Grant agreement No. 822390

MICROPROD

Raising EU Productivity: Lessons from Improved Micro Data

H2020-SC6-TRANSFORMATIONS-2018
Supply and demand-oriented economic policies to boost robust growth in Europe –
Addressing the social and economic challenges in Europe

D1.6
Analysis of the importance of globalization and fragmentation of production (GVCs) for productivity measurement

WP 1 – Firm-level Data and Productivity Measurement

Due date of deliverable	Month 42 – June 2022
Actual submission date	14/06/ 2022
Start date of project	01/01/2019
Duration	42 months
Lead beneficiary	KRTK
Last editor	Balázs Muraközy
Contributors	KRTK

Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

This Project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 822390.

Disclaimer

The content of this deliverable does not reflect the official opinion of the European Union. Responsibility for the information and views expressed herein lies entirely with the author(s).

History of the changes

Version	Date	Released by	Comments
1.0	15-06-2022	Balázs Muraközy	First draft

Table of contents

Key word list

supplier-buyer networks, quality, assortativity, discrete choice, matching frictions

1. Introduction

1.1 General Context

The increasing importance of global value chains has created both scientific and policy interest in supplier-buyer relationships. Despite their importance, we know surprisingly little about how relationships between companies form and operate. Data availability is often a key constraint for such research. The few firm-to-firm datasets that exist contain little information on how relationships form and operate.

1.2 Deliverable objectives

This study introduces uses firm-to-firm administrative (VAT) data to investigate the structure of supplier networks and how the characteristics of the supplier set are related to buyer performance, including productivity. Therefore, the deliverable demonstrates both how these new data sources can be used to study the impact of suppliers on buyer performance and produces substantive results about the role of quality in networks.

2. Methodological approach

We use transaction-level data from Hungary to investigate the effects of quality and distance on the formation of supplier-client networks. We document four facts that suggest a complementarity between supplier and client firm quality (measured with the average wage), and that clients trade off the distance to and the quality of suppliers. We build a model of endogenous supplier choice that incorporates a quality complementarity and a distance-based search friction. The model yields estimating equations which express the choice of supplier and the share spent on each supplier as a function of supplier and client quality, supplier price and distance.

3. Summary of activities and research findings

We use firm-to-firm administrative data from Hungary coming from VAT declarations between 2015-2019. These data cover all firms in Hungary which pay VAT above a small threshold. We link the data to financial statements and proxy the quality of each firm by the average wage. A key novel element of our paper is that we can observe supplier prices from the unit values from the Prodcom production database for a subset of firms. Prodcom includes information on produced and sold quantity and revenue by product, allowing us to proxy supplier prices by the average unit values of the supplier.

We build a partial equilibrium model with endogenous supplier choice to capture two main features derived from these facts: the model includes complementarity between supplier and buyer quality and buyers face a trade-off between supplier distance and supplier quality. In the model, client (buyer) firms sell output to consumers and buy inputs from suppliers. They have production functions for both quantity and quality. Quality production combines client quality and supplier quality and allows for a complementarity between them. Distance from supplier affects quality in two ways: i) fixed search-and-matching cost and ii) input-share-related iceberg trade cost. Under these conditions, client firms choose their suppliers and input quantities taking into account these effects. Importantly, the model makes predictions on both the extensive (supplier choice) and the intensive (supplier shares) margin of supplier choice, which we test.

In our main empirical exercise, we use the equations derived from the model to estimate both the extensive and intensive margin decisions. Importantly, we observe directly the key cost variables, supplier price and distance between the two firms, so we can estimate the tradeoff in terms of choices between quality and these two key variables both for higher- and lower-quality buyers. This approach also allows us to control for a rich set of fixed effects and yields a discrete choice model for supplier choice.

4. Conclusions and future steps

Increasing supplier quality by 10% increases revenue by 2.7 percentage points more for 10% better client. In addition, we find that distance has a stronger effect for the extensive margin compared to the intensive margin, suggesting that search frictions, rather than only transportation costs, drive the trade-off between distance and quality. This trade-off is indeed substantial: a 10\% increase in distance reduces choice probability as much as a 30% increase in price.

5. Publications resulting from the work described

Published as working paper.

Quality Complementarity and Search Frictions in Supplier-client Networks*

Attila Lindner — UCL, IFS, KRTK
Balázs Muraközy — ULMS, KRTK
Adam Szeidl — CEU

June 2022

Abstract

We use transaction-level data from Hungary to investigate the effects of quality and distance on the formation of supplier-client networks. We document four facts that suggest a complementarity between supplier and client firm quality (measured with the average wage), and that clients trade off the distance to and the quality of suppliers. We build a model of endogenous supplier choice that incorporates a quality complementarity and a distance-based search friction. The model yields estimating equations which express the choice of supplier and the share spent on each supplier as a function of supplier and client quality, supplier price and distance. Estimating these equations implies, through the lenses of the model, that quality complementarity and search frictions are quantitatively important: a 10% increase in supplier quality leads to a 2.7-12.6% increase in client's sales, and a 10% increase in distance has a similar effect as a 30% increase in price on the probability of choosing a supplier.

keywords: supplier-buyer networks, quality, assortativity, discrete choice, matching frictions

JEL-codes: D21, D22, D85, L14

*We thank Istvan Boza, Rebeka Cook, Luca Drucker, and Balazs Krusper for research assistance, Miklos Koren and seminar audiences for helpful comments, and the European Union's Horizon 2020 research and innovation programme grant agreement numbers 724501 (for Szeidl) and 822390 (for Murakozy) and 949995 (for Lindner) for support. Lindner acknowledges financial support from the Economic and Social Research Council (new investigator grant, ES/T008474/1).

1 Introduction

Modern economies are complex ecosystems and firm-to-firm trading relationships are one of the key links in this system. The functioning of these ecosystem can affect economic development and the distribution of economic activities across space. For example, geographic concentration could be driven by access to high quality suppliers, as suggested by Marshall (1890).

Granularity and firms networks have been shown to affect significantly how the economy reacts to different shocks (see e.g. Miyauchi et al. 2018, Bernard & Moxnes 2018, Bernard, Dhyne, Magerman, Manova & Moxnes 2019). The broad structure of the network matters in how these shocks propagate (Dhyne et al. 2021) and how firms compete with each other (Kikkawa et al. 2019). Much of previous research has focused on the assortativity of the network in terms of the cost of production and documented a negative assortativity between buyers' and their suppliers' average cost level (Bernard & Moxnes 2018, Bernard et al. 2018, Bernard, Moxnes & Saito 2019).

A recent literature has documented the importance of input quality in firm production (see e.g. Verhoogen 2008, Kugler & Verhoogen 2012). Motivated by that literature, we focus on quality in supplier-buyer linkages. The importance of quality in the formation of supplier-buyer networks has already been documented in Turkey (Demir et al. 2021) and Uganda (Spray 2017). Indeed, our data shows a positive relationship between supplier and buyer prices, rather than a negative one as negative assortativity in terms of productivity would suggest (Figure 1).

In this paper we investigate two questions. First, are quality complementarities between suppliers and clients are key feature of the economy? Second, do distance-based frictions limit efficient formation of supplier-client relationships?

We use firm-to-firm administrative data from Hungary coming from VAT declarations between 2015-2019. These data cover all firms in Hungary which pay VAT above a small threshold. We link the data to financial statements and proxy the quality of each firm by the average wage, following Demir et al. (2021). A key novel element of our paper is that we can observe supplier prices from the unit values from the Prodcom production database for a subset of firms. Prodcom includes information on produced and sold quantity and revenue by product, allowing us to proxy supplier prices by the average unit values of the supplier.

We document four facts to motivate our model and main regression analysis. First, higher quality suppliers have higher quality buyers on average. This positive assortativity in quality is our starting point and our aim is to provide new insights about the mechanisms behind this pattern. The second and third facts focus on the role of distance in supplier (and quality) choice. The second fact establishes that high quality buyers choose more distant suppliers if there are few high-quality suppliers locally. The third fact is that high quality buyers go farther away to find good quality suppliers. These facts are consistent with substantial distance-related frictions in the matching process and also with higher-quality buyers applying a different tradeoff between these frictions and supplier quality compared to lower-quality buyers. The fourth fact documents that buyers with higher quality suppliers perform better. This is true for a number of firm-performance metrics: profit, sales, value added and the per-worker versions of these measures. Having higher quality suppliers — and having access to higher quality suppliers — is associated with higher productivity, larger size and profitability.

We build a partial equilibrium model with endogenous supplier choice to capture two main features derived from these facts: the model includes complementarity between supplier and buyer quality and buyers face a trade-off between supplier distance and supplier quality.

In the model, client (buyer) firms sell output to consumers and buy inputs from suppliers. They have production functions for both quantity and quality. Quality production combines client quality and supplier quality and allows for a complementarity between them. Distance from supplier affects quality in two ways: i) fixed search-and-matching cost and ii) input-share-related iceberg trade cost. Under these conditions, client firms choose their suppliers and input quantities taking into account these effects. Importantly, the model makes predictions on both the extensive (supplier choice) and the intensive (supplier shares) margin of supplier choice, which we test.

In our main empirical exercise, we use the equations derived from the model to estimate both the extensive and intensive margin decisions. Importantly, we observe directly the key cost variables, supplier price and distance between the two firms, so we can estimate the tradeoff in terms of choices between quality and these two key variables both for higher- and lower-quality buyers. This approach also allows us to control for a rich set of fixed effects and yields a discrete choice model for supplier choice.

Controlling for these factors, we estimate supplier choice by a conditional logit model and supplier shares by a two-way fixed effects model. Both approaches show evidence for complementarity. In particular, according to simple calculations, increasing supplier quality by 10% increases revenue by 2.7 percentage points more for 10% better client. In addition, we find that distance has a stronger effect for the extensive margin compared to the intensive margin, suggesting that search frictions, rather than only transportation costs, drive the trade-off between distance and quality. This trade-off is indeed substantial: a 10% increase in distance reduces choice probability as much as a 30% increase in price.

These findings corroborate existing evidence in the literature by highlighting the key role of quality complementarity in production (Verhoogen 2008, Kugler & Verhoogen 2012, Demir et al. 2021). Nevertheless, here we provide direct evidence on quality complementarity by modeling and estimating how quality and prices influence clients' choice of suppliers and the intensity of trade with them. To do that, we exploit the fact that in our data we can uniquely observe all important components of this choice—price, quality and distance—which allows us to estimate and separate the role of these factors from each other. Our paper also provides a number of novel insights into the interaction between supplier choice, quality and geography. We show that there are substantial search frictions in this network, generating a trade-off between geographic proximity of suppliers and match quality in terms of quality.

The remaining part of the paper is structured as follows. Section 2 introduces the data we use, Section 3 presents four facts that motivate our modeling approach. Section 4 describes our model and Section 5 discusses our estimates of the regressions derived from the model. Section 6 concludes.

2 Data

Our main database is based on the VAT declarations of Hungarian firms, submitted between 2015 and 2019. All firms that pay VAT are required to submit these declarations and need to report individually each relationship in which they have sold or purchased above a threshold within each calendar year. This threshold was reduced during the period under study, starting from value around HUF 4 mn (app. 12,000 EUR) in 2015 and declining to HUF 400 thousand (app. 1,200 EUR) by 2019.[1]

VAT declarations include the tax id of both the reporter (supplier) and the partner (buyer), the number and the net value of sales/purchases and the amount of VAT. As a result, these data allow us to see the universe of transactions between firms operating in Hungary with two limitations: i) we don't see very small links and ii) we don's see transactions between two small firms which don't pay VAT. All firm ids are anonymized, but can be linked to other firm-level data sources.

The tax id in the data also allows us to link the VAT information to the firms' financial statements, including balance sheet and Profit&Loss (P&L) statements. Balance sheet data is available for firms who conduct double-entry bookkeeping, practically all firms with employees. This provides an additional opportunity to clean the data by comparing firms' sales and material costs in the P&L to their sales and purchases in the VAT data. Even more importantly, we can calculate firms' average wage from the P&L data which we use as a proxy for quality (similarly to Demir et al. 2021).

[1] These values represent total transaction values in a year between the two firms.

We also merge the VAT data to the Register of Economic Organizations, a dataset including basic information on all economic organizations. This includes firms which do not conduct double-entry bookkeeping. Information includes location, legal forms, employees and turnover. Geographical data is also linked to the VAT information, allowing us to calculate the distance between each pair of firms.

Finally, we link these data to a production survey (PRODCOM) to calculate unit values at the firm level. This survey, harmonized across EU member states, samples manufacturing firms and asks the quantity and value produced from each of the firm's products. This allows us to calculate the firm's unit value, a proxy for price, from each of these products. We normalize log unit values with the weighted average unit value from each product in our sample. Finally, we take a sales-weight average of the firm-product level unit values to calculate a firm-level price.[2]

In its final form, the VAT panel data includes purchases of each firm from their suppliers between 2015 and 2019 (Table 1, Panel A). These five years include a total of 350,607 suppliers, 318,625 buyers, and 4,808,877 supplier-buyer pairs. Then, for our analysis, we must exclude all observations where key variables were missing, such as the reporter or the partner ID, the distance between the pair, industry or location code, or sales value. As a result, we had a total of 148,374 suppliers, 142,410 buyers, and 2,754,019 supplier-buyer pairs (Table 1, Panel B). Finally, it is worth reporting the number of observations which include the PRODCOM price, as this is a notable subsample due to the survey data it comes from. We can observe the price of 4,855 suppliers, 4,965 buyers, and 347,174 supplier-buyer pairs (Table 1, Panel C).

When reporting our results, we include all firms and industries (indicated by NACE codes) and focus on the year 2019.

[2]Note that the VAT data does not include product codes, so we cannot know which products are exchanged in each specific relationship, therefore we can only use this firm-level price.

3 Motivating Facts

In this Section we introduce four facts that will motivate the setup of our model and our main empirical analysis.

3.1 Fact 1: Higher quality buyers have higher quality suppliers on average

Our first question is about assortativity in the network. Are higher quality buyers more likely to buy from higher quality suppliers on average?

In order to answer this question, we regress buyer quality in average supplier quality:

$$\ln \omega_j = \beta_1 \overline{\ln \omega}_j^{I(j)} + nace_j + county_j + \epsilon_i \tag{1}$$

where i denotes buyers and j suppliers.

Our dependent variable is ω_j, supplier quality in 2019, proxied by average wage. Our main variable of interest is the (weighted) average buyer quality of supplier j calculated as $\overline{\ln \omega}_j^{I(j)} = \sum_{k \in I(j)} s_{kj} \ln \omega_k$, where $I(j)$ is the set of buyers buying from supplier j and s_{kj} is buyer k's share in supplier j's total purchases ($s_{kj} = T_{kj} / \sum_{l \in I(j)} T_{lj}$, where T_{kj} is the transaction amount from firm k purchased by j). In the regression above we also control for the location and industry of the suppliers by including county fixed effects ($county_j$), as well as 2-digit buyer industry fixed effects ($nace_j$).

Table 2 reports results from these regressions. Column (1) reports the main results for the whole sample. We find that 10% higher buyer wage is associated with 1.4% higher wage at supplier firms, showing positive quality assortativity in the network. In column (2) we exclude the Budapest, the capital of Hungary, and get very similar results. One possible concern is that the average buyer unit value reflects the industrial or geographic combination of buyers rather than their quality. Therefore, in columns (3) and (4) we first residualise buyers' wages by regressing buyer wage on country and industry (2-digit level) fixed effects and obtaining the residual and then we take the average of these residualized values. With this specification we get similar results.

To sum up, these results show strong positive assortativity between supplier and buyer quality.

3.2 Fact 2: Higher-quality buyers choose more distant suppliers if there are few higher-quality supplier around

Our second question is about whether higher quality buyers source their inputs from farther away. Naturally, sourcing distance depends on the availability of high-quality suppliers at the buyers location. Therefore, we are particularly interested in whether higher quality buyers are willing to source from farther away if only lower quality suppliers are available locally.

To investigate this question, we run buyer-supplier pair level regressions with distance as the dependent variable. The main variables of interest are buyer quality, the local average supplier quality and their interaction, which reflects whether higher quality buyers are more likely to source from far away if only lower quality buyers are available nearby:

$$\ln dist_{ij} = \beta_1 \ln \omega_i + \beta_2 \overline{\ln \omega}_{r(i)}^{s(j)} + \beta_3 \ln \omega_i \times \overline{\ln \omega}_{r(i)}^{s(j)} + \gamma \times X_{ij} + \epsilon_{ij} \tag{2}$$

where $dist_{ij}$ is the distance between buyer i and supplier j, $r(i)$ is the region of buyer i and $s(i)$ is the industry of supplier j. The local quality average in industry $s(i)$ and location $r(i)$ calculated as $\overline{\ln \omega}_{r(i)}^{s(j)} = \frac{\sum_{r(k)=r(i)\&s(k)=s(j)} \ln \omega_k}{\#r(k)=r(i)\&s(k)=s(j)}$. Finally, the X_{ij} term includes different sets of fixed effects.

The results are presented in Table 3. Column (1) includes fixed effects capturing the buyer's location at the NUTS4 level. Higher quality buyers tend to source from more distant suppliers. If local suppliers are higher quality on average, buyers are more likely to source from nearby suppliers.

Similarly, having more potential suppliers locally is associated with a lower average distance between buyers and suppliers. These findings suggest that the quality and variety of local supply affects sourcing decisions significantly.

Column (2) also includes the interaction between buyer quality and the average quality of local potential suppliers. The coefficient of this interaction is negative, showing that higher-quality buyers are more sensitive to local supplier quality, and choose more distant suppliers when there are few high quality suppliers nearby. This is consistent with higher willingness to pay in terms of distance for access to higher quality inputs.

Columns (3)-(4) also include buyer-supplier (2-digit) industry pair fixed effects to control for industry differences in the demand and supply for quality. The results remain largely unchanged. Finally, in columns (5)-(6), we also include buyer firm fixed effects (together with 2-digit supplier industry fixed effects). Therefore here we identify the effects primarily from comparing a suppliers' decisions to source its different inputs, depending on the local supply of each input. This specification controls for such factors as the remoteness of the buyer or its general input demand. The results are similar to the previous ones.

To sum up, Fact 2 is consistent with the presence of distance-related frictions in the matching process. Higher quality buyers are more sensitive to the local availability of high-quality inputs, suggesting a different trade-off between input quality and sourcing distance.

3.3 Fact 3: Higher-quality buyers source from farther away to find good quality suppliers

Our next fact investigates the trade-off between distance/transportation cost and supplier quality from a different angle. Here we ask the question: are higher quality firms more willing to go farther to find high quality suppliers?

Figure 2 shows the descriptive relationship between distance and supplier price for higher- and lower-quality buyers. There is a positive relationship for both groups, suggesting that firms are willing to source better quality inputs even if they have to look farther away. Also, higher quality buyers source higher quality inputs from each distance. This relationship is also flatter for higher-quality firms, suggesting that they are willing to look farther away for a higher quality.

We investigate this further with supplier-buyer pair level regressions where the supplier quality is the dependent variable and the main variables of interest are the distance between the two firms and its interaction with buyer quality. A negative interaction suggests a higher willingness to pay—in terms of distance—for higher input quality by higher quality suppliers.

In particular, we estimate the following regression:

$$\ln \omega_{j(i)} = \beta_1 \ln dist_{ij} + \beta_2 \ln dist_{ij} \times I[\omega_i > med_{s(i)}(\omega)] + \eta_{s(j)} + \epsilon_{ij} \tag{3}$$

where $I[\omega_i > med_{s(i)}(\omega)]$ equals to 1 if the buyer quality, ω_i, is larger than the median in the buyer's industry, $s(i)$, while $\eta_{s(j)}$ are industry fixed effects.

The results are in Table 4. Column (1), including buyer fixed effects, shows a positive correlation between distance and supplier quality, with an elasticity close to 5 percent. The positive elasticity suggests that firms, in general, face a tradeoff between higher transportation costs for higher quality. This elasticity is even larger when we exclude the capital city, Budapest (Column (2)).

Columns (3) and (4) include the interaction between buyer quality and distance. We find a lower elasticity for high quality buyers, suggesting that they are willing to source from farther away to find a supplier with a given quality level. A potential concern is that this relationship reflects different wage levels at different locations. In columns (5) and (6), as a robustness check, we also include an alternative interaction, where the buyer wage dummy is defined relative to the local median wage to capture these local wage differences. Including this interaction does not affect our main result, suggesting that quality, rather than local price levels drive this relationship.

This fact complements Fact 2 by providing evidence for the importance of distance-related frictions and that higher-quality buyers apply a different trade-off between these frictions and supplier quality.

3.4 Fact 4: Buyers with higher quality suppliers perform better

We investigate the relationship between supplier average supplier quality and different measures of buyer firm performance with the following regressions:

$$\ln y_{it} = \beta_1 \ln \omega_{i,2015} + \beta_2 \overline{\ln \omega}_{i,2015}^{J(i)} + \xi_t + county_i + \epsilon_{it} \tag{4}$$

where i denotes buyers, j suppliers, t years, $\ln y_{it}$ denotes buyer performance: sales, profit, value added and these variables divided by the number of workers. In the regression we control for the quality of the buyer, $\ln \omega_{i,2015}$, and we are interested whether the (weighted) average quality of suppliers, $\overline{\ln \omega}_{i,2015}^{J(i)}$, is associated with increase in firms' performance. The weighted average is calculated as follows: $\overline{\ln \omega}_{i,2015}^{J(i)} = \sum_{k \in J(i)} s_{ik,2015} \ln \omega_{j,2015}$, where $J(i)$ is the set of firms supplying firm i and $s_{ik,2015}$ is the share of supplier j-s sell in buyer i total purchases in 2015, formally, $s_{ik,2015} = T_{ik,2015} / \sum_{l \in J(i)} T_{il,2015}$, where $T_{ik,2015}$ is the transaction value between buyer i and supplier k. Furthermore, ξ_t are year effects while $county_i$ are buyer county effects.

The results from these regressions are reported in Table 5. The table shows the all the outcomes considered—sales, profit, value added—are positively associated both with buyer and supplier quality (columns (1), (3), (5)). The elasticity of profits with respect to supplier quality is, for example 0.14. This relationship is only partly driven by firm size: average supplier quality is also positively associated with per-worker measures, even though the elasticities are lower (columns (2), (4) and (6)). Buyers with higher quality suppliers tend to perform better.

4 Model

This section describes our model and Appendix A includes the model derivations. The main aim of our model is to guide our empirical work by providing microfoundations for our main regressions.

Our model incorporates two main features based on the motivating facts. First, it includes a complementarity between supplier and buyer quality. Second, buyers face a trade-off between the quality of suppliers and their distance. We capture these two features in a partial equilibrium model of supplier choice.

4.1 Setup

Demand for the product of firm i is

$$Q_i = A h_i^{\sigma-1} P_i^{-\sigma} \tag{5}$$

so that it depends on both price P_i and quality h_i.

The quantity of output of firm i is

$$Q_i = a_i \cdot L_i^{1-\alpha} \cdot Q_{im}^{\alpha} \tag{6}$$

where a_i is the productivity of firm i and Q_{im} is the quantity of composite input

$$Q_{im} = \left[\sum_{j \text{ supplies } i} (\phi_{ij} Q_{ij})^{1-1/\sigma} \right]^{\frac{\sigma}{\sigma-1}}. \tag{7}$$

Here Q_{ij} is the quantity i purchases from j, and ϕ_{ij} is the importance of the input from the industry of j to producing the output in the industry of i.

A firm i in industry I needs one supplier each in industries J_I.

Firm i chooses one supplier j in each of several exogenously specified industries. Conditional on choosing a set of suppliers S_i, one from each industry $J \in J_I$, output quality is a product of three terms:

$$h_i = \underbrace{\bar{\omega}_i{}^{1-\alpha}}_{\substack{\text{Client}\\ \text{quality}}} \cdot \underbrace{\left[\sum_{j \in S_i} \left(\frac{\phi_{ij} Q_{ij}}{Q_{im}} \cdot \bar{\omega}_{ij}\right)^{1-1/\sigma}\right]^{\frac{\sigma}{\sigma-1}\alpha}}_{\text{Supplier and match quality}} \cdot \underbrace{\left[\prod_{j \in S_i} \left(d_{ij}^{-\gamma_r} e^{\varepsilon_{ij}}\right)^{\beta}\right]}_{\text{Search friction}}. \tag{8}$$

The first term, $\bar{\omega}_i$, measures effective client firm quality and is defined through

$$\log \bar{\omega}_i = \theta_c \log \omega_i \tag{9}$$

where ω_i is our observable measure of client quality.

The second term captures supplier and match quality. This is modeled as the quantity-share $(\phi_{ij} Q_{ij}/Q_{im})^{1-1/\sigma}$ weighted CES average of the effective qualities $\bar{\omega}_{ij}$ of suppliers $j \in S_i$, where

$$\log \bar{\omega}_{ij} = \theta_s \log \omega_j + \theta \log \omega_i \log \omega_j.$$

In this formulation, effective supplier quality $\bar{\omega}_{ij}$ matters more if a higher quantity of input Q_{ij} is purchased from supplier j. θ is the key parameter, measuring the complementarity between supplier and client quality.

The third term captures search frictions. This depends on distance and an error term with standard Gumbel distribution ε_{ij}. The search friction in our specification does not depend on the share of input j and hence can be thought of as a cost in terms of managerial time, with β measuring its importance. Intuitively, having more distant suppliers means that the manager needs to spend more time on them to maintain a good relationship, reducing the time available for managing the production process at the factory and reducing output quality.

Suppose further that the weight of input j in production is

$$\phi_{ij} = \phi_{ij}^0 d_{ij}^{-\gamma_t} e^{\nu_{ij}}. \tag{10}$$

Here γ_t corresponds to an iceberg trade cost, so that distance affects both the fixed cost and the variable cost of purchasing inputs from j. And ν_{ij} is an error term independent of everything, mean zero, and not known at the time of choosing suppliers.

In our empirical exercise we focus on measuring complementarity and the effect of distance. In the model, complementarity is captured by θ as part of the 'supplier and match quality' term in the quality production function. Distance affects client outcomes via two margins: search (γ_r) and trade (γ_t).

4.2 Solution

Under these conditions, the client firm solves the following problem:

$$\max_{S_i, Q_{ij}, L_i, P_i} P_i \cdot Q_i(P_i, h_i) - \omega_i L_i - \sum_{j \in S_i} P_j Q_{ij} \tag{11}$$

subject to constraints (5)-(10), in particular: quality h_i is determined by choice of suppliers S_i, and demand $Q_i(P_i, h_i)$ is determined by price and quality. The challenge when solving this problem is that firm has dual objectives: maximize quality and minimize cost. In particular, input Q_{ij} affects both quantity and quality. In our setup these simplify into single objective: to minimize cost of a different firm with input-quality-weighted production function.[3]

[3]Because composite input Q_{im} is in the numerator of quantity production and the denominator of quality production.

We solve the model in Appendix A. We derive two predictions. First, buyers face the following discrete choice problem when sourcing from industry J:

$$j = \arg\max_{j \in J} \left(\frac{\bar{s}_{IJ}\theta_s}{\beta} \log \omega_j + \frac{\bar{s}_{IJ}\theta}{\beta} \log \omega_i \cdot \log \omega_j - \frac{\bar{s}_{IJ}}{\beta} \cdot \log P_j - \left(\frac{\bar{s}_{IJ}\gamma_t}{\beta} + \gamma_r \right) \log d_{ij} + \varepsilon_{ij} \right). \quad (12)$$

The model yields two key predictions for this equation. First, in this equation quality complementarity ($\theta > 0$) increases choice probabilities. Second, distance has a negative effect both because of the search costs (γ_r) and trade costs (γ_t).

This is a microfoundation for our estimating equation for supplier choice, or the extensive margin. Note that several terms involve $\bar{s}_{IJ}$ which is a measure of the importance of input J for industry I. Intuitively, for a less important input, quality and price matter less relative to the supplier-specific search cost shock. The coefficient of the distance variable has two terms one of which does, the other does not depend on $\bar{s}_{IJ}$. This is because distance affects both the search cost, which does not depend on the input's importance, and the transportation cost, which does.

The second prediction of the model is that the shares of each supplier (intensive margin) among all suppliers will be the following:

$$\log s_{ij} = (\sigma - 1)(\log \phi^0_{ij} + \log \tilde{P}_{im} - \log P_j + \theta_s \log \omega_j + \theta \log \omega_i \log \omega_j - \gamma_t \log d_{ij} + \nu_{ij}). \quad (13)$$

There are two important consequences of this equation. Fist, quality complementarity increases input share. Second, distance has a negative effect on input shares, but only because of trade costs. As a result, comparing the effect of distance in the choice and share equations informs us about the magnitude of search costs.

Note that here ϕ^0_{ij} only depends on I and J. Furthermore, here we do not have the γ_r and ε because those shifters only affect the choice of supplier. But we have ν_{ij}, because that shock is realized after the suppliers are chosen but before the quantities are chosen.

5 Results

5.1 Extensive margin: Supplier choice

Here we are interested in supplier choice. The microfoundation for this estimation is Equation 12. All theoretical variables ω_j, ω_i, P_j and d_{ij} are observable.

To investigate this, for each buyer i and input p (defined as the 2-digit industry of supplier) we define a potential supplier set within a given year and 2-digit supplier industry code. To make sure that the results are not driven by suppliers with little relevance, we drop all industries where the supplier industries sales constitute less than 5% of the buyer industry's total purchases. We define our dependent variable as the dummy, $supplies_{ipj}$, taking the value one if i buys product p from firm j and taking the value zero for all other firms in the choice set. To capture the heterogeneity of choice sets, especially in terms of the number of potential suppliers, we include choice-set fixed effects (buyer firm-supplier industry) in our regressions. We run two different models for supplier choice: a conditional logit and a linear model. While the conditional logit model can capture the nature of the choice problem better, the linear model allows us to include supplier fixed effects besides the choice set fixed effects, controlling for unobserved supplier characteristics.

In about half of choice sets, the buyer purchases from multiple potential suppliers. This may be explained by many factors such as buying multiple different products from the same industry or sourcing from different suppliers for different plants. We report results both for all choice sets and also restricting our sample to choice sets when the buyer sources only from one supplier.

In particular, we run the following regressions:

$$\begin{aligned} supplies_{ij} = \quad & F(\beta_1 \ln \omega_j + \beta_2 \ln \omega_i \times \ln \omega_j + \beta_3 \ln dist_{ij} \\ + \quad & \beta_4 \ln P_j + \zeta_{is(j)} + \epsilon_{ij}) \end{aligned} \tag{14}$$

where i denotes buyers, j potential suppliers, $supplies_{ij}$ is an indicator variable showing whether j supplier i. $dist_{ij}$ is the distance between i and j, ω_i is buyer quality, ω_j is supplier quality. P_j is the price of supplier j, only when supplier fixed effects are not included. $\zeta_{is(j)}$ is choice set fixed effects, representing a buyer firm-supplier industry combination, while $F()$ is either conditional logit or linear model.

Importantly, referring back to Equation (12), a positive coefficient of the quality interaction (β_2) suggests positive complementarity, given that $\bar{s}_{IJ}$ and β are positive. Also, dividing the coefficient of the interaction with the coefficient of price, $\frac{\beta_2}{\beta_4}$, is equal to the structural complementarity coefficient, θ.

Table 6 shows the results of these regressions. Column (1) presents the results from the conditional logit regressions. Distance has a negative coefficient, while supplier wage a positive one, reflecting that buyers are more likely to buy from nearby and higher quality firms. The interaction of buyer and supplier quality is positive and significant showing that higher supplier quality increases the probability of sourcing more when the buyer is also a high quality producer. The supplier price in this specification comes from PRODCOM and it has a negative coefficient, as expected.

As we have discussed, the PRODCOM price can only be observed for a subset of suppliers, reducing the number of observations by 80%. Such strong selection can lead to biased results. To alleviate this issue, we predict the price for all firms by running a regression on the PRODCOM sample with log price as the dependent variable and average wage, productivity, size as well as region and industry dummies on the right hand size. We predict the price for all suppliers based on this regression. In Column (2) we include this predicted price as the Prodcom price is available only for a subsample of firms. This increases the sample size. The coefficients remain quite similar. Finally, in Column (3) we run a linear model where we can include supplier fixed effects in addition to the choice set fixed effects to control for unobserved supplier characteristics, including price. The interaction is positive and highly significant in this specification.

Columns (4)-(6) repeat the same exercise but focus only on cases when the buyer chooses only one supplier from the choice set. The results remain similar.

To sum up, the evidence presented here suggests that higher quality buyers more likely to choose a higher quality suppliers than a lower one.

5.2 Intensive margin: Supplier shares

Here we are interested in understanding on how much firms buy from each of their suppliers (intensive margin). We regress the share of supplier j in buyer i's total purchases, and our main variable of interest is the interaction between buyer and supplier quality, showing whether higher quality buyers buy more from their higher quality suppliers. In some of the specifications we also control for the supplier price.

Our estimating equation is derived from Equation (13). P_j, ω_i, ω_j and d_{ij} are observable in our data, while ϕ_{ij}^0 and $\tilde{P}_{im}$ are not. The former is buyer and supplier-specific while the latter is buyer specific. We control for the latter two unobservables by including buyer and supplier fixed effects.

In particular, we run the following regressions:

$$\ln s_{ij} = \beta_1 \ln \omega_j + \beta_2 \ln \omega_i \times \ln \omega_j + \beta_3 \ln dist_{ij} + \beta_4 \ln P_j + \eta_i + \xi_j + \zeta_{k(i)s(j)} + \epsilon_{ij} \tag{15}$$

where i denotes buyers, j suppliers, s_{ij} is the share of supplier j in buyer i's purchases and $dist_{ij}$ is the distance between i and j. ω_i is buyer quality, ω_j is supplier quality. P_j is the price of supplier j , only when supplier fixed effects are not included. η_i, ξ_j and $\zeta_{k(i)s(j)}$ are buyer, supplier and buyer industry-supplier industry pair fixed effects, respectively.

Importantly, to clean the data from inputs which are relevant for only few firms in an industry, we use the following procedure. First, for each buyer industry k, we calculate the share of each supplier industry's s share in k's purchases, $s_{ks} = \frac{\sum_{i \in k, j \in s} v_{ij}}{\sum_{i \in k} \sum_j v_{ij}}$. Then, we drop all observations when $s_{ks} < 5\%$.

Referring back to Equation (13), the coefficient of the interaction informs us about the strength of complementarity between between supplier and buyer quality, θ. In particular, a positive coefficient for the interaction suggests complementarity. Also, note that the ratio of the interaction and price coefficients equals the theoretical parameter θ.

The results of these regressions are in Table 7. Column (1) includes buyer fixed effects, identifying the coefficients by comparing suppliers' share within the same buyer. The positive coefficient of supplier quality shows that the buyer with average quality buys a higher share from higher quality suppliers (both quality measures are normalized with the average). The interaction of buyer and supplier quality is positive showing that higher quality buyers buy a higher share from higher quality suppliers compared to lower quality buyers. Finally, distance, as expected, has a negative coefficient reflecting that firms are less likely to source from farther away, ceteris paribus.

Supplier quality is likely to be positively correlated with supplier price. In columns (2)-(5) we attempt to disentangle the role of supplier price and quality in this decisions. In column (2) we control for the supplier's price, calculated from PRODCOM. Recall that PRODCOM is a survey, therefore supplier price is available only for a subset of suppliers, reducing our sample size to 8,200 observations. As expected, we find that supplier price is negatively correlated with the supplier's share. The coefficients of supplier wage and the interaction increase substantially, showing that quality conditional on price has an even higher effect.

A concern with these regressions is that quality and price may be correlated with supplier size, and larger suppliers have higher share in their buyers' purchases. To investigate this possibility, in column (4) we also control for supplier sales and employment. The main results does not change. Column (5) includes the predicted price measure to estimate on a larger sample, and yields similar results.

Another way to control for unobserved supplier characteristics, including price in case of firms not sampled by the Prodcom is to run a two-way fixed effects model by including both buyer and supplier fixed effects. Column (6) reports the results of these regressions while (7) also includes supplier industry-buyer industry fixed effects. The interaction remains positive and significant.

The results show that high quality buyers buy more from higher-quality suppliers than lower-quality buyers.

6 Discussion

Figure 3 illustrates our results from the choice and share regressions. Both specifications suggest the presence of complementarity between supplier and buyer quality, with higher-quality buyers valuing supplier quality more both in terms of their supplier choice and the supplier shares. As discussed, we can estimate the degree of complementarity, θ, by dividing the interaction cofficient with the distance coefficient in both equations. The different specifications yields θ estimates between 1.3-6 (Table 8).

The model allows us to quantify the complementarity in terms of log revenue. We can approximate the revenue in the following way:[4]

$$\log \text{Revenue} \approx \alpha(\sigma-1)\theta \log \omega_i \cdot \overline{\log \omega_j} + f_i + f_j$$

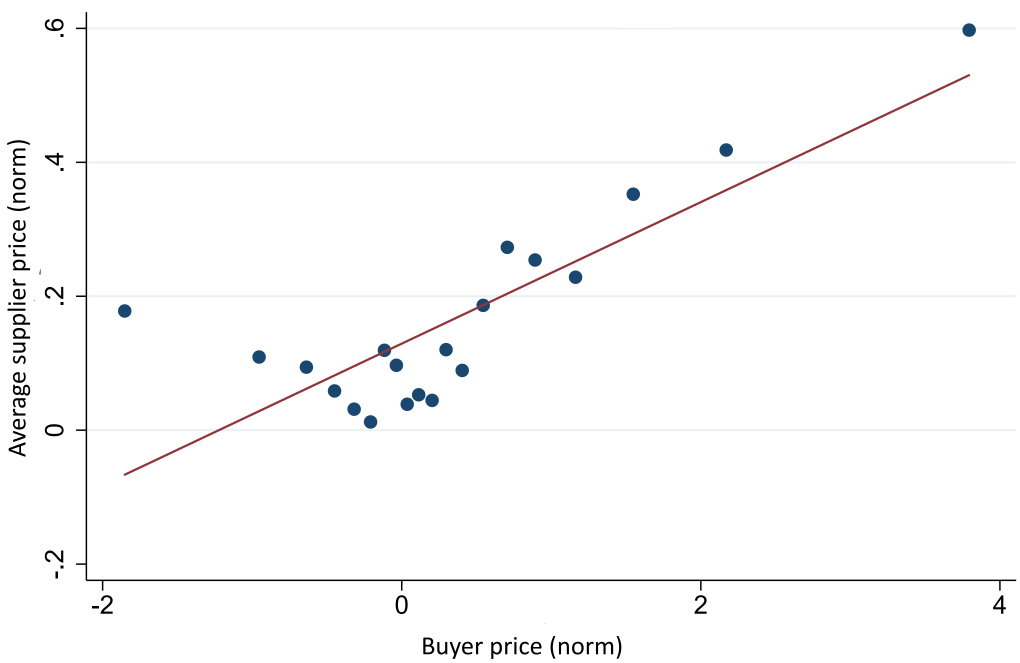
[4] By plugging in Equation (8) into Equation (5)

where $\overline{\log \omega_j}$ is average supplier quality. Using $\sigma = 3$ and $\alpha = 0.7$ as well as the estimates for θ from above, increasing supplier quality by 10% increases revenue by 2.7-12.6% percentage points more for a 10% higher quality client.

We can also quantify the importance of search costs. By comparing the main predictions, Equation (12) and (13), distance affects choice via both search and trade costs, while it only affects share via trade costs. Therefore, the difference between the distance coefficients in the choice and share equations represent the effect of search costs. Based on these comparisons, a 10% increase in distance reduces choice probability as much as a 30% increase in price.

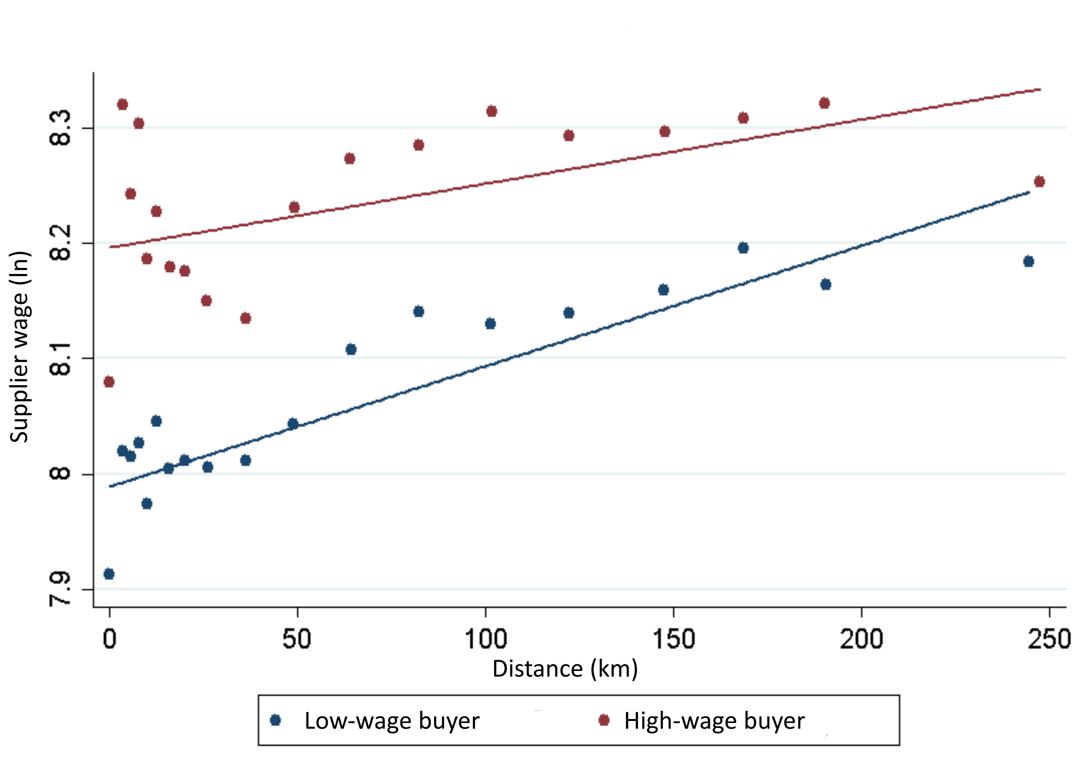
7 Figures

Figure 1: Relationship between buyer price and average supplier price



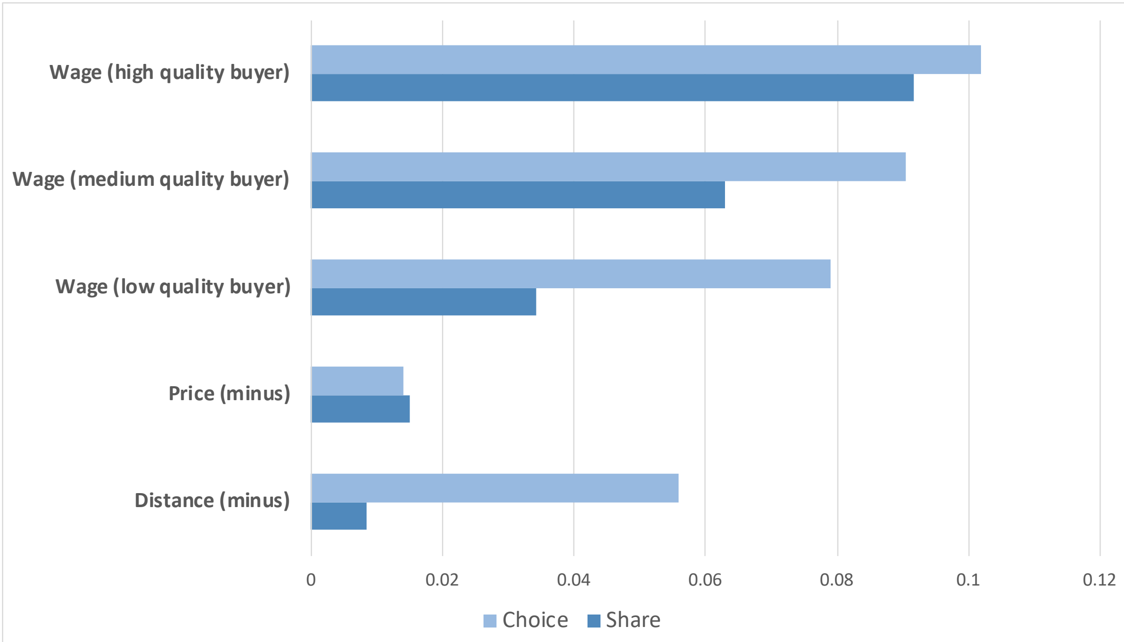
Note: This figure depict the relationship (binscatter) between buyer normalized prices (from Prodcom, normalized by 4-digit product average unit value) and the average normalized price of each buyer's suppliers. 2019.

Figure 2: Relationship between buyer price and average supplier price



Note: This figure shows the relationship (binscatter) between supplier distance and supplier wage for lower- and higher-quality buyers, for supplier-buyer pairs. 2019.

Figure 3: Summary of the estimated extansive and intensive margin effects



Note: The figure is based on Table 7 col (3) and Table 6 col (1). High quality buyer's average wage is 50% higher than the average and low quality buyers' is 50% lower. In case of share, these are effects on share, scaled between 0 and 1, while in case of choice we report are conditional logit coefficients.

Tables

Table 1: Number of observations

Panel A: Observations in VAT

	Year				
	2015	2016	2017	2018	2019
Number of suppliers	101 444	102 839	127 435	235 673	243 839
Number of buyers	96 602	98 427	124 226	196 406	207 847
Number of supplier-buyer pairs	423 387	426 674	536 626	1 519 442	1 902 748

Panel B: Observations where key variables are not missing

	Year				
	2015	2016	2017	2018	2019
Number of suppliers	61 007	60 781	56 470	106 633	108 834
Number of buyers	53 642	54 451	48 839	100 723	105 921
Number of supplier-buyer pairs	260 506	257 672	246 676	884 887	1 104 278

Panel C: Observations with price information

	Year				
	2015	2016	2017	2018	2019
Number of suppliers	3 187	3 323	3 088	3 607	3 608
Number of buyers	3 198	3 314	3 089	3 843	3 791
Number of supplier-buyer pairs	38 240	38 915	36 849	107 387	125 783

Note: Panel A includes all suppliers, buyers, and pairwise combinations which can be found in the VAT panel data. Panel B includes all suppliers, buyers, and pairwise combinations in the VAT panel data, excluding all observations where key variables were missing (reporter ID, partner ID, distance, NACE code, NUTS code, sales). Panel C includes all suppliers, buyers, and pairwise combinations in the VAT panel data, excluding observations without a PRODCOM price.

Table 2: Higher quality buyers have higher quality suppliers on average

	(1) supplier ln wage	(2) supplier ln wage	(3) supplier ln wage	(4) supplier ln wage
Av. buyer ln wage (norm.)	0.141***	0.111***		
	(0.00468)	(0.00583)		
Av. residual buyer ln wage (norm.)			0.123***	0.0950***
			(0.00426)	(0.00523)
Observations	32,444	22,662	32,444	22,662
R-squared	0.090	0.088	0.049	0.042
Budapest excluded		x		x

Note: These supplier-level regressions show how supplier quality is associated with average buyer quality. All regressions include county- and 2-digit industry fixed effects. The main explanatory variable in columns (1)-(2) is the average ln average wage of the buyers while in columns (3)-(4) it is residualized wage after filtering out county and industry fixed effects from buyer wages before calculating the average buyer wage.

Table 3: High quality buyers choose more distant suppliers if there are few high-quality supplier around

	(1) distance (ln)	(2) distance (ln)	(3) distance (ln)	(4) distance (ln)	(5) distance (ln)	(6) distance (ln)
Buyer wage (ln)	0.214***	1.079***	0.142***	0.971***		
	(0.00634)	(0.118)	(0.0153)	(0.185)		
Local average potential supplier wage (ln)	-0.241***	-0.236***	-0.268***	-0.265***	-0.257***	-0.248***
	(0.00953)	(0.00955)	(0.0165)	(0.0164)	(0.0276)	(0.0285)
Number of potential local suppliers (ln)	-0.168***	-0.168***	-0.262***	-0.262***	-0.218***	-0.218***
	(0.00297)	(0.00297)	(0.00942)	(0.00939)	(0.0171)	(0.0171)
Buyer wage (Ln) X Local average potential supplier wage (ln)		-0.113***		-0.108***		-0.0823***
		(0.0154)		(0.0240)		(0.0295)
Observations	226,676	226,676	226,514	226,514	217,461	217,461
R-squared	0.056	0.057	0.115	0.115	0.255	0.255
buyer NUTS4 FE	YES	YES	YES	YES	YES	YES
Industry pair FE			x	x		
Buyer FE					x	x
Supplier ind FE			x	x	x	x

Note: These buyer-supplier pair-level regressions show how supplier quality is associated with buyer quality. The dependent variable is the ln distance between the supplier and the buyer and the independent variables are the ln local wage in the supplier's industry, ln number of firms in the supplier's industry at the same location to the buyer, and the interaction of buyer wage and average wage of firms in the supplier's industry. Local is considered at the NUTS4 level. Columns (1)-(2) include fixed effects capturing the buyer's location. Columns (3)-(4) include additional buyer-supplier industry pair fixed effect. Columns (5)-(6) also include buyer firm and 2-digit supplier industry fixed effects. standard errors are clustered at the buyer-supplier industry level. The regressions are unweighted, and the sample includes all buyer-supplier pairs in year 2019.

Table 4: High quality buyers go farther away to find good quality suppliers

	(1) supplier wage (ln)	(2) supplier wage (ln)	(3) supplier wage (ln)	(4) supplier wage (ln)	(5) supplier wage (ln)	(6) supplier wage (ln)
distance (ln)	0.0429***	0.0753***	0.0603***	0.0792***	0.0600***	0.0784***
	(0.00162)	(0.00122)	(0.00198)	(0.00205)	(0.00203)	(0.00211)
buyer wage>median X dist. (ln)			-0.0220***	-0.00500**	-0.0253***	-0.0118**
			(0.00278)	(0.00251)	(0.00492)	(0.00477)
buyer wage>local med. X dist. (ln)					0.00370	0.00781
					(0.00493)	(0.00479)
Observations	294,358	204,159	294,358	204,159	294,358	204,159
R-squared	0.205	0.220	0.205	0.220	0.205	0.220
Budapest buyers excluded		x		x		x

Note: These buyer-supplier pair-level regressions investigate how supplier quality is associated with the distance from the buyer and its interaction with buyer quality. The dependent variable is ln supplier wage (demeaned) and the independent variables are the ln distance between the pair, and buyer quality, defined as a higher-than-median wage. All specifications include buyer fixed effects and standard errors are clustered at at the buyer firm level. All regressions are unweighted, and the sample includes all buyer-supplier pairs in year 2019. In columns (2)-(4)-(6), Budapest buyers are excluded.

Table 5: Firm performance is positively correlated with average supplier quality

	(1) Buyer's Profit (ln)	(2) Buyer's Profit/worker (ln)	(3) Buyer's Sales (ln)	(4) Buyer's Sales/worker (ln)	(5) Buyer's Value added (ln)	(6) Buyer's Value added/worker (ln)
Buyer wage (ln)	1.308***	0.463***	1.302***	0.489***	1.553***	0.721***
	(0.0171)	(0.0129)	(0.0151)	(0.0105)	(0.0151)	(0.00865)
Av. Supplier wage (ln)	0.140***	0.0611***	0.112***	0.0328***	0.117***	0.0369***
	(0.00307)	(0.00256)	(0.00235)	(0.00163)	(0.00262)	(0.00150)
Observations	90,208	90,208	104,552	104,552	100,193	100,193
R-squared	0.211	0.063	0.284	0.119	0.314	0.201

Note: These buyer-year-level regressions show the relationship between buyer performance and the average quality of its suppliers, including observations between 2015-2019. All regressions include year and buyer county fixed effects and standard errors are clustered at the buyer level. The wages are calculated from the relevant financial statements. Independent variables are the the demeaned buyer and supplier wages of the given year. The dependent variables are the following: ln buyer profit in column (1); ln buyer profit per worker in column (2), calculated as the previous profit per size, ln buyer's sales in column (3); ln buyer's sales per worker in column (4); ln buyer value added in colunm (5), calculated as the difference between sales and material; and ln buyer value added per worker in column (6). All regressions are unweighted.

Table 6: Extensive margin: Supplier choice

	(1) supplies	(2) supplies	(3) supplies	(4) supplies	(5) supplies	(6) supplies
distance (ln)	-0.558***	-0.626***	-0.0119***	-0.690***	-0.756***	-0.00843***
	(0.0107)	(0.00760)	(0.000169)	(0.0162)	(0.0104)	(0.000140)
supplier wage (ln)	0.905***	0.362***		0.916***	0.356***	
	(0.0306)	(0.0156)		(0.0470)	(0.0245)	
supplier wage (ln) X buyer wage (ln)	0.228***	0.138***	0.00329***	0.423***	0.193***	0.00111***
	(0.0610)	(0.0232)	(0.000181)	(0.0950)	(0.0413)	(0.000102)
price (PRODCOM)	-0.141***			-0.150***		
	(0.0122)			(0.0201)		
predicted price		-0.112***			-0.185***	
		(0.0137)			(0.0251)	
Observations	184,662	1,829,638	6,799,315	91,967	793,305	4,157,856
Estimator	clogit	clogit	linear	clogit	clogit	linear
choice set FE	x	x	x	x	x	x
Supplier FE			x			x
Sample	all	all	all	one supplier	one supplier	one supplier

Note: In these regressions the sample consists of pairs of buyers and their potential suppliers (defined as firms in the same industry from which the buyer purchases at least one input). The binary dependent variable shows whether the buyer purchases from the potential supplier or not. All regressions are unweighted and include choice set fixed effects. Standard errors are clustered at the buyer firm level. Column (1) includes the ln distance between the buyer and the potential supplier, ln supplier wage as a proxy for quality, and the interaction between ln buyer wage and ln supplier wage as independent variables. Column (2) follows the same logic, but substitutes the predicted price for the PRODCOM price to increase sample size. Column (3) only includes the ln distance and the interaction as independent variables, but also incorporates supplier firm fixed effects. Columns (4)-(6) are robustness checks following the same logic, but only including buyer-industry combinations from which the buyer chose only one supplier.

Table 7: Intensive margin: Supplier share

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\ln s_{ij}$	$\ln s_{ij}$	$\ln s_{ij}$	$\ln s_{ij}$	$\ln s_{ij}$	$\ln s_{ij}$	$\ln s_{ij}$
supp. wage (ln, norm.)	0.276***	0.731***	0.630***	0.0645	0.290***		
	(0.0150)	(0.0942)	(0.0951)	(0.109)	(0.0278)		
supp. wage X buyer wage	0.246***	0.800***	0.573***	0.562***	0.256***	0.288***	0.276***
	(0.0283)	(0.125)	(0.122)	(0.122)	(0.0304)	(0.0362)	(0.0365)
distance (ln)	-0.0581***	-0.0678***	-0.0836***	-0.0990***	-0.0601***	-0.0874***	-0.0885***
	(0.00633)	(0.0225)	(0.0218)	(0.0221)	(0.00675)	(0.00926)	(0.00925)
Supplier price (ln) - Prodcom		-0.140***	-0.152***	-0.110***			
		(0.0380)	(0.0371)	(0.0355)			
supplier sales (ln)				0.284***			
				(0.0574)			
Supplier employment (ln)				-0.0957			
				(0.0703)			
predicted price (ln)					-0.108***		
					(0.0195)		
Observations	48,301	8,212	8,212	8,212	46,315	35,653	35,605
R-squared	0.550	0.515	0.557	0.565	0.552	0.721	0.725
buyer FE	x	x	x	x	x	x	x
suppllier FE						x	x
supplier ind FE			x	x	x	na	na
industry pair FE							x
larger than 5%	x	x	x	x	x	x	x

Note: In these supplier-buyer level regressions the dependent variable is the share of the supplier in the buyer's total purchases in the VAT data. In column (1), standard errors are clustered at the buyer firm level, in columns (2)-(4), standard errors are clustered at the supplier firm level, and in columns (6)-(7), standard errors are clustered at the supplier-buyer pair level. In column (1), the independent variables are the ln distance between the supplier and buyer, the ln, normalized price of the supplier's wage, and the interaction between the supplier's and the buyer's wage (both ln and normalized) and buyer fixed effects are included. Column (2) additionally includes the ln price of the supplier, and this causes the sample size to drop (PRODCOM data). Column (3) adds supplier industry fixed effects (at the 4-digit level), and column (4) also includes ln supplier sales and ln supplier employment. Column (5) uses the predicted price instead of the PRODCOM price, leading to a larger sample size. Column (6) includes two-way supplier-buyer pair fixed effects, and column (7) includes additional two-way supplier-buyer industry fixed effects (at the 2-digit level). The sample includes supplier-buyer pairs from 2019, where the supplier's industry share exceeds 5%, where industry share is calculated as the supplier's sales value to the total sales value of the industry.

Table 8: Estimated θ values

	Extensive	Intensive
Prodcom price	6	3.75
Predicted price	1.3	2.5

Note: this table shows the estimated vales for the complementarity parameter, θ, from the different specifications reported in Tables 6 and 7. In particular, we estimate θ by dividing the quality interaction coefficient with the distance coefficient in Table 6 columns (1) and (2) and Table 7 columns (3) and (5).

References

Bernard, A. B., Dhyne, E., Magerman, G., Manova, K. & Moxnes, A. (2019), The origins of firm heterogeneity: A production network approach, Technical report, National Bureau of Economic Research.

Bernard, A. B. & Moxnes, A. (2018), 'Networks and trade', Annual Review of Economics **10**, 65–85.

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

Bernard, A. B., Moxnes, A. & Ulltveit-Moe, K. H. (2018), 'Two-sided heterogeneity and trade', Review of Economics and Statistics **100**(3), 424–439.

Demir, B., Fieler, A. C., Xu, D. & Yang, K. K. (2021), O-ring production networks, Technical report, National Bureau of Economic Research.

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

Kikkawa, A. K., Magerman, G. & Dhyne, E. (2019), 'Imperfect competition in firm-to-firm trade', Available at SSRN 3389836 .

Kugler, M. & Verhoogen, E. (2012), 'Prices, plant size, and product quality', The Review of Economic Studies **79**(1), 307–339.

Marshall, A. (1890), 'Principles of economics, 8th edn (1920)', London, Mcmillan .

Miyauchi, Y. et al. (2018), Matching and agglomeration: Theory and evidence from japanese firm-to-firm trade, Technical report, working paper.

Spray, J. (2017), 'Reorganise, replace or expand? the role of the supply-chain in first-time exporting'.

Verhoogen, E. A. (2008), 'Trade, quality upgrading, and wage inequality in the mexican manufacturing sector', The Quarterly Journal of Economics **123**(2), 489–530.

Appendix A Model derivations

Here we discuss the solutions of our model described in Section 4.

Demand is

$$Q_i = A h_i^{\sigma-1} P_i^{-\sigma}.$$

Because quality h_i is homogenous of degree zero in the input quantities, once the relative input quantities are chosen, increasing scale only affects Q_i but not h_i. Thus, fixing the relative input quantities, standard arguments apply. In particular, denoting the total cost by C_i, cost minimization implies $(1-\alpha)C_i = L_i\omega_i$ and $\alpha C_i = \sum_{j\in J} P_j Q_{ij}$. And the constant markup rule applies: $P_i = \sigma/(\sigma-1)C_i/Q_i$.

These observations imply that revenue can be written as

$$R_i = A\left(\frac{h_i}{P_i}\right)^{\sigma-1} = A\left(\frac{h_i Q_i}{C_i}\right)^{\sigma-1}\cdot\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}$$

and profits as

$$\Pi_i = A\left(\frac{h_i Q_i}{C_i}\right)^{\sigma-1}\cdot\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\frac{1}{\sigma}.$$

Solving the firm's problem is equivalent to choosing Q_{ij} and L_i to maximize this expression of profits. The key term is $h_i Q_i/C_i$. Using the definition of h_i,

$$\begin{aligned}
\frac{h_i Q_i}{C_i} &= \frac{\bar{\omega_i}^{1-\alpha}\cdot\left[\sum_{j\in J}\left(\frac{\phi_{ij}Q_{ij}}{Q_{im}}\cdot\bar{\omega}_{ij}\right)^{1-1/\sigma}\right]^{\frac{\sigma}{\sigma-1}\alpha}\left[\prod_{j\in S_i}\left(d_{ij}^{-\gamma_r}e^{\varepsilon_{ij}}\right)^{\beta}\right]a_i\cdot L_i^{1-\alpha}\cdot Q_{im}^{\alpha}}{C_i}\\
&= \frac{\bar{\omega_i}^{1-\alpha}\cdot\left[\sum_{j\in J}\left(\phi_{ij}Q_{ij}\bar{\omega}_{ij}\right)^{1-1/\sigma}\right]^{\frac{\sigma}{\sigma-1}\alpha}\left[\prod_{j\in S_i}\left(d_{ij}^{-\gamma_r}e^{\varepsilon_{ij}}\right)^{\beta}\right]a_i\cdot L_i^{1-\alpha}}{C_i}\\
&= \frac{\bar{\omega_i}^{1-\alpha}\cdot\tilde{Q}_{im}^{\alpha}\cdot a_i\cdot L_i^{1-\alpha}}{C_i}\left[\prod_{j\in S_i}\left(d_{ij}^{-\gamma_r}e^{\varepsilon_{ij}}\right)^{\beta}\right]
\end{aligned}$$

where we used the notation that

$$\tilde{Q}_{im} = \left[\sum_{jinJ} (\phi_{ij} \bar{\omega}_{ij} Q_{ij})^{1-1/\sigma} \right]^{\frac{\sigma}{\sigma-1}}$$

which, like Q_{im}, is a CES aggregate of the input quantities Q_{ij}, but with different weights $\phi_{ij}\bar{\omega}_{ij}$. This observation implies that the original firm problem with endogenous quality is equivalent to the problem of linked firm with exogenous quality. In particular, consider now a hypothetical firm with production function

$$\tilde{Q}_i = a_i L_i^{1-\alpha} \tilde{Q}_{im}^{\alpha}$$

and exogenous quality

$$\tilde{h}_i = \bar{\omega}_i{}^{1-\alpha} \left[\prod_{j \in S_i} \left(d_{ij}^{-\gamma_r} e^{\varepsilon_{ij}} \right)^{\beta} \right].$$

The above derivation then implies

$$\frac{h_i Q_i}{C_i} = \frac{\tilde{h}_i \tilde{Q}_i}{C_i}$$

because the cost of inputs is the same. It follows that the profits of the original firm are equal to the profits of the hypothetical firm. Moreover, because solving the original firm's problem is equivalent to choosing Q_{ij} and L_i to maximize our expression for profits, which in turn equals an analogous expression for profits for the hypothetical firm as a function of the same input choices, the input choices of the hypothetical firm will equal the input choices of the actual firm.

We now proceed by solving the problem of the hypothetical firm. Define the price index for the inputs of this hypothetical firm to be

$$\tilde{P}_{im}^{1-\sigma} = \sum_j (\phi_{ij} \bar{\omega}_{ij})^{\sigma-1} P_j^{1-\sigma}$$

then the spending share on input j in the optimum will satisfy

$$s_{ij} = \frac{P_j Q_{ij}}{P_{im} Q_{im}} = \frac{P_j Q_{ij}}{\tilde{P}_{im} \tilde{Q}_{im}} = \frac{P_j{}^{1-\sigma} (\phi_{ij} \bar{\omega}_{ij})^{\sigma-1}}{\tilde{P}_{im}^{1-\sigma}}$$

where at the second equality we used that input prices and quantities are the same for the actual and the hypothetical firm.

Marginal cost for the hypothetical firm is proportional to $1/a_i$ times Cobb-Douglas composite of the input price index and the wage, that is,

$$\frac{C_i}{\tilde{Q}_{im}} \propto \frac{1}{a_i} \cdot \omega_i^{1-\alpha} \tilde{P}_{im}^{\alpha}.$$

It follows that profits are proportional to

$$\left(\left(\sum_j (\phi_{ij} \bar{\omega}_{ij})^{\sigma-1} P_j^{1-\sigma} \right)^{\frac{\alpha}{\sigma-1}} \cdot a_i \left(\frac{\bar{\omega}_i}{\omega_i} \right)^{1-\alpha} \prod_{j \in S_i} \left(d_{ij}^{-\gamma_r} e^{\varepsilon_{ij}} \right)^{\beta} \right)^{\sigma-1}.$$

Separating the terms that only depend on i, log profits are

$$\log \text{Profit}_i = f_i + \alpha \log \left(\sum_{j \in S_i} (\phi_{ij} \bar{\omega}_{ij})^{\sigma-1} P_j^{1-\sigma} \right) - (\sigma - 1)\beta\gamma_r \log d_{ij} + (\sigma - 1)\beta\varepsilon_{ij}.$$

We take the following first order approximation

$$\begin{aligned} \log \sum_{j \in S_i} \left(\frac{\phi_{ij} \bar{\omega}_{ij}}{P_j} \right)^{1-\sigma} \approx & \log \sum_{j \in S_i} \left(\frac{\hat{\phi}_{ij} \hat{\bar{\omega}}_j}{\hat{P}_{ij}} \right)^{1-\sigma} \\ & + \sum_{j \in S_i} \frac{(\hat{\phi}_{ij} \hat{\bar{\omega}}_j / \hat{P}_{ij})^{\sigma-1}}{\sum_j (\hat{\phi}_{ij} \hat{\bar{\omega}}_{ij} / \hat{P}_{ij})^{\sigma-1}} \cdot (\sigma - 1) \cdot (\log \phi_{ij} - \hat{\phi}_{ij} + \log \bar{\omega}_{ij} - \log \hat{\omega}_{ij} - \log P_j + \log \hat{P}_{ij}) \\ = & \text{const} + \sum_{j \in S_i} \bar{s}_{IJ} \cdot (\sigma - 1) \cdot (\log \phi_{ij} + \log \bar{\omega}_{ij} - \log P_j). \end{aligned}$$

A few remarks are in order. The constant is fixed for possible suppliers in the industry J of supplier j. The underlying logic is that when client i is choosing a supplier in the industry of j, she has an expectation of what the final prices, qualities and importance weights for all its suppliers are going to be: these are the hatted variables. The constant term is determined by these expectations. When contemplating different suppliers in the industry of j, the elasticity is the share of that supplier at the expected (hatted) values. We assume that this share equals the industry share, which amounts to taking the first order approximation at hatted variables which reproduce the industry shares.

Maximizing profits is thus equivalent to maximizing

$$\sum_{j \in S_i} \bar{s}_{IJ} \cdot (\log \phi_{ij} + \log \bar{\omega}_{ij} - \log P_j) - \beta \gamma_r \log d_{ij} + \beta \varepsilon_{ij}$$

and substituting in for $\bar{\omega}$ and for ϕ this is equivalent to maximizing

$$\sum_{j \in S_i} \bar{s}_{IJ} \cdot (\log \phi^0_{ij} - \gamma_t \cdot \log d_{ij} + \nu_{ij} + \theta_s \log \omega_j + \theta \log \omega_i \log \omega_j - \log P_j) - \beta \gamma_r \log d_{ij} + \beta \varepsilon_{ij}.$$

This is separable in j, and, because ν_{ij} is unknown and has mean zero, implies the following discrete choice problem among suppliers in the industry of j:

$$j = \arg\max_{j \in J} \left(\frac{\bar{s}_{IJ} \theta_s}{\beta} \log \omega_j + \frac{\bar{s}_{IJ} \theta}{\beta} \log \omega_i \cdot \log \omega_j - \frac{\bar{s}_{IJ}}{\beta} \cdot \log P_j - \left(\frac{\bar{s}_{IJ} \gamma_t}{\beta} + \gamma_r \right) \log d_{ij} + \varepsilon_{ij} \right).$$

This is a microfoundation for our estimating equation. Note that several terms involve $\bar{s}_{IJ}$ which is a measure of the importance of input J for industry I. Intuitively, for a less important input, quality and price matter less relative to the supplier-specific search cost shock. The coefficient of the distance variable has two terms one of which does, the other does not depend on $\bar{s}_{IJ}$. This is because distance affects both the search cost, which does not depend on the input's importance, and the transportation cost, which does.

Taking logs of the share equation yields

$$\log s_{ij} = (\sigma - 1)(\log \phi^0_{ij} + \log \tilde{P}_{im} - \log P_j + \theta_s \log \omega_j + \theta \log \omega_i \log \omega_j - \gamma_t \log d_{ij} + \nu_{ij}).$$

Note that here ϕ^0_{ij} only depends on I and J. Here we do not have the γ_r and ε because those shifters only affect the choice of supplier. But we have ν_{ij}, because that shock is realized after the suppliers are chosen but before the quantities are chosen.

Appendix B Cleaning the VAT data

In this section we describe the steps we undertake when cleaning the VAT data. We conduct two types of cleaning. The first (one-sided) considers each record in isolation and cleans records. The second step compares the same transaction as reported by the buyer and the seller and makes the data symmetric (two-sided).

One issue is that some firms supplied annual while others monthly or quarterly declarations, depending on the total amount of VAT they paid. To make the data comparable across firms, handle seasonality and be able to link the VAT data to annual financial data, we aggregate the VAT records to the annual level for each buyer-supplier pair.

B.1 One-sided cleaning

During the processing of the raw data, we found the following reporting issues:

(1) Negative values. These only occurred in 0.8% of the records, and probably represent 'corrections' of transactions reported in the previous period. Since these occurred only when firms provided monthly or quarterly reports, the observations were kept as long as the yearly sums remained positive.

(2) Zero reported value with multiple invoices: this can come from redacted invoices and only occurred in very few cases. These were dropped.

(3) The reported value was below the threshold of reporting for 0.25% of the records. While these likely represent actual transactions, we dropped these to apply the same threshold for all firms.

Some smaller, minor issues were also identified, such as an overly large tax rate (above 30%), reporting transactions the same buyer and supplier, and duplicate observations. Such instances were very rare, representing less than 0.1% of transactions and were dropped.

B.2 Two-sided cleaning

A key feature of the VAT data is that all transactions are reported both by the buyer and the seller. The tax id allows us to link declarations between buyers and suppliers and to identify the two records of the same transaction. We rely extensively on comparing these records when we clean the data. Based on these comparisons, we make two types of imputations: (i) extensive corrections, e.g. including a positive value when partner reports a positive value but the reporter's value was missing; or (ii) intensive corrections, when both firms report different positive values.

Extensive imputation cleans records when one firm did not report a transaction while the other did so. This is the situation in 30% of transactions, even though these tend to be small. In these cases we found it plausible that the transaction actually took place but one party did not report it by mistake, therefore we input the positive value.

Intensive imputation is used when the two firms report different positive values, and we assume that the firm reporting the smaller amount is likely to have forgotten to report some of the transactions. Here we consider a hierarchical correction where we accept their larger yearly reported value in a given firm-pair. When creating the hierarchy, we consider which reported values seem the most reliable. The 'hierarchy' goes as follows:

(1) net and tax value and transaction number based on larger net value. Such discrepancies are likely to be from the monthly extensive margin, leading to unreported tax values.

(2) if the net value is the same, we consider tax value and transaction number based on larger tax value. These cases are rare and typically come from rounding errors.

(3) if both the net and tax values are the same, we consider the larger transaction number. Such issues probably come from submitted corrections, which makes the number of invoices less reliable in the data.

We also compare the overall amount of transactions reported in the VAT data to firm's sales costs from the financial statements. Our expectation is that the VAT transactions should be below the balance sheet quantities because of thresholds and international trade. We find that this is the case in 96% of cases.

Note that at the end of this procedure we get a database which includes each transaction twice, once reported by the buyer and once by the seller. We keep only one of these rows.