# The Origins of Firm Heterogeneity: A Production Network Approach*

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

**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 upstream and downstream margins with firm, buyer/supplier and match components. We establish three empirical results. First, downstream factors explain the vast majority of firm size heterogeneity, while upstream factors are only one fifth as important. Second, nearly all the variation on the downstream side is driven by sales to other firms rather than final demand. By contrast, most of the variation on the upstream side reflects heterogeneity in own production capability and a much smaller share comes from input suppliers. Third, the extensive margin of the production network dominates the intensive margin. Most of the variance in the downstream and upstream components of firm size is determined by the number of customers and suppliers, rather than their size or capability.

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 in the same industry.[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. Alternatively, 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. 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 customers of the customers

[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.

(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 downstream and upstream 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 three 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, 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. Together, these patterns suggest that the network of buyer-supplier links is key to understanding the firm size distribution.

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 and inputs from upstream suppliers. Firms sell their output to final consumers, 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: upstream and downstream. The upstream 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 downstream margin can be further decomposed into final demand and sales in the production network, 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 on the production side; sales to final 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 upstream- and downstream 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 (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.

[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).

We report results for 2014 but 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 downstream side and 19% from the upstream side. Second, almost the entire variation on the downstream side (99%) is generated by purchases from downstream firms in the production network rather than by final demand. On the upstream 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 sales to downstream buyers, while the number of suppliers explains 49% of the variation in the 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 upstream side, firms benefit more when their production needs are especially well served by highly capable suppliers. On the downstream 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 upstream and downstream 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 upstream and downstream 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). 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 we confirm some of the findings in these papers about the distributions of buyers and suppliers , 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 Belgian production network data, Magerman et al (2016) and Dhyne et al (2017) 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 The Belgian Production Network

## 2.1 Data

We exploit several comprehensive data sources on annual firm operations in Belgium over the 2002-2014 period. We match these data based on unique firm-level value-added tax identification numbers (VAT ID) that arew 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 Transactions Dataset, administered by the National Bank of Belgium (NBB).[5] This dataset reports the sales relationships between any two VAT-liable enterprises across all economic activities in Belgium.[6] In particular, an observation is the sum $m_{ij}$ of sales invoices (in euro and excluding any value-added tax due) from enterprise $i$ to enterprise $j$ in a given calendar year. Coverage is almost universal, as all annual sales worth at least 250 euros must be reported, and administrative sanctions on late and erroneous reporting ensure high data quality.[7] The NBB B2B data thus documents both

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

[6]We use "enterprise" and "firm" interchangeably in this paper. The unit of observation is the VAT ID number, which corresponds to the legal entity of the firm. In other words, we take the VAT ID as the identification of the firm, and do not consider plants, establishments, or groups of firms that might be (in)directly owned through financial participations. The VAT ID is common and unique for all firms across all the datasets used, providing an unambiguous merge across all datasets.

[7]While it is impossible to compare aggregated micro data to the national accounts of economic activity

the extensive and the intensive margins of domestic buyer-supplier relationships in Belgium.[8]

We obtain information on firm-level characteristics from several other datasets. We use data on total sales (turnover), total input purchases, employment and labor costs from firm annual accounts maintained by the Central Balance Sheet Office (CBSO) at the NBB.[9] Annual accounts are collected by fiscal year and have been annualized to match the calendar year in the NBB B2B data. Since there is a firm-size threshold for reporting turnover and input purchases to CBSO, we access data on these two variables for small firms below the threshold from a separate source, namely firms' VAT declarations. We keep only firms with at least one full-time equivalent employee. We observe the main economic activity of each firm at the NACE 2-digit level (harmonized over time to the NACE Rev. 2 (2008) version) and its geographic location at the zip-code level from the Crossroads Bank of Enterprises at the NBB.

We combine these three data sources to construct several variables necessary for the firm size decomposition. We construct firms' sales to final demand as the difference between their turnover and the sum of all their B2B sales to other enterprises in the domestic production network. Final demand thus contains sales to final consumers at home, potentially unobserved links in B2B with small transaction values, and exports. We likewise measure firms' purchases from outside the observed production network (including imports) as the difference between their total input costs and the sum of all their B2B purchases.

Finally, we compute the labor share in production at the NACE 2-digit level as the sum of total employment expenses across all firms in a sector, divided by total turnover in that sector.[10] Similarly, we calculate average wages by sector as the sum of total labor costs divided by total employment in a sector. We use information on firms' zip codes to calculate the bilateral distance between any two enterprises in Belgium.[11]

due to the different data construction methodologies used by these data sources, the two aggregates are very close to each other and have similar growth rates (see Dhyne et al. (2015) for details).

[8]While we do not observe the specific product content of each transaction, our analysis does not require such information. Our theoretical and empirical approach builds on the premise that firms assemble multiple inputs potentially sourced from multiple suppliers into a single product that they sell to other firms and to final consumers.

[9]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.

[10]Our assumption on the Cobb-Douglas upper tier of the production function implies that these shares are also the elasticity of output with respect to employment at the firm level.

[11]We provide further details on data coverage and cleaning in Appendix B.

## 2.2 Stylized Facts

We document three stylized facts about firm size and firm linkages in the Belgian domestic production network.[12] These facts provide evidence that buyer-supplier relationships are key to understanding the firm size dispersion in an economy, and motivate the subsequent theoretical and empirical analysis. We present cross-sectional evidence for the most recent year in our sample, 2014, but the patterns we establish are stable over the 2002-2014 period.

**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 2014, both overall and within six broad industries (primary and extraction, manufacturing, utilities, construction, market services, and non-market services).[13] Across the 109,908 firms with both balance-sheet and production network data in the matched CBSO-B2B data, average firm turnover was €6.7 million with a standard deviation of €145 million. The cross-sectional distribution is, however, extremely skewed. Firms at the 90th percentile generate turnover over 33 times higher than firms at the 10th percentile. Overall, the top 10% of firms account for fully 84% of aggregate sales. The kernel density graphs in Figure 1 illustrate the full distribution of firm sales in the raw data, as well as demeaned across firms within NACE 2-digit sectors.

Similar patterns hold within each broad industry category, although there is substantial heterogeneity across industries. The biggest number of firms is active in market services, while there are few firms in utilities. At the same time, firms in utilities are on average much larger than those in market services or other industries.

Turning to firm-to-firm connections in the domestic production network, we find that the number of downstream customers per seller (out-degree) and the number of upstream suppliers per buyer (in-degree) are also very skewed. In 2014, we observe 17.3 million sales relationships among 859,733 firms within Belgium.[14] Of these, 590,271 enterprises sell to other firms in the network, while 840,607 buy from other firms in the network. Hence 31.5% of firms sell only to final demand, while a small minority of 2.2% do not purchase inputs from the domestic production network (or do so in an amount less than €250). Conditional

[12] 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)).

[13] See Table 12 in Appendix B for the classification of industry groups at the 2-digit NACE level.

[14] The number of firms we observe in the B2B production network is much larger than the number of firms in the matched B2B-CBSO sample above with turnover data, because B2B contains many small firms that do not have to submit complete balance sheets to CBSO.

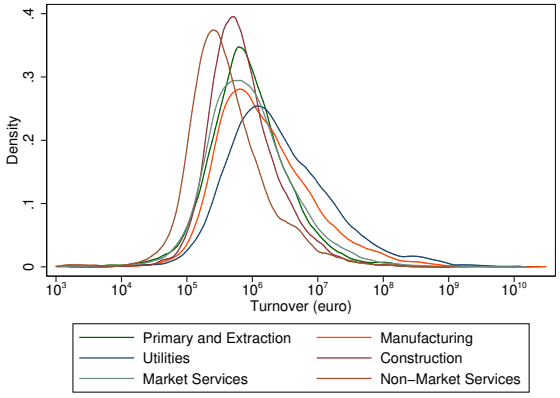
Table 1: Firm sales (€ million, 2014).

| Industry | N | Mean | St Dev | 10th | 25th | 50th | 75th | 90th | 95th | 99th |
|---|---|---|---|---|---|---|---|---|---|---|
| Primary and Extraction (NACE 01-09) | 3,063 | 11.9 | 433 | 0.2 | 0.4 | 0.7 | 1.9 | 4.8 | 9.5 | 52 |
| Manufacturing (NACE 10-33) | 18,086 | 14.4 | 251 | 0.2 | 0.5 | 1.1 | 3.8 | 13.8 | 34.6 | 202 |
| Utilities (NACE 35-39) | 897 | 39.2 | 443 | 0.3 | 0.7 | 1.9 | 6.9 | 25.7 | 68.6 | 496 |
| Construction (NACE 41-43) | 20,206 | 2.3 | 13.4 | 0.2 | 0.3 | 0.6 | 1.4 | 3.6 | 6.9 | 25.9 |
| Market Services (NACE 45-82) | 65,323 | 5.5 | 79.8 | 0.2 | 0.3 | 0.8 | 2.1 | 6.3 | 13.3 | 63.8 |
| Non-Market Services (NACE 84-99) | 2,333 | 2.2 | 26.2 | 0.1 | 0.2 | 0.3 | 0.8 | 2.6 | 5.5 | 24.9 |
| All | 109,908 | 6.7 | 145 | 0.2 | 0.3 | 0.8 | 2.1 | 6.6 | 14.3 | 78.3 |

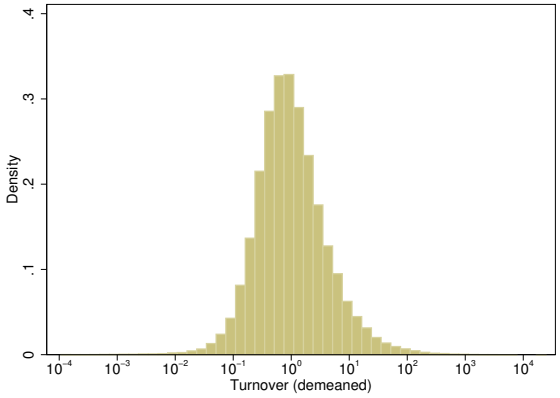
*Notes:* Summary statistics for the matched CBSO-B2B data. 10th, 25th, etc. refers to values at the 10th, 25th, etc. percentile of the distribution.

Figure 1: Firm sales distribution (2014).

(a) Across industry groups.




(b) Within NACE 2-digit industries.




on trading with others in the network, 74% of producers have more than one supplier and 88% have more than one buyer.

Table 2 summarizes the distribution of buyer and supplier connections in the full data, as well as by broad industry. Across all sellers, the average number of customers is 29.3, with a standard deviation of 394. Across all buyers, the average number of suppliers is 20.6, with a standard deviation of 49.5. The average firm thus has more buyers than suppliers, and the distribution of buyers per seller is more dispersed than that of suppliers per buyer. Firm-to-firm links in the network are also highly concentrated among a few very connected participants: 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. The dispersion and skewness across firms within NACE 2-digit industries is also evident in the histograms in Figure 2. 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.

Of note, the in-degree and out-degree distributions have similar features within different broad industries, but they also display some heterogeneity in line with priors. For example, the number of buyers and suppliers is highest for firms in utilities, which are followed closely by manufacturing firms. These numbers are intermediate for producers in primary materials and extraction, and lowest among service providers.

The intensive margin of firm-to-firm bilateral sales is also very dispersed and skewed, with the vast share of economic activity concentrated in a small number of buyer-supplier transactions, as demonstrated in Table 3. The mean transaction across the 17,304,408 buyer-supplier links in 2014 amounts to €28,893. At the same time, the median purchase totals

Table 2: Number of firm buyers and suppliers (2014).

(a) Number of downstream buyers.

| Industry | N | Mean | St Dev | 10th | 25th | 50th | 75th | 90th | 95th | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Primary and Extraction (NACE 01-09) | 50,706 | 12.1 | 60.1 | 1 | 2 | 4 | 8 | 18 | 40 | 1 |
| Manufacturing (NACE 10-33) | 57,976 | 47.5 | 284.9 | 1 | 2 | 7 | 26 | 98 | 192 | 6 |
| Utilities (NACE 35-39) | 2,734 | 192.7 | 3,305 | 1 | 2 | 6.5 | 36 | 154 | 336 | 1, |
| Construction (NACE 41-43) | 104,566 | 14.6 | 107.9 | 1 | 2 | 4 | 10 | 24 | 45 | 1 |
| Market Services (NACE 45-82) | 351,773 | 32.9 | 394.6 | 1 | 1 | 3 | 11 | 48 | 112 | 4 |
| Non-Market Services (NACE 84-96) | 22,516 | 14.1 | 183.1 | 1 | 1 | 2 | 6 | 19 | 38 | 1 |
| All | 590,271 | 29.3 | 394 | 1 | 1 | 4 | 11 | 42 | 98 | 4 |

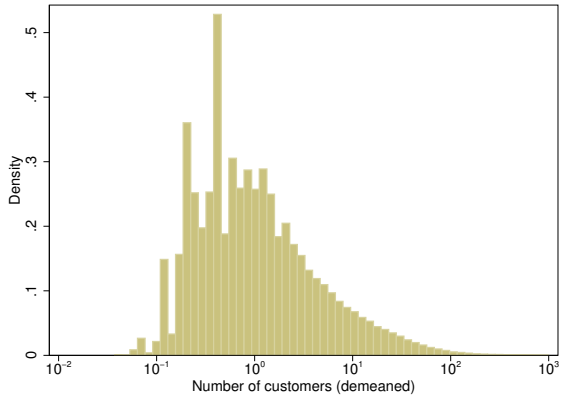
(b) Number of upstream suppliers.

| Industry | N | Mean | St Dev | 10th | 25th | 50th | 75th | 90th | 95th | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Primary and Extraction (NACE 01-09) | 60,508 | 20.5 | 29.6 | 2 | 5 | 13 | 27 | 44 | 57 | 1 |
| Manufacturing (NACE 10-33) | 72,698 | 38 | 89.5 | 2 | 5 | 15 | 38 | 89 | 148 | 3 |
| Utilities (NACE 35-39) | 3,401 | 62.8 | 180.7 | 2 | 4 | 14 | 55 | 146 | 235 | 7 |
| Construction (NACE 41-43) | 130,358 | 24.5 | 48.3 | 2 | 5 | 13 | 29 | 52 | 77 | 1 |
| Market Services (NACE 45-82) | 506,145 | 18.3 | 41.5 | 1 | 3 | 8 | 19 | 42 | 64 | 1 |
| Non-Market Services (NACE 84-96) | 67,497 | 9.7 | 37.1 | 1 | 2 | 4 | 9 | 19 | 30 | 9 |
| All | 840,607 | 20.6 | 49.5 | 1 | 3 | 9 | 22 | 46 | 71 | 1 |

*Notes:* Summary statistics for the B2B data. 10th, 25th, etc. refers to values at the 10th, 25th, etc. percentile of the distribution.

Figure 2: Distribution of firm buyer and supplier connections (2014).

(a) Number of downstream buyers, within NACE 2-digit industries.




(b) Number of upstream suppliers, within NACE 2-digit industries.

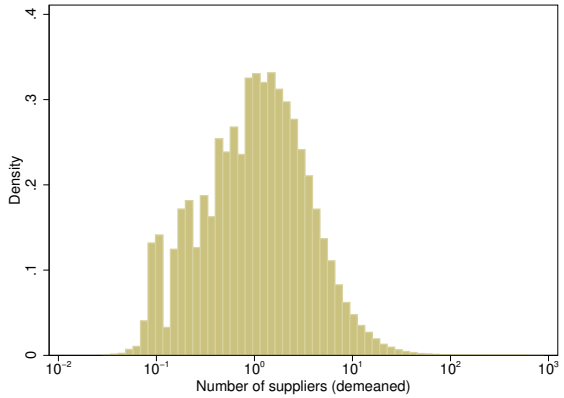


Table 3: Firm-to-firm transaction values (€, 2014).

| Industry | N | Mean | St Dev | 10th | 25th | 50th | 75th | 90th | 95th | 99th |
|---|---|---|---|---|---|---|---|---|---|---|
| Primary and Extraction | 613,868 | 39,898 | 5,409,863 | 419 | 840 | 2,490 | 9,150 | 33,789 | 81,626 | 387,573 |
| Manufacturing | 2,755,457 | 44,303 | 2,007,421 | 359 | 613 | 1,661 | 6,185 | 25,436 | 63,467 | 411,379 |
| Utilities | 526,932 | 59,953 | 7,410,682 | 366 | 615 | 1,388 | 3,744 | 11,560 | 28,382 | 281,181 |
| Construction | 1,529,078 | 24,500 | 386,201 | 375 | 676 | 1,926 | 7,000 | 27,186 | 64,585 | 339,523 |
| Market Services | 11,562,445 | 24,373 | 2,886,213 | 341 | 546 | 1,266 | 4,060 | 15,579 | 37,960 | 224,363 |
| Non-Market Services | 316,628 | 8,036 | 318,863 | 315 | 472 | 996 | 2,736 | 8,396 | 18,920 | 92,732 |
| All | 17,304,408 | 28,893 | 2,988,881 | 348 | 571 | 1,392 | 4,669 | 18,280 | 44,770 | 269,153 |

*Notes:* Summary statistics for the B2B data. 10th, 25th, etc. refers values at the 10th, 25th, etc. percentile of the distribution. Industry refers to the main industry of activity of the seller.

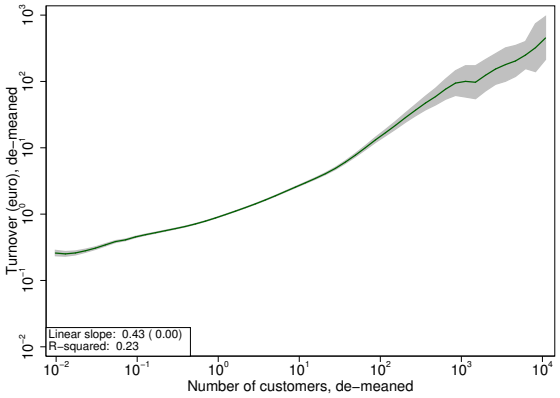
only €1,392, while the standard deviation reaches nearly €3 million and the top 10% of relationships account for 92% of all domestic firm-to-firm sales by value. This dispersion in transaction values in a buyer-supplier production network was first documented in the Belgian data by Dhyne et al. (2015). As with firm size and the extensive margin of firm connections, the intensive margin of firm linkages exhibits qualitatively similar properties within broad industries, with notable variation in magnitudes across industries.

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

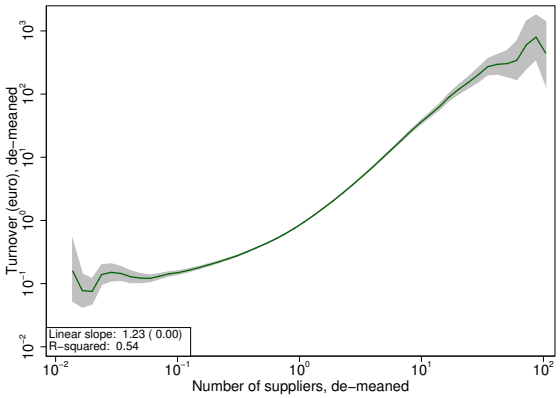
A sharp pattern in the data is that bigger firms interact with more buyers and suppliers in the production network. Figure 3a plots the fitted line and 95% confidence interval based on a local polynomial regression of firm turnover on the number of firm downstream customers, on a log-log scale. Both variables have been demeaned by their NACE 2-digit sector average, such that the latter corresponds to the point with coordinates (1,1) in the graph. Figure 3b repeats the exercise for the relationship between firm sales and number of upstream suppliers. Both figures display tightly estimated upward-sloping lines. Implied elasticities and R-squared from linear OLS regressions with NACE 2-digit industry fixed effects are also reported in the lower left corner of each graph. The estimates indicate that relative to the industry mean, a firm with 10 times more customers has approximately 4.3 times higher sales, while a producer with 10 times more suppliers attains 12.3 times higher sales.

Figure 3: Firm size and number of buyers and suppliers (2014).

(a) Firm sales and number of buyers, within NACE 2-digit industries.




(b) Firm sales and number of suppliers, within NACE 2-digit industries.




**Fact 3.** *The distribution of sales across buyers does not vary with the number of buyers. The distribution of purchases across suppliers widens with the number of suppliers.*

Facts 1 and 2 reveal broadly symmetric patterns in the extensive margin of firms' interactions with upstream suppliers and with downstream buyers in the production network. In contrast, Fact 3 uncovers asymmetry between the input and output sides along the intensive margin of firm-to-firm transactions: While the distribution of a firm's bilateral sales across customers does not vary with the number of customers, the distribution of its input purchases across suppliers widens monotonically with the number of suppliers.

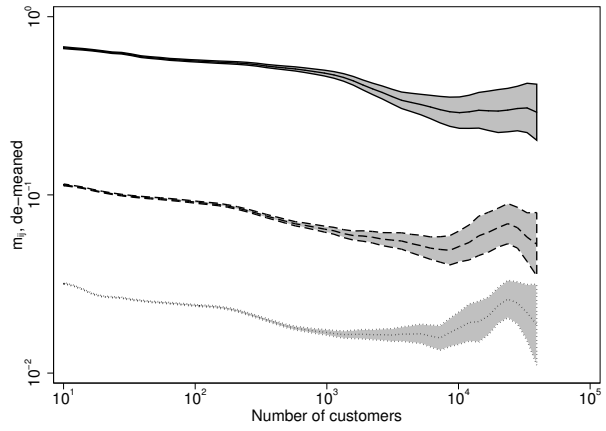
Figure 4a illustrates the dispersion of downstream sales across buyers within a seller. For each firm with at least 10 customers, we take the 10th, 50th and 90th percentile values of its bilateral sales, and demean these by NACE 2-digit industry. We plot the fitted lines from local polynomial regressions of these percentile values against firms' out-degree, including 95% confidence interval bands. The three lines we obtain are almost parallel and slightly declining. In other words, sales to the bottom, median and top customer are essentially the same, or somewhat smaller, for firms with 100 customers and for firms with 10 customers. Together with Fact 2, this suggests that larger sellers have higher sales primarily because they serve more customers, but they neither sell more to their buyers nor vary their sales more across buyers.

Figure 4b demonstrates the distribution of input purchases across upstream suppliers within a buyer. For each firm with at least 10 input providers, we obtain the 10th, 50th and 90th percentile values of its bilateral purchases, and demean by its NACE 2-digit industry. We graph the fitted lines from local polynomial regressions of these percentile values against

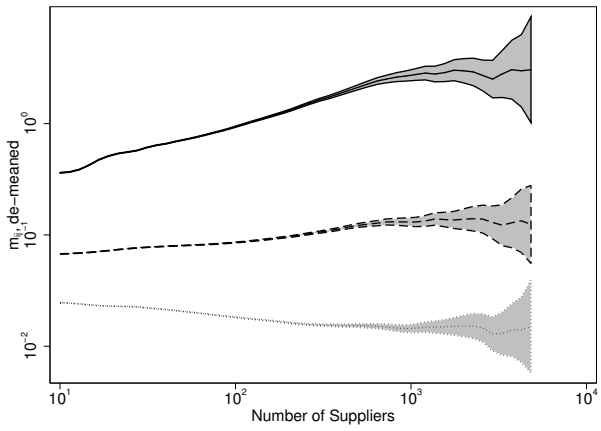
firms' in-degree, with 95% confidence interval bands. While purchases from the median supplier are once again unchanged across firms with broad and narrow supplier bases, however, firms that source inputs from more suppliers systematically buy more from their largest suppliers and less from their smallest. Together with Fact 2, this pattern implies that larger buyers have higher purchases both because they transact with more suppliers and because they vary their purchases more across suppliers.

Figure 4: Sales distribution across buyers and suppliers within firms.

(a) Number of buyers and bilateral sales..




(b) Number of suppliers and bilateral purchases.




*Notes:* Firm-to-firm sales have been demeaned by the NACE 2-digit industry of the seller. Local polynomial regressions for the the value of firm-to-firm transactions at the 10th, 50th and 90th percentile of the distribution.

*Summary.* We have documented three stylized facts which suggest that buyer-supplier linkages in a production network are key to understanding the origins of the firm size distribution. In particular, they signal an important role for (i) downstream input demand relative to final output demand, (ii) the number of buyers and suppliers of a firm, (iii) seller and buyer firm characteristics, and (iv) seller-buyer match characteristics. Motivated by these stylized facts, we next develop a unified theoretical framework that accommodates them by introducing two-sided firm heterogeneity in an input-output production network. Importantly, this model allows us to decompose the variation in the firm size distribution into economically meaningful components related to both own-firm characteristics and the production network. Of note, existing models of one-sided firm heterogeneity such as differentiated firms producing only for final consumers cannot account for (ii)-(iv), while existing models of two-sided firm heterogeneity have so far ignored either (i) or (iv).

# 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 downstream and upstream 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 $i$:

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

where $y_i$ is output, $z_i$ is productivity, $l_i$ is labor, $\alpha$ is the labor share and $\kappa > 0$ is a normalization constant.[15] $v_i$ 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 $\nu_{ki}$ 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. $\phi_{ki}$ 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 allow for heterogeneity in $\alpha$ and $\sigma$ 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 cost of the firm is then

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

[15] $\kappa = \alpha^{-\alpha} (1-\alpha)^{\alpha-1}$

## 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 $S_i$ (including exports) and firm-to-firm sales, and as such it is unnecessary to model it explicitly, see Section 2.1. In this part of the paper, we therefore take final demand as given, while Section 6 extends the model with endogenous final demand. We define $\beta_i^c$ as the ratio between total sales and sales to the network,

$$\beta_i^c \equiv \frac{S_i}{\sum_{j\in\mathcal{C}_i} m_{ij}} \geq 1 \tag{3}$$

where $\mathcal{C}_i$ is the set of network customers of firm $i$. In our data, we only observe firm-to-firm links in the domestic economy. Hence, demand from foreign firms (exports) will be part of $S_i$ but not $\sum_{j\in\mathcal{C}_i} m_{ij}$. In a similar manner, we define $\beta_i^s$ as the ratio between total purchases and purchases from the network,

$$\beta_i^s \equiv \frac{M_i}{\sum_{k\in\mathcal{S}_i} m_{ki}} \geq 1, \tag{4}$$

where $\mathcal{S}_i$ is the set of network suppliers of firm $i$.

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 $\phi_{ij}$ can be written as $\phi_{ij} = \phi_i\tilde{\phi}_{ij}$, where $\phi_i$ 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 cost, $\tau_i$ captures the average mark-up and trade cost of $i$, and $\tilde{\tau}_{ij}$ is the match-specific trade cost/mark-up term.[16] $\tilde{\tau}_{ij}$ can 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.

[16] 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 $i$. Intuitively, this normalization separates the systematic variation across firms from the variation across buyers and suppliers within firms, such that the former is fully loaded on $\phi_i$ and $\tau_i$.

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 $\phi_{ij}$). 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.

## 3.3 Mapping Buyer and Seller Effects to Model Parameters

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

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

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

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

Loosely speaking, the buyer effect $\theta_j$ reflects the magnitude of average purchases controlling for the size of suppliers. We refer to $\theta_j$ as *sourcing capability*. Intuitively, it is an input price aggregate that implicitly reflects the number and production capabilities of input suppliers. 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 such that the input price index $P_j$ in equation (6) is small.

Conversely, the seller effect $\psi_i$ reflects the magnitude of average sales controlling for the size of customers. After adjusting for input costs $P_i^{1-\alpha}$, $\psi_i$ thus identifies the productivity ($z_i$) or average attractiveness ($\phi_i/\tau_i$) of the seller in equation (7). In the following, it will become useful to define the left hand side of equation (7) as $\tilde{z}_i \equiv \left(\phi_i z_i / \left(\tau_i w^\alpha\right)\right)^{\sigma-1}$. 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 is associated with a reduction in the seller effect of $1-\alpha$. 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 test this prediction in Section 3.3.[17]

[17] While the model predicts a relationship between the buyer and seller effects, it has no prediction about the

Finally the match-specific component $\omega_{ij}$ 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 downstream and upstream factors. 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 $\xi_i \equiv \sum_{j \in \mathcal{C}_i} \theta_j \omega_{ij}$.

The components $\psi_i$ and $\xi_i$ represent upstream and downstream fundamentals in explaining firm size, respectively, while $\beta_i^c$ 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 role 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 upstream and downstream margins into various submargins. Starting with the downstream side, the 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 $\bar{\theta}_i \equiv \left( \prod_{j \in \mathcal{C}_i} \theta_j \right)^{1/n_i^c}$.[18] The covariance term $\Omega_i^c$ is defined as

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

Each of these components has an intuitive economic interpretation, 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

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.

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

if the covariance term $\Omega_i^c$ is large. This would be the case if large customers (high $\theta_j$) also happen to be a good match (high $\omega_{ij}$). As with the overall decomposition, we will regress each component in equation (10) on $\ln \xi_i$.

Next, we turn to the upstream 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 components for the number of suppliers, average supplier capability 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 (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) 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 $\Omega_i^s$ 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 ($\alpha$, $\beta_i^s$ and $n_i^s$) or can be estimated from the production network data ($\tilde{z}_i$, $\psi_i$, $\bar{\psi}_i$ and $\Omega_i^s$).

The interpretation of each element is as follows. A firm has a large market share among customers (high $\psi_i$) 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 because attractive suppliers also happen to be a good match (high $\Omega_i^s$). 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.[19]

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 an upstream factor and a downstream factor. The upstream factor comprises own production capability and network supply, where the latter

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

constitutes the number of input suppliers, average production capability across suppliers, and a covariance term. The downstream factor includes final demand and network demand, where the latter contains number of customers, average sourcing capability across customers, and a covariance term.

$$\overset{Size}{\ln S_i} = \overbrace{\overset{ProdCapability}{\ln \tilde{z}_i} + \overbrace{(1-\alpha)\left[\underset{\#Suppliers}{\ln n_i^s} + \underset{AvgSupProdCapab}{\ln \bar{\psi}_i} + \underset{Covariance}{\ln \Omega_i^s} + \underset{NetworkInputShare}{\ln \beta_i^s}\right]}^{NetworkSupply}}^{Upstream} + \overbrace{\ln \xi_i + \ln \beta_i^c}^{Downstream} \tag{13}$$

$$= \overbrace{\underbrace{\underset{\#Customers}{\ln n_i^c} + \underset{AvgCustSourceCapab}{\ln \bar{\theta}_i} + \underset{Covariance}{\ln \Omega_i^c}}_{NetworkDemand} + \overset{FinalDemand}{\ln \beta_i^c}}^{Downstream}.$$

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 with 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 seller, buyer and match effects from the production network data ($\ln \psi_i$, $\ln \theta_j$ and $\ln \omega_{ij}$). In *Step Two*, we use the first-stage estimates and observed firm outcomes to calculate unobserved firm outcomes ($\ln \xi_i$, $\ln \tilde{z}_i$, $\ln \bar{\theta}_i$, $\ln \bar{\psi}_i$, $\ln \Omega_i^c$ and $\ln \Omega_i^s$). In *Step Three*, we perform the variance decomposition itself, regressing each component of firm size on total sales $\ln S_i$, the downstream (demand-side) factor $\ln \xi_i$ and the upstream (supply-side) 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. 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.

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.[20] In practice, the estimation sample retains a substantial portion of the production network. For the baseline year, 2014, we are able to use 17,054,274 firm-to-firm transactions which capture 99% of all links in the data and 95% of

[20]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 4: Full sample vs. first-stage estimation sample (2014).

| Full Sample | | | Estimation Sample | | | |
|---|---|---|---|---|---|---|
| # Links | # Sellers | # Buyers | Links | Value | Sellers | Buyers |
| 17,304,408 | 590,271 | 840,607 | 99% | 95% | 74% | 88% |

*Notes:* Summary statistics for firm-to-firm transactions in the raw B2B data and in the estimation sample in *Step One*.

their sales value. We thus obtain seller fixed effects for 436,715 firms and buyer fixed effects for 743,326 firms. We report the characteristics of the initial and estimation samples in Table 4 below.

Figure 5 summarizes the estimation results for 2014. Three patterns stand out. First, the variation in the seller effect $\ln \psi_i$ is large compared to that in the buyer effect $\ln \theta_j$ (standard deviations of 1.05 and 0.50 respectively). Second, the $R^2$ from the regression is 0.43, and the dispersion in the residual $\ln \omega_{ij}$ (standard deviation of XXX) exceeds that in the buyer and seller effects. 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 $\ln (\theta_i / M_i)$ is -0.13 and significant at 1%.

## 4.2 Conditional Exogenous Mobility

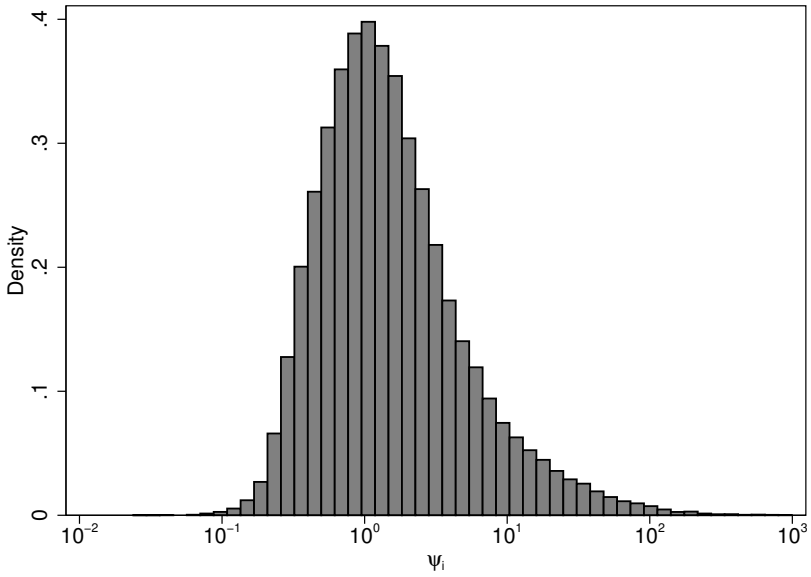
Equation (14) is a two-way fixed effects model similar to the models that are used in the employer-employee literature following Abowd et al (1999). The identifying assumptions needed for ordinary least squares to identify the parameters $\ln \psi_i$ and $\ln \theta_j$ are also similar:

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

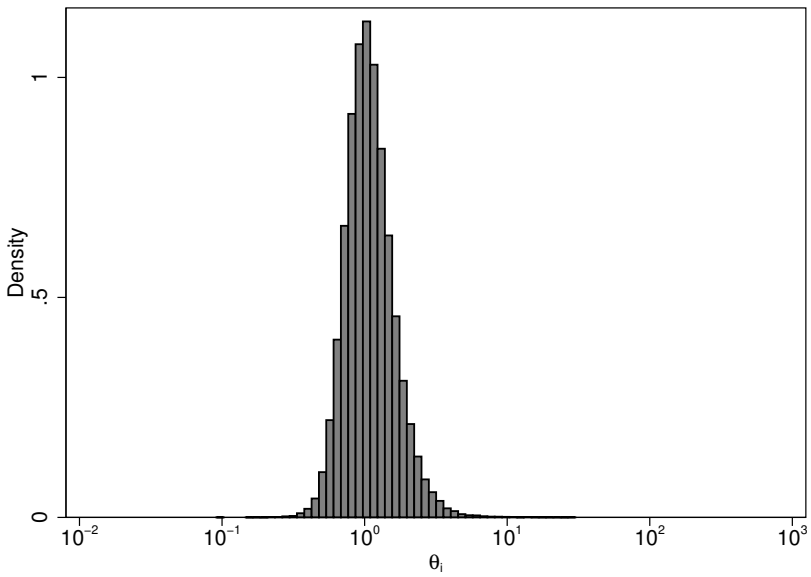
Hence, both the seller and the buyer fixed effects must be orthogonal to the match-specific error term $\omega_{ij}$. As is well known from the employer-employee literature, the key identification assumption is that the assignment of suppliers to customers is exogenous with respect to $\omega_{ij}$, so-called conditional exogenous mobility. Hence, the identification assumption is violated if

Figure 5: Distribution of seller and buyer effects (demeaned by NACE-2 sector, 2014).

(a) Seller Fixed Effect, $\ln \psi_i$. N=436,715.




(b) Buyer Fixed Effect, $\ln \theta_j$. N=743,326.



a positive shock both increases the likelihood of matching and raises $\omega_{ij}$. 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 unobserved fixed costs that do not matter for sales, such as fixed search-and-match costs. 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 that are unrelated to $\omega_{ij}$. Eaton et al (2015) develop a quasi-random matching model which would be consistent with this 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 by 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 linked to one employer at a 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 at least one $q_1$ buyer in $t = 0$ and add at least one $q_4$ buyer in $t = 1$ (year 2006), 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 quartile $q_k$. 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 it $\bar{\Delta}_i^{Up}$. 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 $\bar{\Delta}_i^{Down}$. We find that 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, this assymmetry goes in the opposite direction to what one would expect under

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 94,357 firms with complete data for the purpose. 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 ratio, $\beta_i^c$, is calculated as total sales divided by the sum of all sales to other firms in the domestic network from B2B. Since we observe firm-to-firm links only within Belgium, sales to foreign firms (exports) are classified as part of final demand.[21] We measure firm purchases $M_i$ with total input expenditures from CBSO.[22] The network input ratio, $\beta_j^s$, is then total input purchases divided by the cost of inputs from suppliers in the domestic network . 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.

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 $\alpha$ in the Cobb-Douglas production technology, the wage rate $w$, and the elasticity of

[21] 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.

[22] 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.

Table 5: Firm size components (demeaned by NACE-2 sector, 2014).

| Firm Size Component | Estimated? | N | Mean | Median | St Dev |
|---|---|---|---|---|---|
| Total Sales, $\ln S_i$ | | 94,357 | 0.000 | -0.141 | 1.377 |
| *Overall Decomposition of* $\ln S_i$ | | | | | |
| Usptream Supply, $\ln \psi_i$ | Y | 94,357 | 0.000 | -0.147 | 1.052 |
| Downstream Network Demand, $\ln \xi_i$ | Y | 94,357 | 0.000 | 0.001 | 1.710 |
| 1 - Final Demand Share, $\ln \beta_i^c$ | | 94,357 | 0.000 | -0.359 | 1.285 |
| *Upstream Decomposition of* $\ln \psi_i$ | | | | | |
| Production Capability, $\ln \tilde{z}_i$ | Y | 94,357 | 0.000 | -0.113 | 1.346 |
| # Suppliers, $\ln n_i^s$ | | 94,357 | 0.000 | -0.007 | 0.797 |
| Avg Supplier Capability, $\ln \bar{\psi}_i$ | Y | 94,357 | 0.000 | -0.022 | 0.237 |
| Supplier Covariance, $\ln \Omega_i^s$ | Y | 94,357 | 0.000 | -0.089 | 0.684 |
| Network Supply Share, $\ln \beta_i^s$ | | 94,357 | 0.000 | -0.146 | 0.557 |
| *Downstream Decomposition of* $\ln \xi_i$ | | | | | |
| # Customers, $\ln n_i^c$ | | 94,357 | 0.000 | -0.040 | 1.446 |
| Avg Customer Capability, $\ln \bar{\theta}_i$ | Y | 94,357 | 0.000 | -0.037 | 0.334 |
| Customer Covariance, $\ln \Omega_i^c$ | Y | 94,357 | 0.000 | -0.142 | 0.761 |

substitution $\sigma$. To accommodate the variation in production technologies and factor costs across industries, we proxy $\alpha$ 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 $\sigma$ 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 capability 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 5 below provides summary statistics for these components, while Table 13 in Appendix B reports all two-way correlation coefficients among them.

# 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 upstream and downstream 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 upstream and downstream 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 upstream or downstream 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 $\Delta_S$ 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 variance decomposition therefore estimates the importance of each margin in explaining within-industry size heterogeneity.

## 5.1 Baseline Results

### 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 final demand $\ln\beta_i^c$, the upstream factor $\ln \psi_i$ and the downstream factor $\ln \xi_i$, from equation (9), 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 Table 6. The downstream side accounts for fully 80% of the size dispersion across firms, the upstream fundamentals explain 19%, while final demand explains only 1%.

What is the interpretation of these results? The upstream factor $\ln \psi_i$ represents, loosely speaking, the average market share of $i$ among its customers. Hence, the relatively small role for upstream 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

Table 6: Overall Decomposition (2014).

| | N | Final Demand $\ln\beta_i^c$ | Upstream $\ln\psi_i$ | Downstream $\ln\xi_i$ |
|---|---|---|---|---|
| $\ln S_i$ | 94,357 | .01** | .19*** | .80*** |
| | | (.00) | (.00) | (.00) |

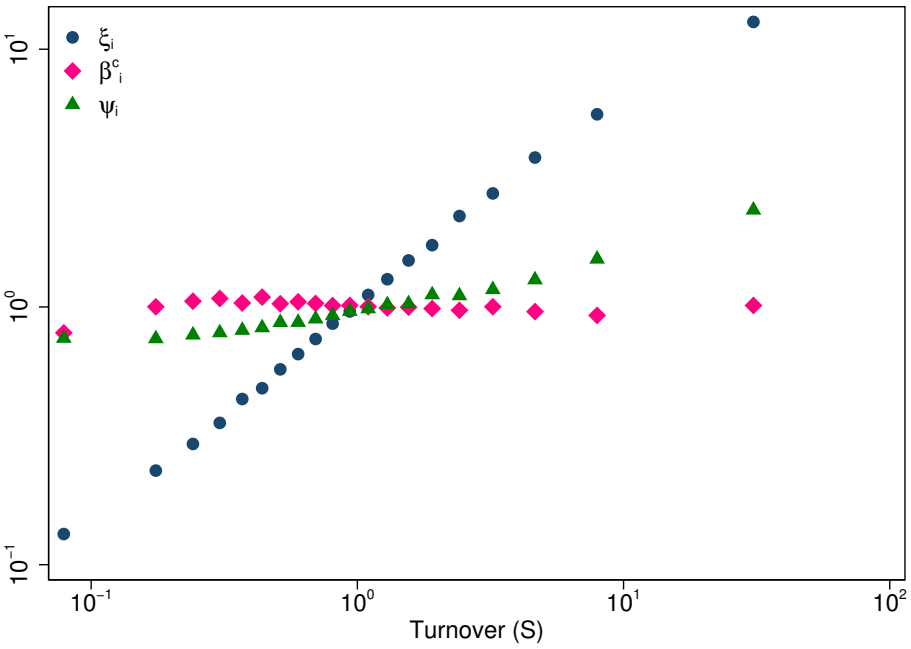
Note: The table reports coefficient estimates from OLS regressions of a firm size margin (as indicated in the column heading) on total firm sales (all variables in logs). Standard errors in parentheses. Significance: * < 5%, ** < 1%, *** <0.1%.

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. 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 1% of the overall variation in firm size. Hence, practically the entire downstream factor is governed by demand from other firms in the production network, rather than from final demand. The small and insignificant effect of final demand $\ln\beta_i^c$ means that large firms are not systematically selling relatively more (or less) to final demand than small firms.

We can also visualize the importance of each component using a binned scatterplot. In Figure 6, we group log sales into 20 equal-sized bins and compute the mean of log sales and the components $\ln \psi_i$, $\ln \xi_i$ and $\ln \beta_i^c$ within each bin, and then create a scatterplot of these data points. The result is a non-parametric visualization of the conditional expectation function, and because the components sum to log sales, the sum of the components on the vertical axis equals log sales on the horizontal axis. Again, we observe the dominance of the downstream component, and furthermore that the relationship is close to linear across the whole distribution of firm sales.

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).

Figure 6: Overall Decomposition (2014).




Note: The binned scatterplot groups the log sales into 20 equal-sized bins, computes the mean of log sales and the components $\ln \psi_i$, $\ln \xi_i$ and $\ln \beta_i^c$ within each bin, then creates a scatterplot of these data points. The result is a non-parametric visualization of the conditional expectation function.

Table 7: Downstream Decomposition (2014).

| | # Customers $\ln n_i^c$ | Avg Customer Capability $\ln \bar{\theta}_i$ | Customer Covariance $\ln \Omega_i^c$ |
|---|---|---|---|
| $\ln \xi_i$ | $0.72^{***}$ | $0.03^{***}$ | $0.25^{***}$ |
| | (.00) | (.00) | (.00) |

Note: The table reports coefficient estimates from OLS regressions of a firm size margin (as indicated in the column heading) on the downstream factor, $\ln \xi_i$. Standard errors in parentheses. Significance: * $< 5\%$, ** $< 1\%$, *** $<0.1\%$.

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 upstream and downstream factors of the firm size suggests that the other facts also implicitly reflect how the production network shapes the firm size distribution.

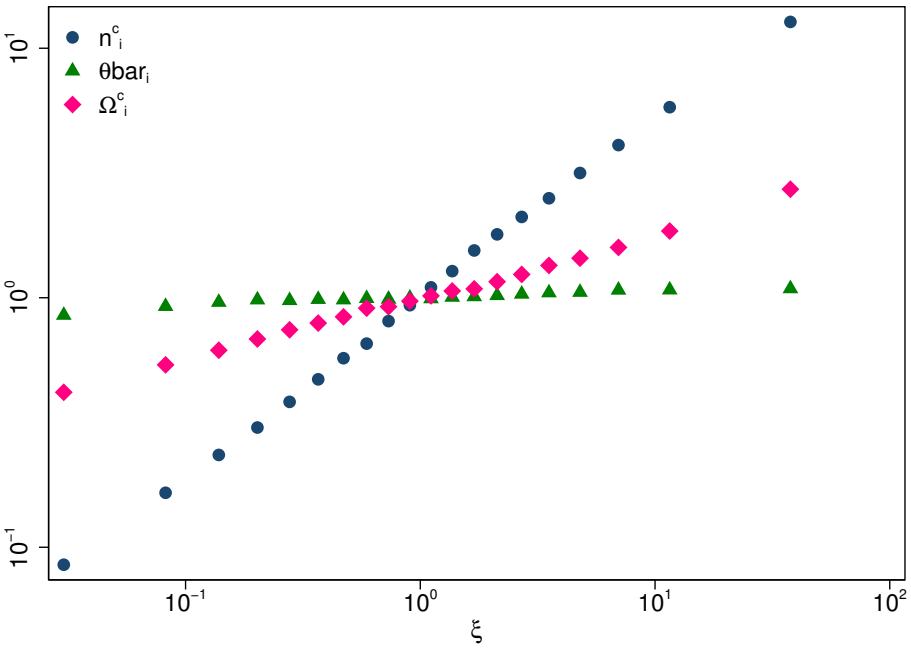
### 5.1.2 Downstream Decomposition

We next decompose the downstream component into their consituent parts, from equation (10), to assess the specific channels through which the production network shapes firm sales. Table 7 reports the results from regressing each downstream subcomponent against $\ln \xi_i$, such that the coefficient estimates quantify the relative importance of each component. An overwhelming 72% of the variation in the downstream component across firms can be attributed to the extensive margin, i.e. the number of (domestic) buyers $\ln n_i^c$ 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 3% and 25%, respectively. As above, we also report the results using a binned scatterplot in Figure 7.

We conclude that on the sales side, the single most important advantage of large firms is that they successfully match with many buyers. The covariance term is also substantial, suggesting that bigger firms also tend to have more skewed sales to large buyers, relative to smaller firms. On the other hand, large firms are not matching with more capable buyers, on average (the $\ln \bar{\theta}_i$ term).

The downstream 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 $\ln \bar{\theta}_i$ reinforces *Fact 3* that the distribution of a firm's sales across customers is

Figure 7: Downstream decomposition (2014).




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

Table 8: Upstream Decomposition (2014).

| | Own capability $\ln \tilde{z}_i$ | # Suppliers $\ln n_i^{s*}$ | Avg Supplier Capability $\ln \bar{\psi}_i^*$ | Supplier Covariance $\ln \Omega_i^{s*}$ | Non-Network Input Share $\ln \beta_i^{s*}$ |
|---|---|---|---|---|---|
| $\ln \psi_i$ | 0.82*** | -0.01*** | 0.05*** | 0.10*** | 0.04*** |
| | (.00) | (.00) | (.00) | (.00) | (.00) |

Note: The table reports coefficient estimates from OLS regressions of a firm size margin (as indicated in the column heading) on the upstream factor, $\ln \psi_i$. $*$ denotes that the variable is multiplied by $(1-\alpha)$. Standard errors in parentheses. Significance: * $< 5\%$, ** $< 1\%$, *** $<0.1\%$.

generally invariant with the number of its customers.

### 5.1.3 Upstream Decomposition

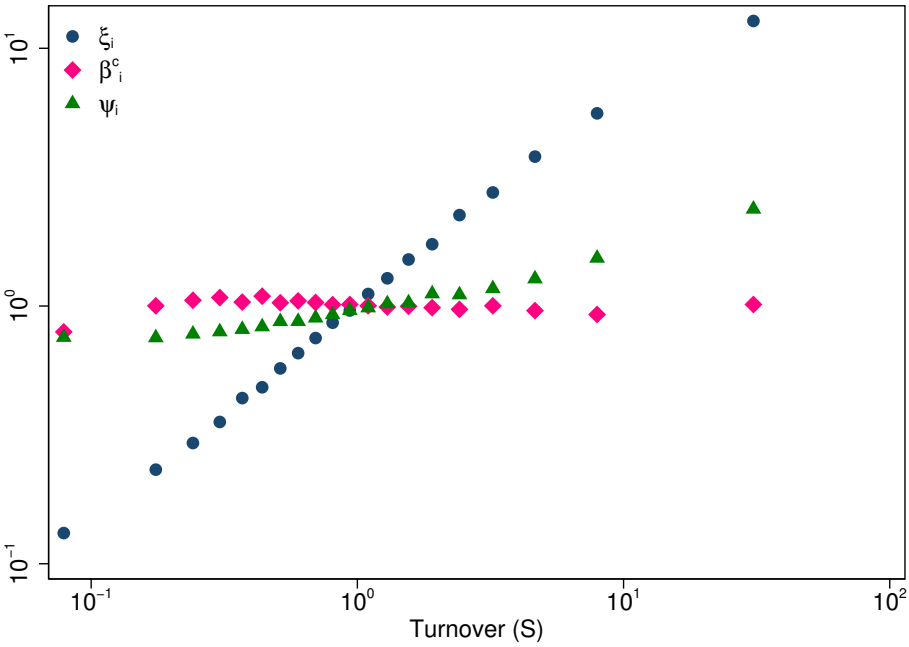
We complete the firm size decomposition by unbundling the upstream margin of firm sales, $\ln \psi_i$, from equation (12). Table 8 reports the results from regressing each subcomponent against the upstream factor $\ln \psi_i$. As above, we also report the results using a binned scatterplot in Figure 8.

The seller-specific production capability $\ln \tilde{z}_i$ drives the overwhelming majority of the upstream factor (82%). The remaining factors are loaded on average supplier capability (5%), the covariance term (10%) and the non-network input share (4%). Differently from the downstream side, the number of suppliers does not explain variation in the firm size.

These results shed light on how successful firms are able to increase their market shares among customers. First and foremost, inherent firm characteristics, such as productivity or quality (the $\ln \tilde{z}_i$ term), explain differences in market shares. According to our results, firms that have good suppliers (the $\ln \bar{\psi}_i$ term), or that tend to source relatively more from good suppliers (the $\ln \Omega_i^s$ term), are also more successful in terms of sales, although the economic magnitude is less pronounced.

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

Figure 8: Upstream Decomposition (2014).




Note: The binned scatterplot groups the log sales into 20 equal-sized bins, computes the mean of log sales and the components $\ln \tilde{z}_i$, $\ln n_i^{s*}$, $\ln \bar{\psi}_i^*$, $\ln \Omega_i^{s*}$ and $\ln \beta_i^{s*}$, where $*$ denotes that the variable is multiplied by $(1-\alpha)$, then creates a scatterplot of these data points. The result is a non-parametric visualization of the conditional expectation function.

Table 9: Results by Year (2002-2014).

| | | N | Final Demand $\ln\beta_i^c$ | Upstream $\ln\psi_i$ | Downstream $\ln\xi_i$ |
|---|---|---|---|---|---|
| $\ln S_i$ | 2002 | 81,410 | .04*** | .19*** | .78*** |
| | 2003 | 83,817 | .03*** | .19*** | .78*** |
| | 2004 | 85,174 | .03*** | .20*** | .77*** |
| | 2005 | 86,617 | .03*** | .19*** | .78*** |
| | 2006 | 88,714 | .03*** | .19*** | .79*** |
| | 2007 | 91,172 | .02*** | .19*** | .79*** |
| | 2008 | 92,465 | .02*** | .19*** | .79*** |
| | 2009 | 92,528 | .03*** | .18*** | .79*** |
| | 2010 | 92,903 | .02*** | .19*** | .79*** |
| | 2011 | 94,282 | .02*** | .19*** | .79*** |
| | 2012 | 95,558 | .02*** | .19*** | .79*** |
| | 2013 | 94,324 | .01*** | .19*** | .79*** |
| | 2014 | 94,357 | .01** | .19*** | .80*** |

Note: The table reports coefficient estimates from OLS regressions of a firm size margin (as indicated in the column heading) on total firm sales (all variables in logs). Standard errors in parentheses. Significance: * < 5%, ** < 1%, *** <0.1%.

could also more effectively allocate their input purchases towards suppliers with both higher production capability and match quality.

## 5.2 Results by Year

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 Table 9. The importance of the upstream side has hovered around 18-20%. The downstream side has gradually risen from 78% to 80%, closely following a decline in final demand from 4% to 1%.

We observe similarly stable patterns when we consider the lower-tier decomposition of downstream and upstream sub-components.[23] These findings suggest that there may be inherent drivers of the firm size distribution whose relative importance remains stable despite

[23]Results available upon request.

the rise in production fragmentation across firm and country boundaries over the last 15 years.

## 5.3 Firm Growth

The baseline decomposition relates variance of sales across firms to variance of the various margins. A related question is what explains the variance of firm growth. We proceed as follows. First, we estimate equation (14) on two cross-sections, the baseline year 2014 ($t = 1$) and year 2002 ($t = 0$). We then calculate the change in every demeaned variable in the decomposition. For example, the overall decomposition from (9) becomes

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

where $\Delta_T$ denotes the change from $t = 0$ to $t = 1$, e.g. $\Delta_T \ln S_i = \Delta_S \ln S_{i1} - \Delta_S \ln S_{i0}$. We then regress each component, e.g. $\Delta_T \ln \psi_i$, on $\Delta_T \ln S_i$. This decomposition allows us to assess the importance of the network in explaining firm-level growth. Note that long differencing is only feasible for firms that are observed with non-missing sales as well as buyer and seller effects in both years, i.e. we cannot perform the decomposition on firms that enter or exit during the sample period. However, the decomposition allows for adding and dropping of customers and suppliers, i.e. the terms $\Delta_T \ln \psi_i$ and $\Delta_T \ln \xi_i$ may change because of extensive margin adjustments.

The results are summarized in Table 10. Interestingly, the contributions of each component are very close to what we found in the cross-sectional analysis in Section 5.1. For example, the downstream component is dominating in the overall decomposition (column 3), whereas the number of customers explains most of the variation in $\xi_i$ (column 4), and own production capability explains most of the variation in $\psi_i$ (column 4).

## 5.4 Results by Industry

Finally, we explore the robustness of our results across different industries. Table 11 reports separate results for six broad industry groups. Across all of these groups, the estimated coefficients relatively close to the baseline results reported in Section 5.1, underscoring the robustness of our results. One exception is construction (NACE 41 to 43), where the final demand term $\beta_i^c$ enters with a coefficient of -0.10. However, this is as expected, as large construction firms typically sell relatively less to final demand compared to small construction firms. We have also performed the decomposition separately for every 2-digit NACE industry. For the large majority of industries, we find that the overall decomposition looks strikingly similar to the baseline results.

Table 10: Firm Growth Decomposition, 2002-2014.

| | | Sales $S_i$ | Downstream $\xi_i$ | Upstream $\psi_i$ |
|---|---|---|---|---|
| Final Demand | $\beta_i^c$ | .10*** | | |
| Upstream | $\psi_i$ | .09*** | | |
| Downstream | $\xi_i$ | .81*** | | |
| # customers | $n_i^c$ | | .62*** | |
| Avg. customer capability | $\bar{\theta}_i$ | | .01*** | |
| Covariance | $\Omega_i^c$ | | .37*** | |
| Production capability | $\tilde{z}_i$ | | | .98*** |
| # suppliers | $n_i^s$ | | | -.03*** |
| Avg. supplier capability | $\bar{\psi}_i$ | | | .01*** |
| Covariance | $\Omega_i^c$ | | | .03*** |
| Outside network supply | $\beta_i^s$ | | | .01*** |
| N | | 41,185 | 41,185 | 41,185 |

Note: The table reports coefficient estimates from OLS regressions of a firm size margin (as indicated in the row heading) on total firm sales (column 3) or $\ln\xi_i$ (column 4) or $\ln\psi_i$ (column 5). All variables in logs. Standard errors in parentheses. Significance: * < 5%, ** < 1%, *** <0.1%.

Table 11: Firm Size Decomposition: Across Industries (2014).

| NACE | Industry | N | Final Demand $\beta_i^c$ | Upstream $\psi_i$ | Downstream $\xi_i$ |
|---|---|---|---|---|---|
| 01-09 | Primary and Extraction | 2,838 | -.03** | .24*** | .79*** |
| 10-33 | Manufacturing | 16,905 | -.01** | .26*** | .75*** |
| 35-39 | Utilities | 852 | .04** | .15*** | .81*** |
| 41-43 | Construction | 19,008 | -.10** | .11*** | .99*** |
| 45-82 | Market Services | 53,604 | .04** | .18*** | .76*** |
| 84-96 | Non-Market Services | 1,150 | .04 | .12*** | .84*** |

Note: The table reports coefficient estimates from OLS regressions of a firm size margin (as indicated in the column heading) on total firm sales. All variables in logs. Standard errors in parentheses. Significance: * < 5%, ** < 1%, *** <0.1%.

# 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 $\sigma$ 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.[24] 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 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.[25] As a consequence, a firm's purchases relative to total sales, $\mu_i \equiv M_i / S_i$ is constant.[26] We therefore use data on $\mu_i$ 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 $\Pi$ 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 $\rho$ is a constant term.

[24] Since final demand is modeled as a representative consumer, there is by construction no match specific component $\phi_{ki}$. 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 $j$.

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

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

*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 $\omega_{ij}$. We solve the fixed point using data on $\alpha$ and our estimates of $\tilde{z}_i$ and $\omega_{ij}$ 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 $\tilde{\mathcal{P}} \equiv \mathcal{P}^{1-\sigma}$:

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

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 $\omega_{ij}$. We solve the fixed point using (i) data on $\alpha$, final demand $X$ and $\mu_i$, (ii) our estimates of $\tilde{z}_i$ and $\omega_{ij}$, 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 $\sigma$.

*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 $\omega_{ij}$ 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, 12, 2.2

Bernard, A., Moxnes, A., and Ulltveit-Moe, K. (2015b). Two-sided heterogeneity and trade,. *NBER Working Paper*, 20136. 1, 12, 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 $\tilde{z}_i \equiv \left(\phi_i z_i / \left(\tau_i w^\alpha\right)\right)^{\sigma-1}$. Substituting for $\ln M_i$ from equation (11) and using equation (5) yields

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

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

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

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 downstream 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

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

*Donwstream 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 and 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: At each purchase transaction, firms pay their input suppliers VAT on top of the value of the inputs sourced. At each sales transaction, firms charge their buyers VAT on top of the sales value, and in effect transfer to the tax authorities only taxes due on the valueadded at that stage. The tax is neutral to the firm (other than potentially through its effect on firms' pre-tax pricing strategy), 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.[27]

VAT-liable firms have to file periodic VAT declarations and VAT listings with the tax administration.[28] 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 firm size, 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

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

[28] Sample VAT declaration forms can be found at here (French) and here (Dutch). Sample VAT listings forms can be found at here (French) and here (Dutch).

observation in this listing refers to the yearly values of total sales from firm $i$ to firm $j$. The reported value is the sum of the value of all invoices from $i$ to $j$. Whenever this aggregate value is larger than or equal to 250 euro, the relationship has to be reported. Sanctions for incomplete and erroneously reporting guarantee the high quality of the data.

## B.2 Data sources

The empirical analysis draws on three main 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. 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 sales relationships among 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 information on firms' annual accounts from the Central Balance Sheet Office at the NBB for the years 2002 to 2014. Firms above a certain size threshold have to file annual accounts at the end of the fiscal year.[29] We retain information on the firm identifier, turnover (total sales in euro, code 70 in the annual accounts), input purchases (total material and services inputs in euro, codes 60+61), labor cost (total cost of wages, social securities and pensions in euro, code 62) and employment (average number of full-time equivalent (FTE) employees, code 9087). Small firms submit abbreviated annual accounts and do not have to report turnover and intermediate input purchases . We annualize all flow variables in 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.[30]

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

[30] 78% of firms have annual accounts that coincide with calendar years, while 98% of firms have fiscal years of 12 months.

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 codes over time to the NACE Rev. 2 version to deal with changes in the NACE classification over our panel from Rev. 1.1 to Rev. 2. Table 12 lists industry groups at the NACE 2-digit level. We also extract the main location of the firm at the postal code level from the Crossroads Bank.

## B.3 Data construction and cleaning

Wages are calculated as labor cost over FTE employment. 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 NACE 2-digit industry, by first summing over all firms' labor costs in that industry and then dividing by total FTE employment or total turnover in that industry. We drop firms that have missing employment information or less than one FTE employee.

Geographical distance between firm pairs is calculated as the distance in kilometers between the firms' 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 squared kilometers.

Firm pairs are indexed by the Cantor pairing function to keep the pairing identity consistent over the panel.[31]

Throughout the paper, we report statistics on both the full sample in the raw data and the estimation sample used in the firm size decomposition. For the full sample, we keep all firm-to-firm relationships in the NBB B2B dataset, even if there is missing firm-level information, as these contribute to the decomposition exercise. We thus keep all firms that show up in the network as either a buyer or a seller. For the estimation sample, in *Step One* we first estimate the two-way fixed effects regression on the full sample. Note that if a buyer or seller has only 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 that have at least two sellers or buyers remain. Finally, for the decomposition exercise to contain the same number of observations across all (sub-)components, in *Step Two* and *Step Three* we keep only firms that show up as both buyer and seller in the network.

## B.4 Supplementary empirical results

This sub-section contains additional empirical results that we refer to in the main text.

[31] 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.

Table 12: NACE classification of industry groups.

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

Table 13 reports all two-way correlation coefficients among the various firm size components that we use in the decomposition analysis.

Table 13: Correlation among firm size components (demeaned by NACE-2 sector, 2014).

| | Firm Size Component | $\ln S_i$ | $\ln \psi_i$ | $\ln \xi_i$ | $\ln \beta_i^c$ | $\ln \tilde{z}_i$ | $\ln n_i^s$ | $\ln \bar{\psi}_i$ | $\ln \Omega_i^s$ | $\ln \beta_i^s$ | $\ln n_i^c$ | $\ln \bar{\theta}_i$ | $\ln \Omega_i^c$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total Sales, $\ln S_i$ | 1 | | | | | | | | | | | |
| $\ln S_i$ | Usptream Supply, $\ln \psi_i$ | 0.25 | 1 | | | | | | | | | | |
| | Downstream Network Demand, $\ln \xi_i$ | 0.65 | 0.14 | 1 | | | | | | | | | |
| | 1 - Final Demand Share, $\ln \beta_i^c$ | 0.01 | -0.36 | -0.52 | 1 | | | | | | | | |
| $\ln \psi_i$ | Production Capability, $\ln \tilde{z}_i$ | -0.52 | 0.64 | -0.58 | -0.30 | 1 | | | | | | | |
| | # Suppliers, $\ln n_i^s$ | 0.76 | -0.02 | 0.62 | 0.00 | -0.61 | 1 | | | | | | |
| | Avg Supplier Capability, $\ln \bar{\psi}_i$ | 0.19 | 0.28 | 0.01 | -0.04 | 0.02 | 0.02 | 1 | | | | | |
| | Supplier Covariance, $\ln \Omega_i^s$ | 0.65 | 0.18 | 0.39 | 0.04 | -0.40 | 0.36 | 0.17 | 1 | | | | |
| | Network Supply Share, $\ln \beta_i^s$ | 0.10 | 0.08 | 0.00 | 0.04 | -0.08 | -0.14 | -0.13 | -0.27 | 1 | | | |
| $\ln \xi_i$ | # Customers, $\ln n_i^c$ | 0.47 | -0.31 | 0.85 | -0.37 | -0.61 | 0.55 | -0.07 | 0.26 | -0.02 | 1 | | |
| | Avg Customer Capability, $\ln \bar{\theta}_i$ | 0.21 | 0.21 | 0.15 | -0.15 | 0.05 | 0.07 | 0.06 | 0.11 | 0.09 | -0.18 | 1 | |
| | Customer Covariance, $\ln \Omega_i^c$ | 0.46 | 0.18 | 0.57 | -0.41 | -0.18 | 0.33 | 0.13 | 0.32 | -0.01 | 0.09 | 0.24 | 1 |

*Notes:* All correlations are significant at 5% except those strictly below 0.01.