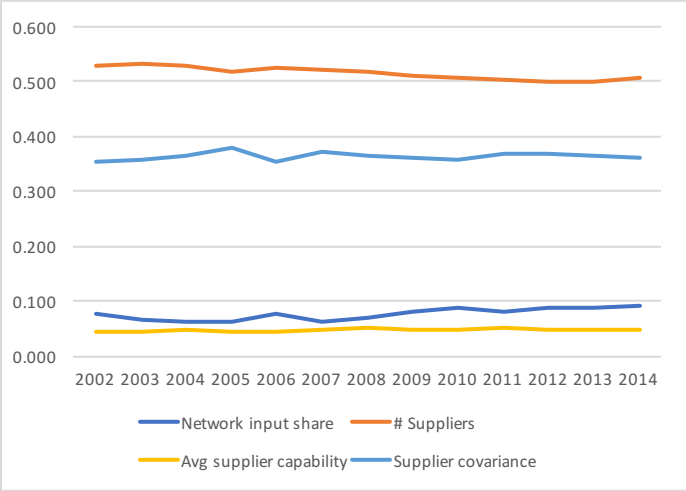


We observe similarly stable patterns when we consider the lower-tier decomposition of the network demand and network supply margins in Figure 8. On the supply side, the contribution of the number of upstream input suppliers has slowly trended down, yet consistently added 50-54% to the variation in marginal costs across firms. The role of average supplier capability has steadily stood at 4.4-5.2%, while that of the supplier covariance term has fluctuated between 35-37%. The share of inputs acquired from within the domestic production network has accounted for 6-9%. On the demand side, 69-70% of the variation in network sales in each year can be attributed to the number of downstream buyers. Average customer capability and the customer covariance term can continuously explain 13-14% and 17-18%, respectively.

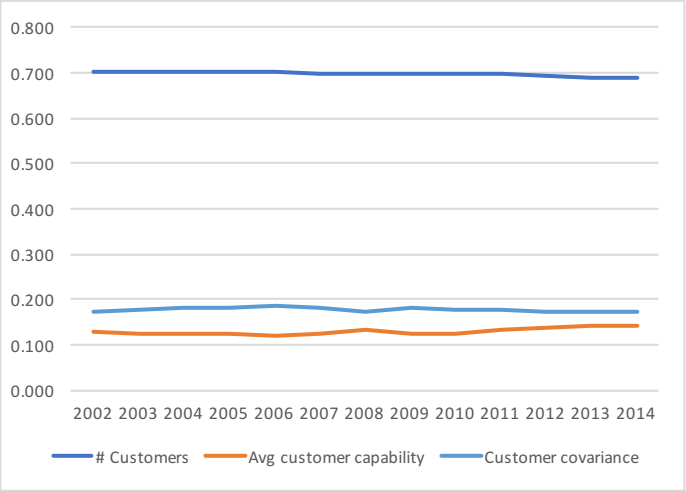
These findings suggest that there may be inherent drivers of the firm size distribution whose relative importance remains stable despite the rise in production fragmentation across firm and country boundaries over the last 15 years.

Figure 8: Network supply and network demand decomposition, 2002-2014.

(a) Network supply decomposition, 2002-2014..



(b) Network demand decomposition, 2002-2014.



5.3 Sensitivity Analysis

We perform several sensitivity exercises to explore the stability of the resuts. In this section, we discuss estimates from alternative approaches to the firm size decomposition in the cross-section for 2014. Similar patterns hold for other years in the 2002-2014 panel, as well as for sales growth within firms over time.

We first consider the pairwise seller-buyer match characteristic $\ln \omega_{ij}$, which we estimate in *Step One*. In the baseline, we specify $\ln \omega_{ij} = \gamma \ln d_{ij} + e_{ij}$, where d_{ij} is observed match frictions proxied by bilateral distance and e_{ij} is unobserved ones. One rational for this choice is that in the model, $\ln \omega_{ij}$ combines bilateral trade costs, demand shocks and mark-ups. To the extent that distance proxies trade costs, the residual e_{ij} isolates the sum of match-specific demand and mark-up, and allows us to compare the relative importance of different drivers of match quality. Regressing e_{ij} on $\ln \omega_{ij}$, we find that ***% of the variation in match quality across buyer-supplier pairs is due to match-specific demand and mark-ups. Another rational for including distance in the first-stage OLS regression is that bilateral trade costs likely affect both the selection into a buyer-supplier match and the value of firm-to-firm sales conditional on such a match. Controlling for d_{ij} thus helps ensure that the identifying assumption of conditional exogenous mobility is satisfied (see Section 4.2). As a robustness check, we have nevertheless conducted the firm size decomposition without including bilateral distance in the first step. Our results are virtually unchanged.[21]

[21]Small firms are less likely to report location and sales data in the BSO data. Dropping distance from the analysis thus increases the estimation sample in the first stage from 32.3% to 98.6% of all links in the

We next consider the parameter assumptions we make in order to back out all firm size components in *Step Two*. Our estimation methodology requires parameter values for the labor share α in firms' Cobb-Douglas production technology and for the wage rate w (see Section 4.3). In the baseline, we allow α and w to vary across industries, and proxy them with measures based on industry-wide employment, wage bill and total production costs. As a robustness exercise, we use firm-specific values for α and w to accommodate the possibility that different firms within the same industry can adopt different production techniques, hire workers with different skills, and/or pay different wages for identical workers. We find that this does not affect the overall contribution of supply- and demand-side factors to firm size heterogeneity. It also does not change the breakdown of the demand factor into final vs. network demand, nor the decomposition of network demand into subcomponents. In all these cases the estimates change less than 0.2 percentage points. The only difference we observe is on the supply side: Whereas own production capability accounts for 13.9% of the overall variation in firm sales in the baseline, its share more than doubles to 29.5% under firm-specific α and w. This is fully compensated by a reduction in the role played by network supply (i.e. in marginal costs that reflect inputs sourced from the suppliers in the network) from 0.6% to -15.1%. Yet the decomposition of the variation in network supply into its subcomponents is practically identical. One potential explanation is that firms with high production capability tend to hire more skilled workers and/or pay higher wages for a given worker, while also sourcing inputs from more suppliers and more effectively allocating input purchases across different suppliers. While the former force would tend to increase marginal costs, the latter would tend to reduce them.

Finally, we explore the possibility that the sources of firm size heterogeneity vary along the firm size distribution in *Step Three*. Put simply, we evaluate whether various supply and demand factors are more or less important to the sales of small vs. large firms. This would be the case, for example, if relatively small firms depend more on sales to final consumers than downstream buyers, or if their marginal costs are less sensitive to the cost of inputs sourced from the production network and driven primarily by own production capability. To this end, we complete steps one and two of the analysis as in the baseline, but perform the firm size decomposion in step three separately for each decile of the firm sales distribution.

We present the results in Figure 9. A striking pattern is that the top-tier decomposition of the variation in firm size into supply- and demand-side factors looks similar across deciles of the size distribution, but the underlying sources of variation in these supply and demand

raw B2B data because we can include small firms without location information and because the avalanching process of removing firms with only one buyer or seller is mitigated. However, because we require sales data in the second and third stage, the number of firms in the firm size decomposition depends little on whether we specify distance in the first stage.

factors do not. On the supply side, own production capability is proportionately more important than network supply for smaller firms. On the demand side, final demand is proportionately more important than network demand for smaller firms. Thus production network activity appears more important to firm size for firms at the top end of the size distribution.

Figure 9: Top-tier firm size decomposition by size decile, 2014.

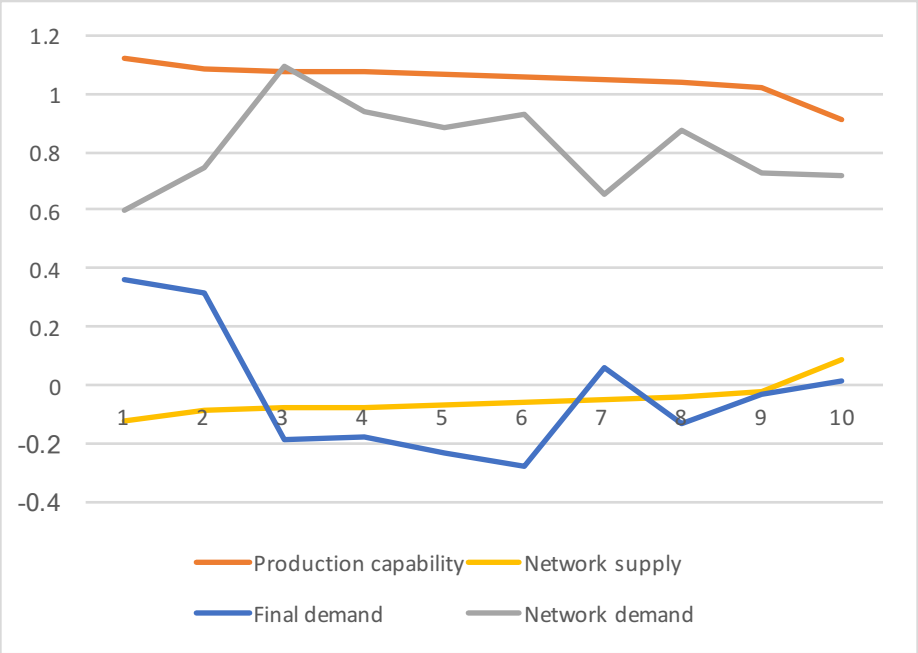
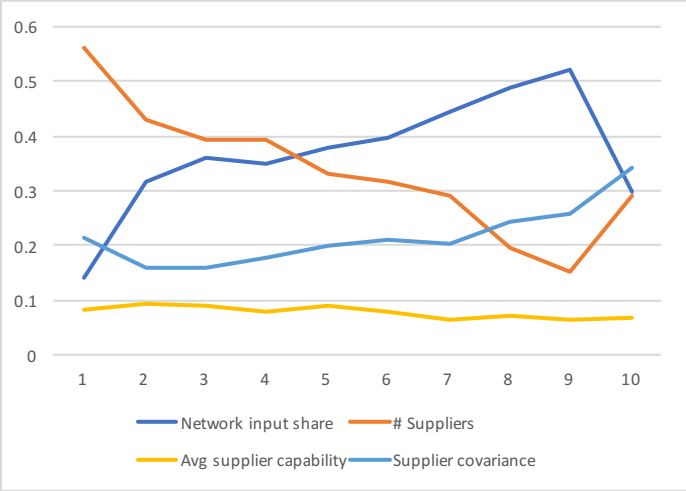


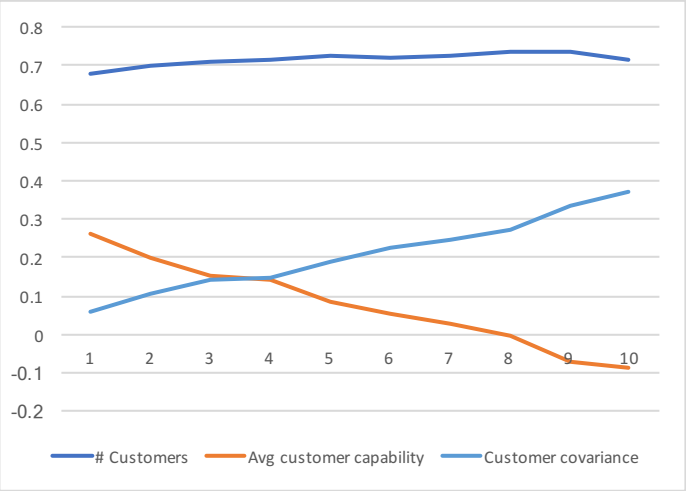
Figure 10 unpacks the sources of variation in network supply and network demand by size decile, and reveals further differences across size deciles. On the network supply side, the heterogeneity in marginal costs among larger firms is largely driven by their share of inputs acquired from other firms in the network and their ability to allocate input purchases towards more capable and better-matched suppliers. By contrast, the number of input suppliers is a much more important source of cost heterogeneity among smaller firms. On the network demand side, average customer capability plays a bigger role for firm size at lower size deciles, while the allocation of sales across downstream buyers is much more important for firm size at top size deciles. Overall, these findings suggest that the production network offers different challenges and opportunities for small and large firms in terms of how they search, match and transact with buyers and suppliers.

Figure 10: Network supply and network demand decomposition by size decile, 2014.

(a) Network supply decomposition by size decile, 2014.



(b) Network demand decomposition by size decile, 2014.



6 General Equilibrium

The estimation and decomposition presented in Sections 3 and 4 provide parameter values for firm-level fundamentals. What remains is to close the model and solve for the general equilibrium. This will become useful in the counterfactual experiments presented below.

Final Demand. To close the model, two additional assumptions are required. First, we need an assumption about final demand. We choose the simplest possible case and assume CES utility with the same elasticity of substitution σ across firms:

$$U = \left(\sum_k \left(\phi_k \nu_k \right)^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)}.$$

Using the same functional form for final demand and firm demand enables us utilize the estimates for production also for final demand. We consider the final consumer as an average input consumer, so that the terms $\tilde{\phi}_{ki}$ and $\tilde{\tau}_{ij}$ do not appear in final demand.[22] The value of final demand is $\mathcal{F}_i = \left(\frac{\phi_i}{\tau_i p_i} \right)^{\sigma-1} \mathcal{P}^{\sigma-1} X$, where X is overall income. $\mathcal{P}$ is the CES price

[22]Since final demand is modeled as a representative consumer, there is by construction no match specific component ϕ_{ki}. Since $\phi_{ki} = \phi_k \tilde{\phi}_{ki}$ and $(1/n_i^c) \sum_i \tilde{\phi}_{ki} = 1$, this implies that the perceived quality of a firm i is identical for the final consumer and the average firm j.

index:

$$\begin{aligned}\mathcal{P}^{1-\sigma} &= \sum_i \left(\tau_i p_i/\phi_i\right)^{1-\sigma} \\ &= \sum_i \tilde{P}_i^{1-\alpha}\tilde{z}_i, \end{aligned} \tag{15}$$

where $\tilde{P}_j \equiv P_j^{1-\sigma}$.

Mark-ups. Second, we need an assumption about mark-ups. So far, we have been completely agnostic about market structure and price determination. To allow for maximum flexibility, we assume that the mark-up potentially varies across firms, but that it is constant across equilibria.[23] As a consequence, a firm's purchases relative to total sales, $\mu_i \equiv M_i/S_i$ is constant.[24] We therefore use data on μ_i when simulating the model. The set of firms is fixed and there is no free entry. We assume that the final consumer is the shareholder of the firms, so that aggregate profits Π become part of consumer income. Income X is forefore the sum of labor income and aggregate profits, $X = wL + \Pi$, where w is the wage and L is inelastically supplied labor. It can be shown that in equilibrium, $X = \rho w L$, where ρ is a constant term.

Backward fixed point. We need to determine how input costs of firm j relates to the input costs of the suppliers of j. This can be solved by iterating on a backward fixed point problem. The backward fixed point relates the price index of firm j to the price index of its suppliers, using equations (1) and (8):

$$\begin{aligned}\tilde{P}_j &= \sum_{i\in\mathcal{S}_i}\left(\frac{p_{ij}}{\phi_{ij}}\right)^{1-\sigma} \\ &= \sum_{i\in\mathcal{S}_i}\tilde{P}_i^{1-\alpha}\tilde{z}_i\omega_{ij}. \end{aligned} \tag{16}$$

Firm j's input costs depend on the production capability of its suppliers, $\tilde{z}_i$, the suppliers' input costs, $\tilde{P}_i$, as well as the match term ω_{ij}. We solve the fixed point using data on α and our estimates of $\tilde{z}_i$ and ω_{ij} from above.

Forward fixed point. We also need to characterize how sales of firm i relates to the sales of the customers of i. Total firm sales are $S_i = \mathcal{F}_i + \sum_{j\in\mathcal{C}_i} m_{ij}$. The forward fixed point is then, using equations (1), (8) and (2) and defining $\tilde{\mathcal{P}} \equiv \mathcal{P}^{1-\sigma}$:

$$S_i = \tilde{z}_i\tilde{P}_i^{1-\alpha}\left(\frac{X}{\tilde{\mathcal{P}}} + \sum_{j\in\mathcal{C}_i}\frac{\mu_j S_j}{\tilde{P}_j}\omega_{ij}\right). \tag{17}$$

[23] An alterantive, and in our opinion less flexible, approach would be to add more assumptions about market structure and pricing behavior.

[24] $\frac{M_i}{S_i} = \frac{1-\alpha}{S_i/TotalCosts_i} = \frac{1-\alpha}{Markup_i}$, which is constant given that $Markup_i$ is constant.

A detailed derivation is found in Appendix A.2. Firm i's sales depend on final demand, $X/\tilde{\mathcal{P}}$, the production and sourcing capability of the firm itself, $\tilde{z}_i$ and $\tilde{P}_i$, as well as the sales, sourcing capabilities and match effects among its customers, S_j, $\tilde{P}_j$ and ω_{ij}. We solve the fixed point using (i) data on α, final demand X and μ_i, (ii) our estimates of $\tilde{z}_i$ and ω_{ij}, and (iii) $\tilde{P}_i$ and $\tilde{\mathcal{P}}$ using the solution to the backwards fixed point in equation (16). Note that we can solve for the equilibrium distribution of sales without imposing any assumption about the elasticity of substitution σ.

Welfare. Indirect utility equals the inverse of the final demand price index $\mathcal{P}$. Hence, welfare can we evaluated by equation (15), using estimates of production capability $\tilde{z}_i$ and match effects ω_{ij} as well as the solution to $\tilde{P}_i$ from the backward fixed point.

7 Conclusions

This paper quantifies the origins of firm size heterogeneity when firms are interconnected in a production network. We first document new stylized facts about a complete production network using data on the universe of buyer-supplier relationships among all firms in Belgium during 2002-2014. These stylized facts suggest that the network of buyer-supplier links is key to understanding the firm size distribution. Specifically, they signal the important roles played by downstream input demand as distinct from final demand, by both seller- and buyer-specific firm characteristics, and by seller-buyer match characteristics.

Morivated by these facts, we outline a model in which firms buy inputs from upstream suppliers and sell to downstream buyers and final demand. In the model, firms can be large for the standard reason that they have high production capability (i.e. productivity or product quality). However, firms can also be large because they interact with more, better and larger buyers and suppliers and because they are better matched to their buyers and suppliers. This framework delivers an exact decomposition of firm size into supply and demand margins with firm, buyer/supplier and match components. We design a three-stage estimation methodology that makes it possible to back out these firm size components from data on firm-level balance sheets and firm-to-firm transactions in a production network. We implement the methodology using detailed data for Belgium, and quantify the contribuion of each component to the overall dispersion in firm size in the economy.

We establish three empirical results for the origins of firm size heterogeneity. These patterns hold in the cross-section of firms in each year of the panel, as well as in the evolution of firm size within firms over time. First, demand factors explain 81% of firm size heterogeneity, while supply factors only 19%. Second, nearly all the variation on the demand side (99%) is driven by sales to other firms rather than by final demand. By contrast, 88% of the variation

on the supply side reflects heterogeneity in own production capability and 12% comes from input suppliers. Third, the extensive margin of the production network dominates the intensive margin. Most of the variance in the network demand and supply components of firm size is determined by the number of customers and suppliers, rather than their size or capability. Second most important is that firms transact more with partners that are simultaneously more capable and better bilateral matches.

These theoretical, methodological and empirical contributions open interesting avenues for future research. We have taken the production network as given in order to assess its role in shaping the firm size distribution. Our results nevertheless shed light on the various challenges and opportunities that firms face in the presence of input-output linkages in the economy. Future work can examine how firm-specific characteristics determine the matching of buyers and suppliers in the production network in light of our findings. Separately, we have dissected the origins of firm size heterogeneity, but not explored its implications for the aggregate economy. Future studies can analyze whether different sources of the dispersion in firm size have different implications for aggregate outcomes such as growth or income inequality. Finally, we have focused on the relationship between the production network and firm size heterogeneity in steady state. Future analyses can explore how this relationship affects the propagation process and aggregate welfare impact of both firm-specific and macroeconomic shocks.

References

Bernard, A., Moxnes, A., and Saito, Y. (2015a). Production networks, geography and firm performance. *NBER Working Paper*, 21082. 1, 2.1, 10, 2.2

Bernard, A., Moxnes, A., and Ulltveit-Moe, K. (2015b). Two-sided heterogeneity and trade,. *NBER Working Paper*, 20136. 1, 10, 2.2

Dhyne, E., Magerman, G., and Rubinova, S. (2015). The belgian production network 2002-2012. *National Bank of Belgium Working Paper*, 288. 2.2

Hottman, C. J., Redding, S., and Weinstein, D. E. (2016). Quantifying the sources of firm heterogeneity. *Quarterly Journal of Economics*, 131(3):1291–1364. 1, 3.4

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature*, 49(2):326–365. 1

Appendix

A The Model

A.1 The Supply Side Decomposition

From equation (7), we get

$$\ln \psi_i = \ln \tilde{z}_i + (1-\alpha)\left(\ln M_i - \ln \theta_i\right),$$

where $\tilde{z}_i \equiv \left(\phi_i z_i / \left(\kappa \tau_i w^{\alpha}\right)\right)^{\sigma-1}$. Replacing $\ln M_i - \ln \theta_i$ from equation (11) yields

$$\ln \psi_i = \ln \tilde{z}_i + (1-\alpha)\left(\ln \sum_{k \in \mathcal{S}_i} \psi_k \omega_{ki} - \ln \eta_i\right).$$

The term $\sum_{k \in \mathcal{S}_i} \psi_k \omega_{ki}$ can be further decomposed into

$$\ln \sum_{k \in \mathcal{S}_i} \psi_k \omega_{ki} = \ln n_i^s + \ln \bar{\psi}_i + \ln \frac{1}{n_i^s} \sum_{k \in \mathcal{S}_i} \frac{\psi_k}{\bar{\psi}_i} \omega_{ki},$$

where $\bar{\psi}_i = \left(\prod_{k \in \mathcal{S}_i} \psi_k\right)^{1/n_i^s}$. Combining the last two equations yields equation (12) in the main text.

A.2 Forward Fixed Point

Total firm sales are $S_i = \mathcal{F}_i + \sum_{j \in \mathcal{C}_i} m_{ij}$. We first derive expressions for final demand and then demand from other firms.

Final demand. Using equation (1) and defining $\tilde{P}_i \equiv P_i^{1-\sigma}$ and $\tilde{\mathcal{P}} \equiv \mathcal{P}^{1-\sigma}$, the final demand price index is

$$\begin{aligned} \tilde{\mathcal{P}} &= \sum_i \left(\frac{\tau_i p_i}{\phi_i}\right)^{1-\sigma} \\ &= \sum_i \tilde{P}_i^{1-\alpha} \tilde{z}_i. \end{aligned}$$

Using equation (1), final demand is

$$\begin{aligned} \mathcal{F}_i &= \left(\frac{\phi_i}{\tau_i p_i}\right)^{\sigma-1} \mathcal{P}^{\sigma-1} w L \\ &= \tilde{z}_i \tilde{P}_i^{1-\alpha} \frac{wL}{\tilde{\mathcal{P}}}. \end{aligned}$$

Firm Demand. Using (1), (8) and (2), firm demand is

$$\sum_{j\in\mathcal{C}_i} m_{ij} = \sum_{j\in\mathcal{C}_i} \left(\frac{\phi_{ij}}{p_{ij}}\right)^{\sigma-1} P_j^{\sigma-1} M_j$$
$$= \tilde{z}_i \tilde{P}_i^{1-\alpha} \sum_{j\in\mathcal{C}_i} \frac{\mu_j S_j}{\tilde{P}_j} \omega_{ij}.$$

Combining the two sources of demand, we get total sales:

$$S_i = \tilde{z}_i \tilde{P}_i^{1-\alpha} \left(\frac{wL}{\tilde{\mathcal{P}}} + \sum_{j\in\mathcal{C}_i} \frac{\mu_j S_j}{\tilde{P}_j} \omega_{ij} \right).$$

B Data sources and construction

B.1 The Belgian VAT system and VAT filings

The Belgian value-added tax (VAT) system requires that the vast majority of firms located in Belgium and across all economic activities charge VAT on top of the delivery of their goods and services. This also includes foreign companies with a branch in Belgium or those firms whose securities are officially listed in Belgium. Firms that only perform financial transactions, medical or socio-cultural activities are exempt. The tax is levied in successive stages of the production and distribution process: with each transaction, firms charge VAT on top of their sales and pay VAT on inputs sourced from their suppliers, in effect only paying taxes on the value-added generated at that stage. The tax is neutral to the firm and the full burden of the tax ultimately lies with the final consumer. The standard VAT rate in Belgium is 21%, but for some goods a reduced rate of 12% or 6% applies.[25]

VAT-liable firms have to file periodic VAT declarations and VAT listings to the tax administration.[26] The VAT declaration contains the total sales value, the VAT amount charged on those sales (both to other firms and to final consumers), the total amounts paid on inputs sourced and the VAT paid on those inputs. This declaration is due monthly or quarterly, depending on some size criteria, and it is the basis for the balance of VAT due to the tax authorities every period. Additionally, at the end of every calendar year, all VAT-liable firms have to file a complete listing of their Belgian VAT-liable customers over that year. An observation in this listing refers to the yearly values of a sales relationship from firm i to firm j. The reported value is the sum of the values of all invoices from i to j.

[25] See ec.europa.eu/taxation_customs for a complete list of rates. These rates did not change over our sample period.

[26] Sample forms of the VAT declarations can be found here (French) and here (Dutch). Sample forms of the VAT listings can be found here (French) and here (Dutch).

Whenever this aggregate value is larger than or equal to 250 euro, the relationship has to be reported. Sanctions for incomplete and mis-reporting guarantee a high quality of the data.

B.2 Data sources

The empirical analysis draws from three data sources administered by the National Bank of Belgium (NBB): (i) the NBB B2B Transactions Dataset, (ii) annual accounts from the Central Balance Sheet Office at the NBB, (iii) and the Crossroads Bank at the NBB. We additionally use geographic information on postal codes in Belgium to calculate distances. Firms are identified by their VAT number, which is unique and common across these databases and allows for straightforward merging across datasets.

Firm-to-firm relationships The confidential NBB B2B Transactions Dataset contains the values of yearly relationships between all VAT-liable Belgian firms for the years 2002 to 2014, and is based on the VAT listings collected by the tax authorities. An observation in this dataset refers to the sales value in euro of firm i selling to firm j within Belgium, excluding the VAT amount due on these sales. This value is the sum of invoices from i to j in a given calendar year. Note that the relationship is directed, as the observation from i to j is different from the observation from j to i. A detailed description of the collection and cleaning of this dataset is given in Dhyne et al. (2015).

Firm-level characteristics We extract firm-level information from the annual accounts from the Central Balance Sheet Office at the NBB for the years 2002 to 2014. Firms above a certain threshold have to file annual accounts at the end of their fiscal year.[27] We retain information on the firm identifier, turnover (total sales in euro, code 70 in the annual accounts), inputs (total material and services inputs in euro, codes 60+61), labor cost (total cost of wages, social securities and pensions in euro, code 62 in annual accounts) and employment (average number of FTE in given fiscal year, code 9087). Small firms with abbreviated accounts do not have to report turnover and intermediate inputs in the annual accounts. We annualize all flow variables from the annual accounts from fiscal years to calendar years. This transformation ensures that all firm-level information in our database is consistent with observations in the VAT listings data.[28]

We obtain information on the main economic activity of the firm at the NACE 4-digit level from the Crossroads Bank of Belgium for the years 2002 to 2014. We concord NACE

[27] See here for filing requirements and exceptions. See here for the size criteria and filing requirements for either full-format or abridged annual accounts.

[28] 78% of firms have annual accounts that coincide with calendar years, while 98% of firms have fiscal years of 12 months.

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. We also extract the main location of the firm at the postal code level from the Crossroads Bank to calculate geographic distance between pairs of firms in our sample.

B.3 Data construction and cleaning

We calculate final demand of firm i in year t as the residual of turnover minus the sum of outgoing firm-to-firm relationships of i. This implies that final demand contains final consumption, exports and business transactions not observed in our B2B dataset. Similarly, we calculate the observed share of inputs as total inputs minus the sum of incoming firm-to-firm relationships of i. The residual unobserved part of inputs then contains imports and unobserved B2B inputs to i.

Wages are calculated as labor cost over FTE. Labor shares are calculated as labor cost over turnover. We set the labor share equal to one if it is larger than one. For most of the analysis, we use wages and labor shares at the 2-digit NACE industry, by first summing over all firms' labor costs within that industry and then dividing by total FTE in that industry. Similarly for labor shares. We drop firms that have missing employment information or less than one FTE.

Geographical distance between firm pairs is calculated as the km distance between the two postal codes. For within-postal code trade, we follow Head and Mayer (2000) and calculate internal distance as $0.4 \times \sqrt{area}$, where $area$ is the surface area of the postal code in km^2.

Firm pairs are indexed by the Cantor pairing function to keep the pairing identity consistent over the panel.[29]

Throughout the paper, we report information and statistics on both the full collected sample or the estimation sample used in the decomposition. For the full sample, we keep all relationships in the dataset, even if there is missing firm-level information, as these contribute to the decomposition exercise. We only keep firms that show up in the network as either a buyer or a seller. For the estimation sample, we start from the full sample and estimate the two-way fixed effects regression. Note that if a buyer or seller only has one business relationship, the fixed effect is not identified. This firm, together with its connections, is then dropped from the sample. This is done iteratively, until only firms remain that have at least two sellers or buyers. Finally, for the decomposition exercise to contain the same number of observations per (sub-)component, we keep only firms that show up as both buyer and seller in the network.

[29] In particular: $p_{ij} = \frac{1}{2}(a+b) \times (a+b+1) + b$, where p_{ij} is the pair ID and a and b are the seller and buyer ID respectively.