ISSN 2042-2695

CENTRE *for* ECONOMIC PERFORMANCE

CEP Discussion Paper No 1435

June 2016

Production Networks, Geography and Firm Performance

Andrew B. Bernard
Andreas Moxnes
Yukiko U. Saito

Abstract
This paper examines the importance of buyer-supplier relationships, geography and the structure of the production network in firm performance. We develop a simple model where firms can outsource tasks and search for suppliers in different locations. Low search and outsourcing costs lead firms to search more and find better suppliers. This in turn drives down the firm's marginal production costs. We test the theory by exploiting the opening of a high-speed (Shinkansen) train line in Japan which lowered the cost of passenger travel but left shipping costs unchanged. Using an exhaustive dataset on firms' buyer-seller linkages, we find significant improvements in firm performance as well as creation of new buyer-seller links, consistent with the model.

Keywords: production networks, trade, productivity, infrastructure
JEL Classifications: F14; D22; D85; L10; L14; R12

This paper was produced as part of the Centre's Trade Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

Acknowledgements
Thanks go to Adam Kleinbaum, Yaniv Dover, and Jonathan Dingel as well as seminar participants at CEPR ERWIT, DINR, NOITS, NBER ITI, RIETI, World Bank and EITI-Phuket for helpful comments. We gratefully acknowledge the financial support from the Japan Society for the Promotion of Science (No. 25780181). We thank the Center for Spatial Information Science, University of Tokyo (CSIS) for their address matching service. This study is conducted as a part of the Project "Inter-organizational and Inter-inventors Geographical Proximity and Networks" undertaken at the Research Institute of Economy, Trade and Industry (RIETI).

Andrew Bernard, Tuck School of Business and Centre for Economic Performance, London School of Economics. Andreas Moxnes, University of Oslo, CEPR and NBER. Yukiko Saito, Research Institute of Economy, Trade and Industry, Tokyo, Japan.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

1 Introduction

In spite of the widespread perception that firms' success in part depends on their connections with suppliers, relatively little work has been done on the structure, performance and importance of production networks. Even less is known about how geography and trade costs affect links in production networks. Finally, in spite of a large literature on the role of infrastructure on economic outcomes, there is almost no evidence on how infrastructure affects supply chains and firm-level productivity. This paper examines the importance of buyer-supplier relationships and the structure of the production network in firm performance.

While there has been an explosion of research on social and economic networks and their formation, to date little of that work has considered the supplier-customer relations between firms. In addition, existing studies are often limited to a particular industry or geography within a country.[1] In this paper we use a comprehensive, unique data set on the production network in Japan. Our data provide supplier-customer links between firms for over 950,000 firms in Japan. This set of firms accounts for the large majority of private sector economic activity in the country. For the large majority of firms in Japan, we can determine their location, suppliers, customers and measures of performance.

We develop a set of stylized facts about the Japanese production network to guide our model. Large and productive firms have more suppliers than small firms. Geographic proximity plays an important role in the matching of suppliers and customers. Most connections are local; the median distance to a supplier is 30 kilometers. Larger firms not only have more suppliers, but, on average, have suppliers that are farther away. The production network displays negative degree assortivity; the trading partners of well-connected firms on average are less-well connected themselves. Consider two firms, one with many suppliers, the other with few. The suppliers to the well-connected firm have on average relatively few customers. The suppliers to the less-connected firm have on average many customers. Many of these facts are also present in cross-border trade networks, e.g. negative degree assortivity is also found in exporter-importer networks in international trade, see Bernard et al. (2013).

We build a parsimonious model of a domestic economy motivated by the stylized facts. Downstream firms require a continuum of tasks as inputs into the production process, e.g. materials processing, accounting, printing, and distribution services. They can produce the tasks themselves or outsource them. Finding suppliers is costly, however, and therefore it may not be profitable for all firms to outsource a given task, even if the market price of a task is lower than the firm's marginal cost of supplying the same task. Our model is closely related to the international sourcing framework in Antràs et al. (2014), but we modify it to allow for the possibility that firms can supply

[1] See, for example, the seminal work of Uzzi (1996)

a given task within the boundary of the firm. Downstream firms can observe broad characteristics of potential upstream locations, i.e. average productivity and trade costs, but need to expend resources to observe the prices of individual tasks in a location. In addition, outsourcing is costly because of trade costs. Trade costs are here broadly defined; we have in mind both shipping costs and efficiency losses in the buyer-supplier relationship. In equilibrium, a higher efficiency firm will search across more locations, source more inputs, and have better performance. If variable trade costs or the fixed costs of search fall, firms will search more, source more inputs from more distant locations, and firm sales will rise. These effects will be larger in input-intensive industries where the marginal benefit of finding better suppliers is greater. For the aggregate economy, locations with low trade and search costs will have higher performing firms, even if productivity is ex ante identical across all locations. Our framework therefore offers a supply-side microfoundation for why measured productivity varies widely across locations, as documented in Sveikauskas (1975), Glaeser and Maré (2001) and Combes et al. (2012).

To examine the predictions of the model we use the 2004 opening of the southern portion of the high-speed rail lines in Japan (Kyushu Shinkansen) as a quasi-natural experiment. The route of this particular extension had been planned at least since 1973 but the actual construction was subject to substantial timing uncertainty due to numerous budgetary and administrative delays, thus limiting the scope for anticipation effects. We examine whether firms near new Shinkansen stations improved their performance after the opening. Estimating a triple difference specification, we find that performance was better for firms near the new stations after the opening and that firms in industries with greater purchased input shares performed better compared to firms in industries with lower purchased input shares.

The model suggests that the firm-level performance improvement is due to the increased number of suppliers and sourcing locations. We draw on a second cross-section of the Japanese production network in 2010 to examine whether firms in localities near the new stations increase their number of suppliers and the number of source locations more than firms in localities that did not become better connected with the high-speed rail extension. The results show support for the mechanisms emphasized in the model; the number of connections and the number of source locations both increase for firms near the new stations.

This paper is naturally related to a growing literature on the determinants of domestic and foreign sourcing and the impact on firms. Amiti and Konings (2007), Goldberg et al. (2010), Halpern et al. (2011) and Bøler et al. (2014) examine the role of imported inputs in firm productivity where foreign and domestic inputs are imperfect substitutes. Our work is closer to Antràs et al. (2014) who develop and structurally estimate a model where firm performance is positively related to the intensive and extensive margins of purchased imported inputs. In the domestic production network, we find systematic relationships between distance to domestic suppliers and firm performance that are analogous to those in the international trade context. In this regard our work is related to Fort (2014) who finds an important role for firm heterogeneity and location in the decision to use domestic contract manufacturing services.

The paper is also related to a large literature on the effects of infrastructure on economic development. Governments typically allocate a large fraction of their budgets to infrastructure projects and multilateral institutions similarly emphasize infrastructure in the expenditure allocation. Most research on the effects of infrastructure concentrates on the location of economic activity, income and aggregate welfare effects. For example, Donaldson (forthcoming) examines the effects of railroads on income and welfare in India, while Duranton et al. (2013) consider the effects of interstate highways on the level and composition of trade for US cities. Redding and Turner (forthcoming) survey the literature on the effects of infrastructure on economic activity. This area of research focuses on the role of infrastructure in reducing transport time and costs for goods between cities and in reducing the travel time for individuals within a city, i.e. commuting time. Our research points to another role for infrastructure in reducing travel time for individuals (as opposed to goods) between regions and the resulting firm-level improvements coming from the supply chain.

Another related strand of recent work studies the geography of knowledge transmission across locations. Davis and Dingel (2012) model costly idea exchange as the agglomeration force in a system of cities. Our framework focuses on the cost of connecting to others (firms) and the resulting improvements in performance. Cristea (2011) considers the importance of face-to-face meetings in international trade and finds that increased exports raises the demand for business class air travel. Comin et al. (2012) study technology diffusion over time and find that technology diffuses more slowly to locations that are farther away from technology leaders. Keller and Yeaple (2013) measure the cross-country spatial barriers to the transmission of embodied or disembodied knowledge. They find that person-to-person communication costs increase in distance. Hillberry and Hummels (2008) examine trade in intermediate goods as an explanation for highly localized shipments in the U.S.

Our work is also related to Giroud (2013) who examines the effect of new airline connections on within-firm performance of and investment in manufacturing plants. Related work in finance argues that proximity matters for monitoring and relationships.[2] In contrast to his study which examines reductions in travel costs between headquarters and plants for multi-plant firms, we broaden the scope by exploring all buyer-supplier connections among all firms in the economy. Moreover, our model and empirical strategy emphasize the creation and destruction of linkages in response to infrastructure shocks.

In the literature on firm-to-firm connections, Oberfield (2013) develops a network theory of search and production where producers potentially sell to many customers but have only one supplier. Downstream firms consider match-specific productivity and price when choosing among available

[2]See Lerner (1995), Coval and Moskowitz (1999), Coval and Moskowitz (2001) and Petersen and Rajan (2002).

techniques. As in our model, the share of purchased inputs matters for the propagation of shocks in the economy although our focus is on the supplier side rather than the downstream links. Acemoglu et al. (2012) relate these types of microeconomic shocks to aggregate fluctuations in a model of sectoral input-output linkages, while Carvalho et al. (2014) use the Japanese production network to study the supply chain disruptions occurring in the aftermath of the 2011 earthquake in Japan.

While our focus is on buyer-supplier matches in the domestic supply network, it is closely related to the nascent literature using matched importer-exporter data. Bernard et al. (2013) consider exporter-importer connections using Norwegian transaction trade data. They find, as we do, negative assortivity in buyer-seller matches, and in-degree and out-degree distributions that largely follow power laws. Blum et al. (2012) examine characteristics of trade transactions for the exporter-importer pairs of Chile-Colombia and Argentina-Chile and also find that small suppliers (exporters) typically sell to large (importers) and small importers (buyers) source from large suppliers (exporters).

The rest of the paper is structured as follow. We describe the data in Section 2 and develop a set of stylized facts about buyer-supplier relationships in Section 3. In Section 4, we develop our multi-location model of domestic sourcing. We describe and estimate our natural experiment along with various robustness checks in Section 5 and provide concluding remarks in Section 6.

2 Data

The data employed in this paper comes from two main sources. First, production network data for two moments in time, 2005 and 2010, are assembled by Tokyo Shoko Research, LTD. (TSR). TSR is a credit reporting agency and firms provide information to TSR in the course of obtaining credit reports on potential suppliers and customers or when attempting to qualify as a supplier. The resulting database contains information on more than 950,000 firms in each cross-section, represents more than half of all the firms in Japan and covers all sectors of the economy. The TSR sample is close to the full population of firms with more than 4 employees.[3]

Each firm provides rank-ordered lists of the most important suppliers (up to 24) and customers (24). TSR also collects information on employment, the number of establishments, the number of factories, up to three (4-digit) industries, sales, profits and a physical address. In addition, the database records TSR's credit score for the firm. Using an address matching service provided by the Center for Spatial Information Science at the University of Tokyo, we are able to match a firm's address to longitude and latitude data.[4] We use the geo-coded data to create a measure of great circle distance between firms. The top 3 prefectures by counts of firms are Tokyo, Osaka and Aichi

[3]Firms with 1 to 4 employees are underrepresented in TSR compared to Census data, while for firms with 5 or more employees, the firm size distribution in TSR is very similar to the distribution in Census data.

[4]As each firm only reports one address, the geographic information for multi-establishment firms is likely to reflect the location of the headquarters.

(Nagoya) while the top three 2-digit industries by counts are General Construction Work, Specialist Construction Work and Equipment Installation.

Second, firm-level balance sheet data comes from Kigyou Katsudou Kihon Chousa Houkokusho (The Results of the Basic Survey of Japanese Business Structure and Activities), henceforth Kikatsu, for the period 1998 to 2008. Kikatsu is an annual survey that gives detailed information about firm activities such as sales, employment, capital stock, intermediate purchases and industry affiliation. It covers the full population of manufacturing and non-manufacturing firms with more than 50 employees and with capital of more than 30 million yen.

2.1 Supplier and Customer Connections

The TSR data has both advantages and disadvantages relative to other production network data sets. Among the advantages is the inclusion of firms of all sizes and industries including both publicly listed and unlisted firms. In addition, the TSR firms self-report their most important suppliers and customers; there is no cutoff in terms of sales or purchases.[5] However, the 24-firm limit for suppliers and customers potentially causes a truncation in the number of relationships in the self-reported data relative to the actual number of such connections.

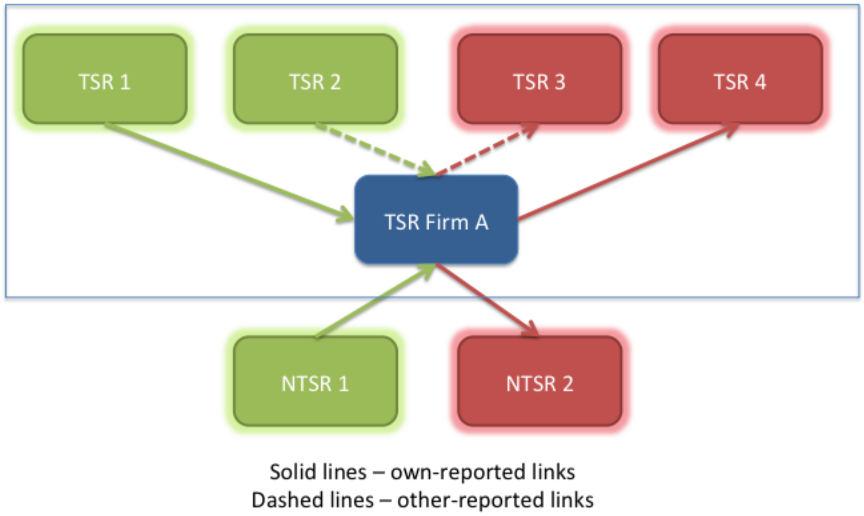
To mitigate this issue, we combine both self-reported and other-reported information for each firm in the data and use the union of own-reported and other-reported information. For firms A and B, we consider A to be a supplier of B if both firms are in the TSR data *and* either (i) A reports B as customer *or* (ii) B reports A as supplier. Note that some firms that are reported as suppliers and customers are outside the TSR set of firms (NTSR), i.e. they are domestic Japanese firms but are not customers or clients of TSR.

In Figure 1 we show possible suppliers and customers for a firm (Firm A) in the TSR database. Firm A reports that it has two customers, TSR4 and NTSR2, and two suppliers, TSR1 and NTSR1. Other firms also report connections to Firm A: TSR2 reports Firm A as a customer while TSR3 reports Firm A as a supplier. In determining Firm A's in-degree, the number of suppliers, and its out-degree, the number of customers, we ignore the NTSR links and include both own-reported and other-reported connections. Thus, Firm A has an in-degree of 2 (TSR1 and TSR2) and an out-degree of 2 (TSR3 and TSR4).

The 24-firm limit will be binding for very large firms and therefore most of their customers and suppliers are other-reported. One may suspect that this implies that large-to-large linkages are underreported, because both firms may not report the other. Recall, however, that firms provide a *rank-ordered* list of their connections; connections to large firms are therefore likely to end up high on the list. Moreover, the degree distributions reported in the next Section (Figure 2) do not

[5] In their analysis of US production networks, Atalay et al. (2011) use Compustat data on publicly listed firms and their major customers defined as firms that purchase more than 10 percent of the seller's revenue.

Figure 1: Supplier and Customer Connections: An Example



indicate any discontinuity around the 24-firm threshold.

A number of firms report either no suppliers and/or no customers among the TSR firms. This does not mean they recorded no suppliers or customers on their forms but instead all their reported connections are outside the TSR set of firms. A report of no TSR suppliers or no TSR customers might occur for several reasons. A firm might appear to have no TSR customers because all the domestic firms that are customers are outside the TSR database, all its customers are foreign firms or all its customers are non-firms, e.g. the public or government. A firm might appear to have no TSR suppliers because all the domestic firms that are suppliers are outside the TSR database or all its suppliers are foreign.[6] We choose to work only with the set of TSR firms with a positive in-degree or positive out-degree (links to other TSR firms) and find no evidence of systematic bias in the the sample of firms with positive TSR degree. Using NTSR+TSR data, the distribution of firms with TSR degree equal zero is virtually identical to the overall sample of firms, i.e. the mean and variance of NTSR+TSR out-degree and in-degree distributions are the same.

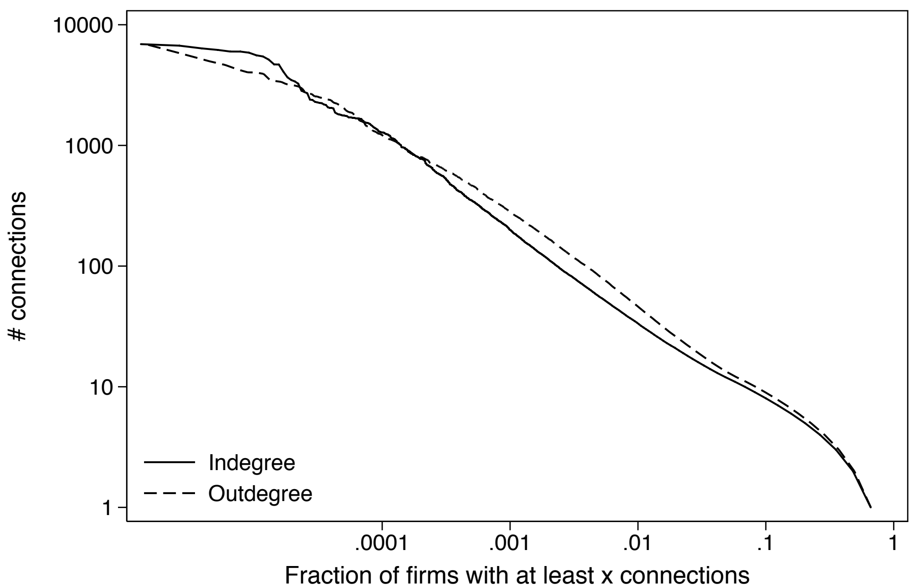
3 The Production Network

In this section we begin to explore the domestic production network in Japan.[7] There are 961,318 firms (nodes) in the TSR production network with 3,783,711 supplier-customer connections (directed edges). Of those nodes, 771,107 (676,320) nodes have positive in-degree (out-degree) among TSR firms. For firms with positive in-degree, the mean number of suppliers is 4.9 and the median is 2.

[6]It seems implausible to imagine that an operating firm has no actual domestic suppliers. This is supported by the fact that more TSR firms report no customers than report no suppliers.

[7]All descriptive statistics refer to the 2005 cross-section. Some of these network characteristics are also presented in Saito et al. (2007) and Ohnishi et al. (2010).

Figure 2: In-degree and Out-degree CDFs



Note: 2005 data, log scales.

For firms with positive out-degree, the mean number of customers is 5.6 and the median is one.

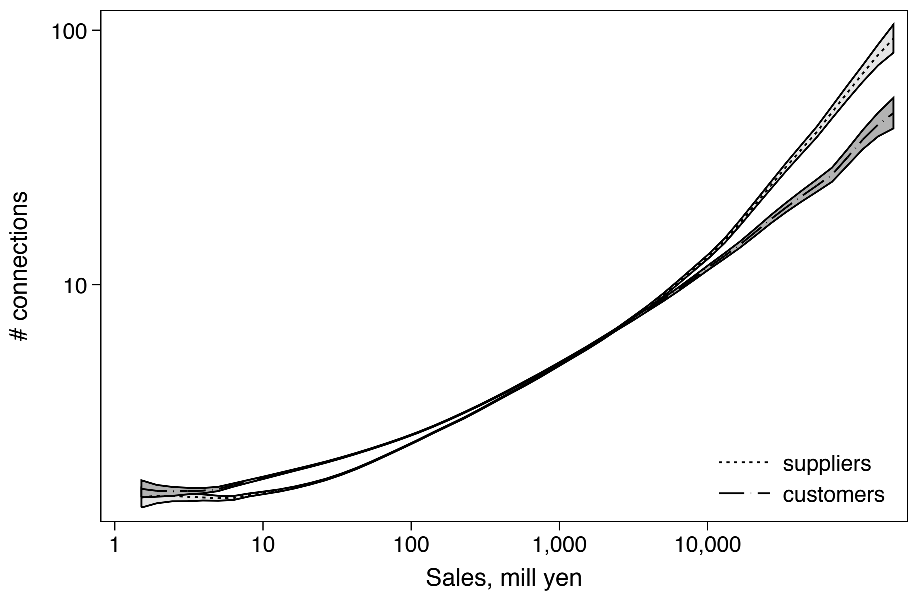
The cdfs of the in-degree and out-degree distributions are given in Figure 2. The distributions are well-approximated by a Pareto (power law) distribution. The estimated Pareto shape parameter is -1.32 for the in-degree distribution and -1.50 for the out-degree distribution. Deviations from the Pareto are found in the extreme tails of the distribution. Firms with a very large number of connections are somewhat under-represented while firms with few connections appear in greater numbers. These deviations from a power law distribution are comparable to those found in exporter-importer degree distributions by Bernard et al. (2013) but are much smaller in magnitude compared to those found by Atalay et al. (2011) for supplier-customer connections derived from data on large US firms and their large customers.

3.1 Stylized Facts

In this section, we document four facts from the data that will guide the development of the model in Section 4. We explore the relationship between firm characteristics, connections in the production network and geography.

Fact 1: Larger firms have more suppliers. Higher sales are associated with a larger number of supplier connections. Figure 3 plots the kernel-weighted local polynomial regression of a firm's in-degree (vertical axis) on sales (horizontal axis), both in logs. The linear regression slope is 0.36,

Figure 3: Size, in-degree and out-degree.



Note: 2005 data. The figure shows the kernel-weighted local polynomial regression of firm-level log degree (vertical axis) on log sales (horizontal axis). The two lines represent in-degree and out-degree as separate regressions. Gray area denotes the 95 percent confidence bands. Sample is first trimmed by excluding the 0.1 percent lowest and highest observations of sales.

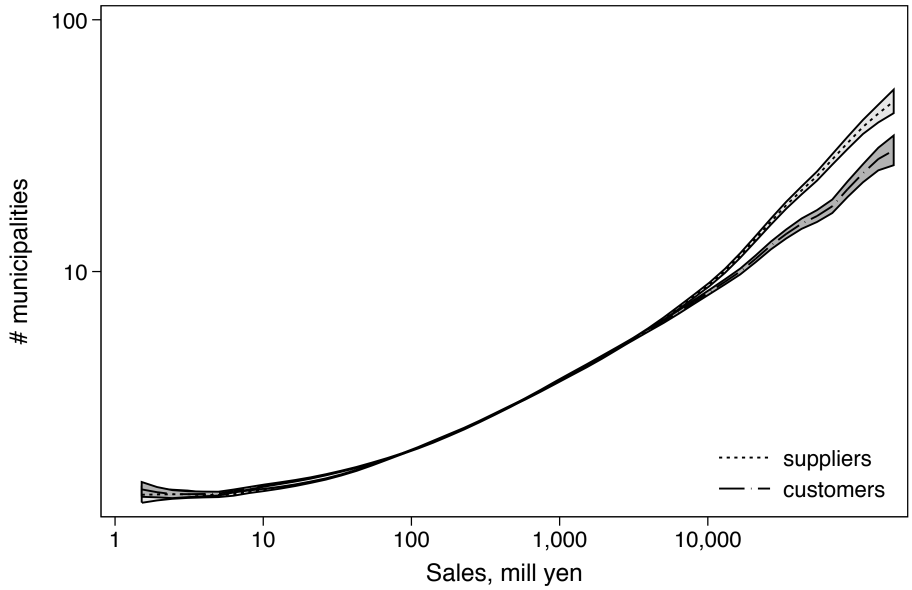
meaning that a 10 percent increase in sales is associated with a 3.6 percent increase in the number of suppliers. A similar positive relationship exists between a firm's sales and out-degree, mirroring the findings in Bernard et al. (2013).

Fact 2. Larger firms have suppliers in more locations and their distance to suppliers is higher. Figure 4 shows that larger firms tend to have suppliers in more municipalities. A firm in the 1st decile of the sales distribution has suppliers in 1.5 locations while a firm in the 9th decile has suppliers in roughly 4 locations.[8] At the same time, larger firms have more remote connections; Figure 5 plots the fitted values from a kernel-weighted local polynomial regression of a firm's median distance to its suppliers on its sales (both in logs). The median distance to suppliers is around 20 km for firms in the 1st decile of the sales distribution, while median distance is roughly 50 percent higher (32 km) for firms in the 9th decile of the sales distribution. A similar positive relationship also exists between a firm's sales and median distance to its customers, as well as between distance and the number of municipalities a firm is supplying.

We also compare buyers that have matched to the *same* supplier. The same pattern arises here; the distance to the supplier is increasing in the performance of the customer. Table 1 reports

[8]There are in total 1410 municipalities in our dataset, see also Section 5.1.

Figure 4: Size and number of supplier & customer locations.



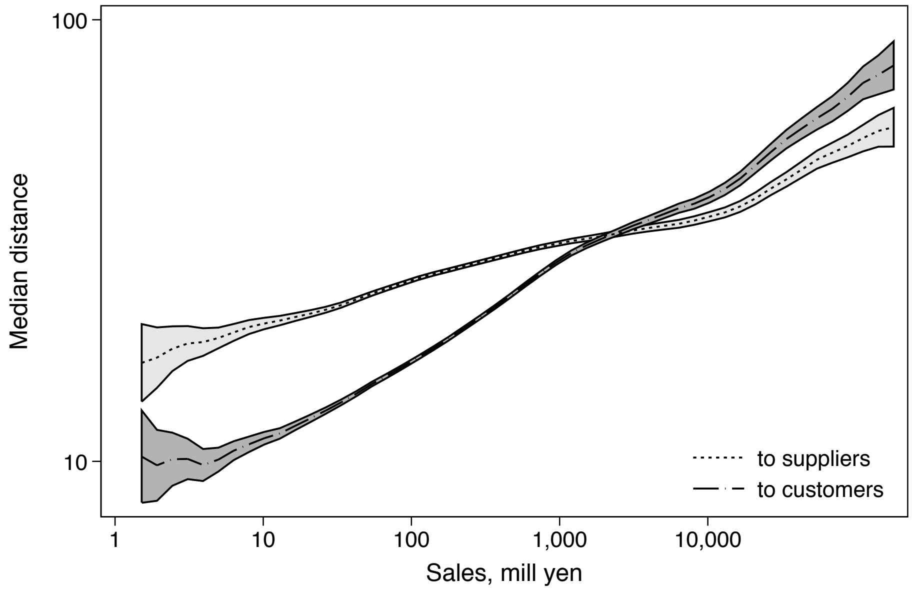
Note: 2005 data. The figure shows the kernel-weighted local polynomial regression of firm-level log number of municipalities with connections (vertical axis) on log sales (horizontal axis). The two lines represent the supplier and customer side as separate regressions. Gray area denotes the 95 percent confidence bands. Sample is first trimmed by excluding the 0.1 percent lowest and highest observations of sales.

results from a regression of firm performance (in-degree, sales, employment and labor productivity) on distance to the supplier, controlling for supplier fixed effects and seller prefecture fixed effects. The relationship is strongly positive; increasing distance by ten percent to the supplier is associated with 1.9 percent higher sales in the buyer firm.

Robustness: A potential concern is that Facts 1 and 2 are partly driven by differences across industries. For example, large firms may belong to certain types of industries that for various reasons require many suppliers. Hence, we also explore within-industry correlations between size and our various outcome variables. Table 2 shows results when regressing firm characteristics on log size and 3-digit JSIC industry fixed effects. The outcome variables are the same as above: in-degree, median distance to suppliers and number of sourcing municipalities. For completeness, we also report results on the customer side: out-degree, median distance to customers and number of municipalities with customers.

An additional concern is that larger firms have more plants, so that distance from the relevant plant to a supplier may not be higher for larger firms. We investigate this by including an interaction term between log size and a dummy variable for whether the firm is single-plant or not. The interaction is close to zero, indicating that there is a positive relationship also for single-plant

Figure 5: Size and median distance to connections.



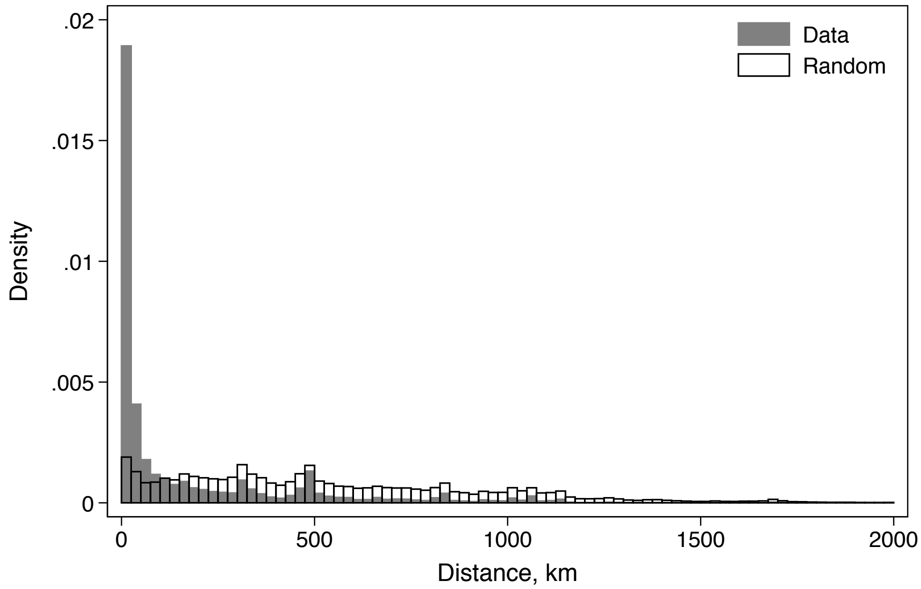
Note: 2005 data. The figure shows the kernel-weighted local polynomial regression of firm-level median log distance to the firm's connections (vertical axis) on log sales (horizontal axis). The two lines represent distance to suppliers and customers as separate regressions. Gray area denotes the 95 percent confidence bands. Sample is first trimmed by excluding the 0.1 percent lowest and highest observations of sales.

firms. We also plot the size-median distance relationship for single-plant firms only; Figure 11 in the Appendix shows the same polynomial regression plot for this group. Also, the within-supplier relationship between distance and performance continues to be positive when considering single-plant firms only (Table 1, rows 5 and 6).

In the model (Section 4), small and large firms require the same number of tasks, but large firms optimally decide to outsource more of them. An alternative hypothesis is that large firms offer more products and therefore also require more tasks than smaller firms. We do not observe the tasks performed within the firm, so we cannot directly test this hypothesis. What we can do, however, is to check whether Facts 1 and 2 also hold for firms that belong to a single 3-digit JSIC industry (in the data, each firm can belong to up to 3 industries). If the positive relationship between size and in-degree also holds for single-industry firms (within an industry), then this suggests that differences in the range of tasks produced is not driving the empirical relationships. Table 2 includes the interaction between log size and a dummy variable for whether the firm is single-industry or not. The interaction is typically negative but small; hence our results survive when considering this group of firms.

Fact 3: The majority of connections is formed locally. Distance is important in the formation

Figure 6: Density of distance across buyer-seller pairs.



Note: 2005 data. The figure shows the density of distance in km for all buyer-seller pairs. The gray bars represent the density from actual linkages whereas the white bars represent the density from random linkages.

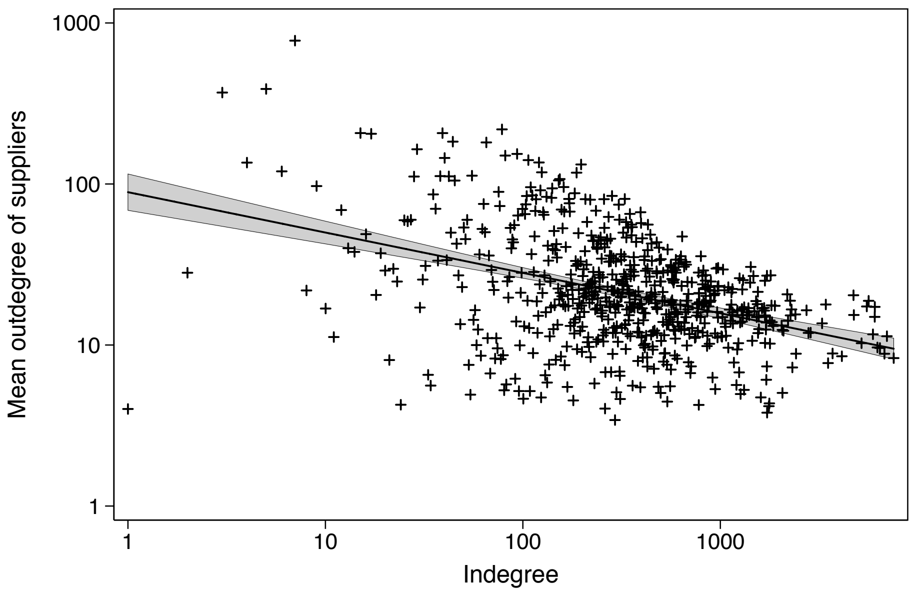
of links. We start by calculating the distance between any supplier-customer pair ij and show the density of distance in Figure 6. As above, geolocation is based on a firm's headquarters, so for multi-plant firms the interpretation is distance between headquarters. The gray bars represent the density based on actual linkages. The white bars represent the density based on random linkages.[9] The median (mean) distance is 30 (172) km. Hence, the majority of connections is formed locally. Even so, a few connections span very long distances, so that the average distance is much greater than the median. Moreover, the actual distances between firms are much smaller than what would emerge in a random network. In the network with randomly drawn connections, median (mean) distance is 464 (540) km.

Fact 4: *There is negative degree assortativity among sellers and buyers.* One distinguishing feature of networks is the extent to which a well-connected node is linked to other well-connected nodes, known as degree assortativity. While there is an extensive body of research on degree assortativity in technical and social networks, these relationships are less well documented in economics networks. We find that the better connected a firm, the less well-connected is its average connection. Figure 7 provides an overview of degree assortativity in the Japanese production network. The figure shows

[9] A random production network is generated by drawing n_i random customer links for firm i, where n_i is based on the actual out-degree of firm i. Then, distance between all random links are calculated based on the geocode of the firms.

all possible values of the number of suppliers per Japanese firm, a, on the x-axis, and the average number of (customer) connections of these suppliers, $b(a)$, on the y-axis. The interpretation of a point with the coordinates (10,1) is as follows: For a Japanese firm sourcing from 10 suppliers, the average supplier has one customer. The fitted regression line has a slope of -0.19, so a 10 percent increase in number of suppliers is associated with a 2 percent decline in the average supplier's number of customers.[10]

Figure 7: Degree Assortivity - Suppliers and Customers of Suppliers



Note: 2005 data. The figure shows all possible values of the number of suppliers per firm, a, on the x-axis, and the average number of customer connections of these suppliers, $b(a)$, on the y-axis. Axes scales are in logs. The fitted regression line and 95% confidence intervals are denoted by the solid line and gray area. The slope coefficient is -0.19.

This results suggests that the best firms, those with many connections, are selling to firms who on average have fewer connections themselves. Interestingly, social networks typically feature positive assortative matching, that is, highly connected nodes tend to attach to other highly connected nodes, while negative correlations are usually found in technical networks such as servers on the Internet (Jackson and Rogers, 2007).[11] In a recent paper, Bernard et al. (2013) also find negative assortivity between trading firms using Norwegian exporter data matched to foreign importers and

[10]The correlation between degree and mean degree of connections is a standard measure of assortativity in networks (Jackson and Rogers, 2007).

[11]In the friendship network among prison inmates considered by Jackson and Rogers (2007), the correlation between a node's in-degree and the average in-degree of its neighbors is 0.58. The correlation in our data is -0.31. Serrano and Boguna (2003) find evidence of negative sorting in the network of trading *countries*; i.e. highly connected countries, in terms of trading partners, tend to attach to less connected countries.

Colombian importer trade data matched to foreign exporters.

The finding of negative degree assortivity may be influenced by the location of the firm. To check this we control for the location of both the firm and its connections. Specifically, we ask whether firms in a prefecture i (e.g., Tokyo) with many suppliers in prefecture j (e.g., Osaka) are trading with less well-connected firms in Osaka, in terms of their number of customers in Tokyo. We include fixed effects for both the location of firm as well as its suppliers,

$$\overline{Supplier\ Outdegree}_{ij} = \alpha_i^1 + \alpha_j^2 + \beta Indegree_{ij} + \epsilon_{ij} \qquad \beta = -0.120 \quad \text{s.e. } (0.003)$$

Again we find negative assortivity: firms with more suppliers in a destination market are sourcing from firms with fewer customers. Controlling for destination countries, Bernard et al. (2013) estimate a comparable coefficient of -0.13 when considering buyer-seller matches in Norwegian exporter data and -0.20 in matches from Colombian importer data. Our findings of negativity assortivity is not limited to this specific configuration of in- and out-degree. We find similar relationships when using customer out-degree or an undirected measure of the total number of connections.[12]

4 The Model

We develop a parsimonious model of outsourcing in a domestic economy motivated by the facts in the previous section. The basic structure is as follows. Firms require a continuum of tasks as inputs into the production process, e.g. materials processing, accounting, printing and mailing services. They can produce the tasks themselves or outsource them. Finding suppliers is costly, however, and therefore it may not be profitable for all firms to outsource a given task, even though the market price of a task is lower than the firm's marginal cost of supplying the same task. This setup will produce theoretical predictions that are consistent with the empirical regularities documented in Section 3 and will guide the development of the empirical methodology in Section 5. Our model is closely related to the framework in Antràs et al. (2014), but we modify it in several directions. First, we allow for the possibility of in-house production, i.e. that firms can supply a given task within the boundary of the firm. This margin of adjustment is crucial in order to match the empirical regularities. Second, geography in the model will be continuous. We combine this with distributional assumptions which allow us to obtain sharp analytical results. Third, our model is a framework for understanding domestic, and not international, sourcing. Since productivity differences are typically much smaller within a country than across countries, and since labor is typically much more mobile within a country than across countries, we assume that wages and productivity are common across locations.

[12]In our data, the correlation is negative and significant for (in-degree,average in-degree of suppliers), (in-degree,average in-degree of customers), (in-degree,average out-degree of customers), (out-degree,average out-degree of suppliers), (out-degree,average in-degree of suppliers), (out-degree,average in-degree of customers), (out-degree,average out-degree of customers).

4.1 Setup

Geography, sectors and market structure. The economy consists of a unit continuum of locations $i \in \mathcal{S}$. Each location consists of an upstream and a downstream sector. Downstream firms combine labor and a unit continuum of tasks and sell their output to final consumers. Upstream firms produce a single task using labor only. Within a location i and for a given task ω, there are many identical firms producing ω at the same marginal cost. Hence, the upstream sector is characterized by perfect competition. Downstream firms are monopolistically competitive and produce a differentiated good with efficiency z which varies across firms.

Production technology. The production function of a downstream firm is

$$y = zl^{\alpha}v^{1-\alpha},$$

where l is labor, α is the labor share and v is a CES composite of the unit continuum of tasks. The CES price index is

$$P^{1-\rho} = \int_0^1 p_u(\omega)^{1-\rho}\, d\omega,$$

where $p_u(\omega)$ is the price of an individual task ω and ρ is the elasticity of substitution between tasks. The firm can potentially produce all tasks in-house. If so, the firm's efficiency in producing a task ω is a realization of a random variable ϕ from the Frechet distribution $F(\phi) = e^{-T_0\phi^{-\theta}}$, where T_0 determines the average efficiency in producing a task and $\theta > \rho - 1$ is inversely related to dispersion in task productivity. $F(\phi)$ is identical across all downstream firms, hence, total factor productivity z is the only source of firm-level heterogeneity. As we will see, in equilibrium, the price $p_u(\omega)$ will depend on whether the firm outsources a task or not and, if outsourced, which location it will outsource from.

The production function of an upstream firm in location i is $y_u(\omega, i) = \phi(\omega, i)\, l$. The efficiency of producing a task ω is a realization of a random variable ϕ from the Frechet distribution $F_u(\phi) = e^{-T\phi^{-\theta}}$. The parameter T governs the average productivity. To keep the model tractable we assume that average productivity T is identical across locations i. Upstream firms in i selling to j are subject to iceberg trade costs $\tau(i,j) \geq 1$. Trade costs are here broadly defined; we have in mind both shipping costs and efficiency losses in the buyer-supplier relationship. The cost of supplying ω from i to j is therefore $w(i)\,\tau(i,j)\,/\phi(\omega, i)$, where $w(i)$ is the nominal wage in i. For tractability, we assume that final goods are costlessly traded. This makes the price index of final goods identical in every location.[13]

Labor. Each location is inhabited by $L(i)$ workers, and the aggregate (exogenous) supply of workers is $\bar{L} = \int_{\mathcal{S}} L(i)\, di$. Consumers derive utility from consumption of the downstream goods.

[13] Without this assumption, nominal wages $w(i)$ would vary across locations because the final goods price index would vary across locations (given that labor mobility equalizes real wages across locations). Costlessly traded final goods allows us to abstract from this.

They have identical CES preferences with an elasticity of substitution σ. There is perfect labor mobility across regions. Since finals goods are costlessly supplied to consumers across locations, nominal wage equalization is sufficient to leave workers indifferent between locations. Henceforth, we denote the common nominal wage by w.

Entry. There is a fixed measure of downstream firms in each location, $m(i)$. As there is no free entry, the production of final goods leaves rents. We assume that consumers derive income not only from labor but also from the dividends of a mutual fund. Each consumer owns w shares of the fund and profits are redistributed to them in units of labor. Total worker income in location i is then $w(1+\psi)L(i)$, where ψ is the dividend per share of the mutual fund.

Outsourcing. The downstream firm located in j can choose to produce a task ω itself or outsource it. The firm can observe average productivity T and trade costs $\tau(i,j)$ from source i. Observing individual prices for all ω, however, requires effort. Specifically, the firm must incur a fixed cost $f(j)$ paid in terms of labor to observe individual prices in a location i.[14] As we will see, more productive firms find it optimal to search a wider range of locations because the marginal profits from search are higher for high z firms, while the marginal cost $f(j)$ is constant. Given that $f(j)$ does not vary by source, each location i can be ranked according to its attractiveness as a supplier location, where attractiveness is defined by $\tau(i,j)^{-\theta}$ (see Antràs (2014)). A firm in j will therefore search all locations i where $\tau(i,j)$ is lower than some threshold value (to be defined below). As in Eaton and Kortum (2002), conditional on a set of search locations, firm z's share of purchases from location i is

$$\chi(z,i,j) = \frac{T\tau(i,j)^{-\theta}}{\Phi(z,j)}.$$

$\Phi(z,j)$ is a measure of market access,

$$\Phi(z,j) = T_0 + \int_1^{\bar{\tau}(z,j)} T\tau^{-\theta} g(\tau,j)\, d\tau, \tag{1}$$

where $\bar{\tau}(z,j)$ is the highest cost location that z located in j is willing to search. g is the density of trade costs to location j.

The share of tasks outsourced is

$$o(z,j) = 1 - \frac{T_0}{\Phi(z,j)}.$$

Adding more locations to search will raise $\bar{\tau}$ and Φ. More search therefore gives more outsourcing o. As in Eaton and Kortum (2002), the task price index is $P(z,j) = \lambda w \Phi(z,j)^{-1/\theta}$ where λ is a constant.[15] Hence, more outsourcing leads to lower input costs P with an elasticity $1/\theta$. Searching an additional location means that the firm can observe a new set of prices for all tasks ω. The

[14] In order to keep the problem tractable, we do not allow an (i,j) specific f.

[15] $\lambda^{1-\rho} = \Gamma\left(\frac{\theta-(\rho-1)}{\theta}\right)$ where Γ is the Gamma function.

probability of finding at least one task with a lower price than the existing one is strictly positive, and therefore the price index $P(z,j)$ must go down.

4.2 Optimal Search

The maximization problem of the firm is

$$\max_{\bar{\tau}} \left\{ \pi(z,j) - wf(j)\, n(z,j) \right\},$$

where $\pi(z,j)$ is gross profits of firm z located in j and $n(z,j)$ is the measure of locations to search. Total sales of the downstream firm can be written $r = Ap^{1-\sigma}$ where A is a demand shifter and p is the firm's price. Profits are proportional to sales, $\pi = r/\sigma$. Appendix A derives the solution to the problem of the firm as well as the second order condition. The solution to $\bar{\tau}$ is

$$\bar{\tau}(z,j) \quad = \quad \kappa_1 \left(\frac{T}{w^{\sigma}} \frac{A}{f(j)} \right)^{1/\theta} \Phi(z,j)^{-k/\theta} z^{(\sigma-1)/\theta} \tag{2}$$

where $k = 1 - (\sigma - 1)(1 - \alpha)/\theta$ and κ_1 is a constant.[16] For an arbitrary geography $g(\tau, j)$, one can jointly solve equations (1) and (2), which is a system of two equations and two unknowns $\bar{\tau}(z,j)$ and $\Phi(z,j)$.

The expression for the hurdle $\bar{\tau}$ has a number of interesting features. First, better market access Φ leads to more search when $k < 0$ and less search when $k > 0$. The model of Antràs (2014) has the same property and describes these as the complements and substitutes cases respectively. Keeping Φ constant, lower search costs $f(j)$ lead to more search (higher $\bar{\tau}$). Higher efficiency z and more demand A also lead to more search (higher $\bar{\tau}$).

4.3 Model and Data

We now return to the stylized facts presented in Section 3 and relate them to the model. The proofs are found in Appendix B.

First, more productive firms outsource more tasks and therefore have more suppliers:

$$\frac{\partial o(z,j)}{\partial z} > 0,$$

because $\partial \Phi(z,j)/\partial z > 0$. Given that more productive firms search more, they are more likely to find a sourcing option for a given task ω at a lower cost than the cost of producing in-house. This is consistent with the evidence in Figure 3, that larger firms tend to have more suppliers. Note that, according to the model, higher efficiency z leads to both increased sales and in-degree, while higher in-degree itself leads to greater sales. Hence, the level of sales for a given firm is determined by both

[16] $\kappa_1 = \left(\frac{\left(\bar{m}\lambda^{1-\alpha}\right)^{1-\sigma}}{\sigma} \frac{(\sigma-1)(1-\alpha)}{\theta} \right)^{1/\theta}$

the direct effect of core efficiency z and the indirect effect of in-degree. The positive correlation shown Figure 3 is a result of both the direct and indirect effects.

Second, more productive firms search more and costlier locations:

$$\frac{\partial \bar{\tau}}{\partial z} > 0.$$

High z firms have a greater incentive to search more locations because the potential cost savings are larger for more productive firms. As a consequence, more productive firms have higher maximum and average trade costs to suppliers. This is consistent with the evidence in Figures 4 and 5, that larger firms tend to have suppliers in more locations and higher distance to their suppliers. A corollary is that, when comparing a supplier's customers in different markets, the average customer in a more remote market is more productive than the average customer in a nearby market. This is shown formally in Appendix B.4 and is consistent with the empirical evidence in Table 1, that remote customers of a given supplier tends to be larger and have many connections.

Third, more productive firms reach suppliers in markets that are on average less well connected. Specifically, consider a firm with efficiency z in location j, sourcing from the marginal location $\bar{\tau}(z, j)$. Denote the expected measure of customers from j among upstream firms in z's marginal location $\bar{c}(z, j)$. Then

$$\frac{\partial \bar{c}(z, j)}{\partial z} < 0.$$

This reflects the fact that higher z firms reach costlier locations and the suppliers there are on average not very competitive in z's home market. This is consistent with the evidence in Figure 7 on negative degree assortativity.[17]

4.4 The Density of Trade Costs

So far, we have not imposed any structure on the density of trade costs $g(\tau, j)$. In this section, we will choose a functional form for $g()$ that will allow us to derive closed form expressions for key relationships in the model (e.g. equations (1) and 2)). This in turn helps us to derive theoretical predictions that will later be tested in the data (Proposition 1 in the next section).

We assume that $g(\tau, j)$ is inverse Pareto with shape $\gamma > \theta$ and support $[1, \tau_H]$, $g(\tau) = \gamma \frac{\tau_H^{-\gamma}}{1-\tau_H^{-\gamma}} \tau^{\gamma-1}$. An inverse Pareto captures the notion that a location has few nearby markets and many remote markets. The upper bound τ_H is the maximum trade cost within Japan. One can show that a higher τ_H shifts the trade cost distribution to the right. In other words a distribution with a high τ_H first-order stochastically dominates a distribution with a low τ_H (Appendix F).

[17] Note that the dependent variable in Figure 7 is *average* out-degree whereas $\bar{c}(z, j)$ is *marginal* out-degree, i.e. expected out-degree in z's least profitable location. In practice, this makes little difference because in the model the average is pinned down by the marginal out-degree.

Therefore, τ_H is a convenient metric for average trade costs in the economy.Appendix F provides empirical evidence that the inverse Pareto is a good approximation of the empirical distance density in our dataset.

4.5 Testable Predictions

As we discuss later in Section 5, we will exploit a natural experiment where a large shock to infrastructure lowered passenger travel time (but not goods travel time) between many location-pairs in Japan. This empirical exercise allows us to to quantify the impact of a large-scale infrastructure project on firm performance and to evaluate the importance of the theoretical mechanism emphasized in this paper. In order to guide the subsequent empirical work, this section details the consequences of such a shock according to the model.

First, consider the impact on firm sales of lower search costs $f(j)$. Lower $f(j)$ leads to sales growth of a downstream firm in j. Holding final goods demand A constant,

$$\frac{\partial \ln r(z,j)}{\partial \ln f(j)} = \frac{(\sigma-1)(1-\alpha)}{\theta} \frac{\partial \ln \Phi(z,j)}{\partial \ln f(j)} < 0.$$

The elasticity $\partial \ln \Phi(z,j) / \partial \ln f(j) < 0$ measures the fall in market access from an increase in $f(j)$ (Appendix B.2). Now consider how $\partial \ln r(z,j) / \partial \ln f(j)$ varies across industries with different labor intensities α:

$$\frac{\partial^2 \ln r(z,j)}{\partial \ln f(j) \partial \alpha} = \frac{\sigma-1}{\theta} \left(-\frac{\partial \ln \Phi(z,j)}{\partial \ln f(j)} + (1-\alpha) \frac{\partial^2 \ln \Phi(z,j)}{\partial \ln f(j) \partial \alpha} \right).$$

The cross elasticity is the sum of direct and indirect effects. The direct effect is that a percent reduction in input costs $P(z)$ will have a stronger positive effect on sales in industries where inputs constitute a large share of total costs. The indirect effect is that input-intensive firms may search more or less intensively relative to labor-intensive firms when $f(j)$ falls (the cross elasticity $\partial^2 \ln \Phi(z,j) / \partial \ln f(j) \partial \alpha$). Appendix D shows that both the direct and indirect effect have the same sign when $g(\tau,j)$ is inverse Pareto. Hence, the total effect is

$$\frac{\partial^2 \ln r(z,j)}{\partial \ln f(j) \partial \alpha} > 0,$$

so that sales growth is stronger for input-intensive firms relative to labor-intensive firms when search costs fall.

Second, consider the impact on firm sales of lower variable trade costs. Using the same inverse Pareto parameterization of the trade cost distribution $g(\tau,j)$, Appendix D shows that $\partial \ln r(z,j) / \partial \ln \tau_H < 0$ and $\partial^2 \ln r(z,j) / \partial \ln \tau_H \partial \alpha > 0$. Hence, lower variable trade costs, e.g. more efficient buyer-supplier relationships or lower shipping costs, increase sales, and sales growth is stronger for input-intensive firms. We summarize this in the following proposition.

Proposition 1. *(i) Lower search costs $f(j)$ and average trade costs lead to growth in sales among downstream firms in j. (ii) Sales growth is stronger in input-intensive (low α) relative to labor intensive (high α) industries.*

Proof. See Appendix D. □

Part (ii) of Proposition 1 forms the basis of our identification strategy in Section 5.1.

Third, consider the impact on supplier connections among firms in j of lower search costs $f(j)$. Lower $f(j)$ leads to more outsourcing and suppliers from more locations among downstream firms in j, see Appendix B.2,

$$\frac{\partial o(z,j)}{\partial f(j)} < 0 \quad \text{and} \quad \frac{\partial n(z,j)}{\partial f(j)} < 0.$$

Lower $f(j)$ means that the cost of obtaining information about prices is lower. Firms therefore search additional locations ($\bar{\tau}$ and n increases). There is a positive probability of finding a task at a lower price compared to the price of in-house production. Hence, outsourcing must also increase.

Fourth, consider the impact on supplier connections among firms in j of lower average variable trade costs. Again, using the Pareto parameterization, one can show that

$$\frac{\partial o(z,j)}{\partial \tau_H} < 0 \text{ and } \frac{\partial n(z,j)}{\partial \tau_H} < 0.$$

Intuitively, lower average trade costs, e.g. due to more efficient buyer-seller relationships, induce firms to search more markets and outsource a larger share of the tasks. We summarize this in the following proposition.

Proposition 2. *Lower search costs and lower average variable trade costs lead to more outsourcing and suppliers from more locations among downstream firms in j.*

Proof. See Appendix D. □

4.6 Closing the Model

Given the assumptions above, both product markets and the labor market clear. Labor market clearing can be seen as follows. Expenditure by final consumers, E, equals total wage income plus profits from the monopolistic sector. Moreover, consumer income must equal the value of output in upstream and downstream production, $Sales^{Up}$ and $Sales^{Down}$, respectively. Hence, we have

$$\begin{aligned} E &= w(1+\psi)\bar{L} \\ &= Sales^{Up} + Sales^{Down}. \end{aligned} \tag{3}$$

The aggregate value of labor demand equals aggregate sales in upstream and downstream production minus aggregate profits. Inserting equation (3) gives us

$$wL^D = Sales^{Up} + Sales^{Down} - w\psi\bar{L} = w\bar{L}.$$

Therefore, aggregate labor demand L^D equals supply $\bar{L}$.

5 Production Networks and Productivity: A Natural Experiment

This section details our identification strategy for estimating the impact of lower trade and search costs on firm performance and linkages in the production network. We start by providing some background on the natural experiment.

The southern portion of the high-speed (bullet) train network in Japan was expanded in March 2004 (Kyushu Shinkansen). This resulted in a dramatic reduction in travel time between major cities in the area. For example, travel time between Kagoshima and Shin-Yatsushiro declined from 130 minutes to 35 minutes, and travel time between Hakata and Kagoshima declined from 4 hours to just 2 hours. Figure 8 gives an overview of the geography. The black dots are locations within 30 km of a new Shinkansen station, whereas the gray dots are all other localities in the dataset. Although the geographical scope is somewhat limited, the new rail line extended Shinkansen service to two prefectures with a total population of 3.5 million, roughly that of Connecticut.

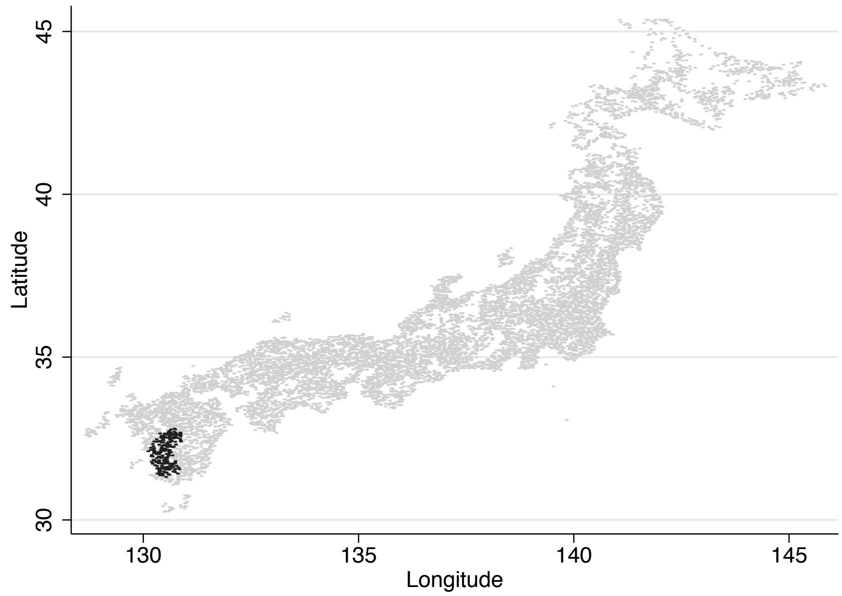
The Shinkansen expansion offers several advantages for assessing the impact on infrastructure on linkages and firm performance. First, the plan of the expansion started already in 1973, making it relatively unlikely that firms in our sample could influence the timing and location of stations. Moreover, the timing of completion was subject to substantial uncertainty starting in 1991, limiting the scope for anticipation effects. Nevertheless, our empirical methodology addresses endogeneity concerns in a variety of ways as discussed below. Second, goods in Japan do not travel on the Shinkansen and there was no contemporaneous reduction in travel time for goods along this southern route, as discussed in Section 5.1.4.

5.1 Economic Integration and Firm Performance

In this section we ask whether and to what extent the Shinkansen expansion improved performance among firms in affected regions. As shown in Proposition 1, the model suggests a simple identification strategy. Lower variable trade costs, due to more efficient buyer-supplier relationships, and lower search costs, $f(j)$, improve firm sales because they reduces marginal costs directly and indirectly, by enabling firms to find lower cost or higher quality suppliers. Moreover, the impact is greater for input-intensive firms (low α firms) relative to labor-intensive firms.[18] Intuitively, improved travel time has no impact on marginal costs in an industry that does not rely on inputs, i.e. when $\alpha = 1$. Of course, sales may improve in $\alpha = 1$ industries as well, because improved travel

[18] A disconnect between the theory and the data is that, in the data, virtually all firms are both suppliers and customers. As such, there is no clear empirical distinction between upstream and downstream firms. One could extend the model to incorporate a round-about production structure, so that suppliers are themselves sourcing from other firms, with an industry-specific labor share α.

Figure 8: Kyushu Shinkansen treated cells.



time allows firms to find new customers. Our empirical strategy will difference out this mechanism, i.e. the methodology only identifies marginal cost effects and not demand side effects.

Consider the following regression,

$$\ln y_{fjrt} = \alpha_f^1 + \alpha_{rt}^2 + \beta Station_f \times H_j \times I\left[t \geq 2004\right] + \beta X_{fjrt} + \epsilon_{fjrt}, \tag{4}$$

where y_{fjrt} is a measure of firm performance for firm f in industry j located in region r at time t, relative to average performance in the same industry-year.[19] We focus on the 8 year period 2000 to 2008, i.e. 4 years before and after the infrastructure shock.

In addition to using sales as the outcome variable, we also use sales per employee (LP) and revenue TFP ($TFPR$). In monopolistic competition models, LP and $TFPR$ are constant across firms within an industry. However, if output prices among treated firms do not adjust immediately in response to the infrastructure shock, then LP and $TFPR$ are expected to increase, see Appendix E. For $TFPR$, this occurs because $TFPR$ controls for the firm's *value* of inputs (in the absence of firm-level input deflators), so that if firm-level input prices decline, then $TFPR$ will rise.[20]

The main independent variable is the interaction between $Station_f$, which is one if firm f is within 30 km of a new station, H_j which is the input intensity of the industry in 2003 and $I\left[t \geq 2004\right]$

[19] Appendix G.1 details the construction of the dependent variable.

[20] Firm performance is measured relative to industry (3 digit)-year means. There are in total 315 3 digit industries in the data. $TFPR$ is estimated using the Olley and Pakes (1996) methodology, see Appendix G.1.

which is an indicator variable taking the value 1 from 2004 and onwards. The 30 km threshold for station is chosen so that total travel time is significantly affected and that Shinkansen dominates alternative modes of transport. For example, for a firm 60 km from a station, car travel time to the station would amount to 40 to 60 minutes, and hence the percentage drop in total travel time would be significantly less compared to a firm located near the station. We also check the results with other thresholds in Section 5.1.2. Input intensity is defined as 1 minus the labor share of the 3-digit industry in 2003. The labor share is the industry's wage costs relative to total costs, see Appendix G. α_f^1 and α_{rt}^2 are firm and prefecture-year fixed effects. There are 47 prefectures in Japan.

The covariates in X_{fjrt} are the remaining interactions $Station_f \times I\,[t \geq 2004]$ and $H_j \times I\,[t \geq 2004]$. In addition, since prefectures are relatively large geographic areas, we introduce a second geographic control. Each prefecture is further divided into local administrative units called municipalities. We have in total 1410 municipalities in the dataset, making the average population of a municipality roughly 90,000 (Japan's total population was 127.8 million in 2005) The high number of municipality-year pairs means that municipality-year fixed effects are computationally infeasible. We can, however, include a variable for average performance in a municipality-year.[21] For example, if $\ln y_{fjrt}$ is sales, we include average log sales excluding firm f in the municipality-year as a control variable.

The regression in equation (4) is a triple differences model and the intuition for identification is as follows. The Shinkansen expansion is expected to bring higher performance gains for an input-intensive firm located close to a new station compared to a labor-intensive firm located close to a new station. The empirical strategy is to compare the growth of input-intensive firms before and after 2004 ($1st$ difference) to the growth of labor-intensive firms ($2nd$ difference), and compare this differential effect in locations with a new station relative locations without a new station ($3rd$ difference).

The triple differences approach resolves a number of potential concerns. First, performance growth due to demand-side effects (i.e. growth among labor-intensive firms due to new customers) is differenced out because demand side effects are expected to affect labor-intensive and input-intensive firms similarly.[22] Second, a potential concern is that input-intensive firms may grow faster than labor-intensive firms even in the absence of the Shinkansen expansion. The methodology controls for this because the triple interaction coefficient β will only capture the differential impact (input intensive relative to labor intensive) for firms close to a station relative to the differential impact for firms that are far from a station. Hence, if input-intensive firms grow faster in every location, then β will be zero. Third, a potential concern is that the new Shinkansen line was introduced in high growth regions. As we only compare the differential growth for input relative to

[21] This approach is similar to Giroud (2013).

[22] The common demand side effect is captured in $Station_f \times I\,[t \geq 2004]$.

labor-intensive firms, endogeneity is not a concern as long as the Shinkansen line was not targeted particularly for input-intensive firms.

5.1.1 Results

Table 3 shows regression results from estimating equation (4). Column (1) uses log sales relative to the industry-year as the dependent variable. The triple interaction term β is positive and significant at the 5 percent level, indicating that the Shinkansen expansion boosts firm sales for input-intensive firms relative to labor-intensive firms. The magnitudes are economically significant: a coefficient of 0.47 means that a Shinkansen stop increased sales by 0.47 log points more for a firm with $H_j = 1$ relative to a firm with $H_j = 0$. A firm in the $9th$ decile of the H_j distribution ($H_j = 0.92$, e.g., industrial plastic products, JSIC 183) increased sales by roughly 0.10 log points more than a firm in the $1st$ decile of the H_j distribution ($H_j = 0.70$, e.g., general goods rental and leasing, JSIC 701). Columns (2) and (3) use labor productivity and $TFPR$ as the dependent variable. Again, the triple interaction term is positive and significant, suggesting that the infrastructure shock improved firm's productivity. The magnitudes are slightly smaller compared to sales, a firm in the $9th$ decile of the H_j distribution improves labor productivity by 0.09 log points faster than a firm in the $1st$ decile.

The fact that an infrastructure project unrelated to transportation of goods can improve firm performance by this magnitude is indeed remarkable. More broadly, our findings suggest that domestic trade costs dampen economic activity by limiting buyer-supplier linkages and that reducing these barriers will help development and growth. From a policy perspective, the results point to important positive effects of large-scale infrastructure projects that are typically neglected from cost-benefit analyses - that infrastructure projects can bring efficiency gains from freer flow of information across firms.

5.1.2 Robustness

In this section, we explore a number of robustness checks. First, a potential concern is that input-intensive firms near a new station tend to grow faster than labor-intensive firms near a new station (but not in other locations), i.e. that there are pre-trends in the treatment relative to the control group. A simple way to check for this is to conduct a falsification test. We estimate equation (4) on the five year period 1998 to 2002 and incorrectly set the Shinkansen expansion to 2000, i.e. we replace $I[t \geq 2004]$ with $I[t \geq 2000]$. The results are shown in Table 4. For all three dependent variables, the triple interaction term is not significantly different from zero, hence there are no pre-trends in the data.

Another concern is that the results are sensitive to the chosen 30 km threshold for whether a firm belongs to a new station or not. We therefore estimate the model with $Station_f$ taking the value one if firm f is within 10 km of a new station. The results in Table 5, columns (1) to (3) show that the results are relatively close to the baseline - the impact on sales is slightly stronger and the impact on sales per employee and TFP is roughly similar. We also investigate whether firms in cities near those served by the Shinkansen were affected. A common concern is that firms in adjacent cities lose market share because they become less competitive relative to firms in cities with Shinkansen service, the so-called "straw effect". In Table 5, columns (4) to (6), we add an indicator variable for whether a firm is between 30 to 60 km of a new station and the set of interactions with H_j and $I[t \geq 2004]$. The triple interaction term is slightly positive when sales is the dependent variable and slightly negative when sales per employee and TFP is the dependent variable, indicating that the impact on firms in adjacent cities is mixed.

Finally, a potential issue is that the results are primarily driven by the construction sector, possibly because construction firms grow due to increased demand related to building the new infrastructure. Note, however, that the triple differences approach should control for this, since identification is based on comparing industries with different input intensities.[23] Nevertheless, we re-run the regression excluding firms belonging to construction industries (Table 6, columns (1) to (3)). Overall the results are very similar to the baseline results in Table 3.

5.1.3 Commuting

The infrastructure shock also benefited firms due to more efficient commuting. First, the reduction in travel time for existing commuters benefits both workers and firms by freeing up more time for work and leisure. Second, firms in the treated areas can potentially attract and hire more workers or find workers with skills that are better matched to the firm. Both effects will improve firm sales and performance. If the positive impact coming from commuting is identical across industries, then our results will be unaffected, since changes that affect all industries at the same time are differenced out. Still, it could be that some industries benefit more from commuting than others.

One mechanism might be that labor intensive industries gain more than input- or capital intensive industries simply because labor constitutes a larger share of the costs of production. This would mean that the true effect coming from supplier linkages is larger than what we have estimated, because input intensive industries benefit less from commuting than labor intensive industries. This would tend to bias our results towards zero.

Another mechanism might be that skilled workers consider the cost of travel time to be higher than unskilled workers, so that a large drop in travel time is more beneficial for skill-intensive industries. We test this hypothesis by constructing a measure of skill intensity for each industry, and then including triple interaction terms for skill intensity, in addition to the existing triple interaction terms for input intensity. In the data, there is no direct measure of skill intensity. There

[23] Moreover, any potential bias would be negative because the construction demand shock occurred before 2004, not after.

is, however, a variable for the number of R&D workers. We define an industry's R&D intensity as the number of R&D workers relative to total workers in the industry. Across JSIC 3-digit industries, almost 20 percent of the industries report zero R&D workers. Our preferred measure of skill intensity is therefore an indicator variable equal to one if the industry has higher than median R&D intensity (the median is 0.013). The results are shown in Table 6, columns (4) to (6); the input intensity interaction term is close to the baseline results, while the skill intensity indicator is close to zero. In sum, we conclude that, although commuting is certainly an important factor, it does not significantly affect our estimates of the supply chain effect.

5.1.4 Congestion

Because the bullet train line only carries people, the main interpretation of our results is that the infrastructure shock facilitated face-to-face meetings, leading to more suppliers and more efficient buyer-seller relationships. A potential concern, however, is that the bullet train line freed up capacity across other modes of transport, leading to less congestion for trucks and regular trains. If that were the case, then the interpretation of our results would be different. We investigate this by exploring data from the Net Freight Flow Census from Japan's Ministry of Land, Infrastructure, Transport and Tourism (MLIT). The census provides data on average freight time across Japan's prefectures. Table 7 shows the percent change in freight time from 2000 to 2010 across the prefectures on Japan's southern main island (Kyushu). The cells in bold are the prefecture pairs that were affected by the bullet train. Although the data is noisy, there is no evidence that freight times fell, or increased by less, in the affected prefectures relative to the unaffected prefectures. The average (median) increase in freight time across all pairs were 18 (22) percent, while the increase in the affected pairs were 21 and 24 percent (for Kagoshima-Fukuoka and Kagoshima-Kumamoto respectively).

5.2 Economic Integration and Firm Linkages

The empirical results from Section 5.1 show that the infrastructure shock improves firm performance and that the performance effects are stronger among input-intensive firms, consistent with the model. In this section, we explore in more detail the economic mechanism behind this decline in marginal costs. Our model suggests that the input price index of the firm, $P(z)$, falls because treated firms work more efficiently with existing suppliers, outsource more tasks and find better suppliers for existing tasks (Sections 4.3 and 4.5). This section provides additional evidence that the performance gains are driven by the supplier channel.

Recall that according to Proposition 2, lower trade and search costs lead to more supplier connections of downstream firms in j. The aim of this section is therefore to test whether the Shinkansen expansion affected the growth in supplier connections between the 2005 and 2010 cross-sections of the TSR data.

A problem with the TSR data is that a firm surveyed in 2005 may not be surveyed in 2010. Hence, at the firm-level, the number of supplier connections could change simply due to variation in sampling. In order to mitigate this issue, we aggregate the TSR data as follows. We divide Japan into a grid consisting of 500 $\times$ 500 cells; each cell is a square roughly 5.6 kilometers on a side. Next, we define C_{ijt} as the number of suppliers in i serving customers in j at time t, where $t = \{2005, 2010\}$. As many cells are covering water and other non-populated areas, the dataset is reduced to roughly 8,000 cells, or localities, after removing these regions. We assign a dummy variable $Treat_i = 1$ to a locality if one or more firms in i are within 30 kilometers of a new Shinkansen stop. We also calculate great circle distances between the center of cells i and j, $Dist_{ij}$. Note that the sample selection issue described above is greatly reduced because C_{ijt} is the sum of supplier connections among all firms in j, so any sample selection noise in the in-degree of a given firm is likely to be averaged out.[24]

In the data, a location j may get new connections for a variety of reasons which may be correlated with the infrastructure shock. The empirical strategy is therefore to test whether location-pairs ij where either i or j or both gets a new station increase their number of connections relative to location-pairs where neither i nor j gets a new station. We therefore estimate the following model,

$$\Delta \ln C_{ij} = \xi_i^1 + \xi_j^2 + \beta_1 Both_{ij} + \beta_2 One_{ij} + \gamma X_{ij} + \epsilon_{ij}, \tag{5}$$

where the dependent variable is the change in the log number of connections between suppliers in i and buyers in j, $\Delta \ln C_{ij} = \ln C_{ij2010} - \ln C_{ij2005}$. The main independent variables are $Both_{ij}$ which equals one if both i and j get a new station and One_{ij} which equals one if either i or j (but not both) gets a new station.[25] Location-pairs where neither i nor j gets a new station are the omitted group. ξ_i^1 and ξ_j^2 are source and destination fixed effects respectively and X_{ij} is a vector of covariates (see below).

The inclusion of $Both_{ij}$ allows for the possibility that the impact on location-pairs with both ends being newly connected to the Shinkansen train network may be stronger than if only one of them is connected. The inclusion of One_{ij} allows for a new Shinkansen station close in i (but not j) to impact connections from i to j and j to i.

The empirical framework has the flavor of a gravity model of the extensive margins of trade. Consider the non-differenced version of equation (5), i.e. a specification with $\ln C_{ij}$ as the dependent variable and with source-year, destination-year and location-pair fixed effects. That model would identify the impact on the number of connections from the change in travel time for an affected location-pair relative to unaffected location pairs. The source-year and destination-year fixed effects would capture trends in economic activity which may differ across locations. The location-pair fixed

[24] The average number of firms in a cell is 104, see Table 8.

[25] Formally, $Both_{ij} = I\left[Treat_i = 1 \cap Treat_j = 1\right]$ and $One_{ij} = I\left[(Treat_i = 1 \cap Treat_j = 0) \cup (Treat_i = 0 \cap Treat_j = 1)\right]$.

effects would control for time-invariant determinants of bilateral trade between locations. Due to the large number of fixed effects, we use the log change in C_{ij} and obtain the estimating equation (5).

The production network is observed at two moments in time, 2005 and 2010. The timing is not ideal for our purposes, because the Shinkansen extension occurred in March 2004. The underlying assumption is therefore that the impact of the expansion had not fully materialized when firms were surveyed in 2005. Our hypothesis is that finding new suppliers is a slow and costly process, so that it is unlikely that firms had fully adjusted after one year. Note that our estimates are biased towards zero if firms partially adjusted before 2005.

5.2.1 Results

We start by documenting a few basic facts about the locality connections dataset. Table 8 show descriptive statistics for the number of connections in 2005 and 2010. The average (median) number of connections between a locality-pair was 7.05 (2) in 2005 and increased slightly to 7.51 (2) in 2010. There are roughly 8,000 localities L and almost 400,000 locality-pairs with positive flows in both 2005 and 2010. This implies that many locality-pairs have zero connections, i.e. the number of locality-pairs with positive transactions is much smaller than the theoretical maximum $\left(L^2 - L\right)$.[26]

Table 9 presents the results from estimating equation (5). Column (1) estimates the model without any fixed effects, while columns (2) includes source and destination fixed effects. The inclusion of these fixed effects controls for changes in the number of firms in the localities as well as region-specific shocks (e.g., productivity shocks, population growth, internet and cell phone coverage, and so on). Column (3) and (4) includes log distance and log distance interacted with $Both_{ij}$ and One_{ij} as additional independent variables. Our preferred specifications are (3) and (4), as there is evidence of agglomeration over time, i.e. that local connections grow faster than remote connections (the distance coefficient being negative). It is important to control for this because $Both_{ij}$ is negatively correlated with distance.

Overall, the results indicate that localities becoming connected by new stations ($Both_{ij} = 1$) increased their number of connections by roughly 40 percent relative to unconnected localities. Recall that the average number of connections is 7, so for the average locality-pair, the Shinkansen extension caused 3 new connections between newly connected localities. Our preferred specifications in columns (3) and (4) suggest that the impact is roughly half as large when only one of the locations in a pair is connected ($One_{ij} = 1$). Perhaps surprisingly, the interaction terms are close to zero. We find no evidence that the infrastructure shock benefited remote connections more than local

[26] As the dependent variable is $\Delta \ln C_{ij}$, we drop pairs with missing $\Delta \ln C_{ij}$ ($C_{ij2005} = 0$ or $C_{ij2010} = 0$). The 8000 localities i in the data therefore have either $C_{ij} > 0$ or $C_{ji} > 0$ for $j \neq i$. ii pairs are also dropped from the dataset because distance is zero for these pairs and log distance is used as an independent variable in several regressions.

connections.[27]

Robustness. In the baseline results, we used the threshold of 30 kilometers from a new station to classify cells as treated or untreated. To check the sensitivity of the results, we instead use a threshold of 10 kilometers. The results are presented in Table 10 and indicate that the results are robust to this change. A potential concern in the baseline results is that location-pairs with either $C_{ij2005} = 0$ or $C_{ij2010} = 0$ are dropped because of the log transformation. We address this by replacing the dependent variable $\Delta \ln C_{ij}$ with a $\Delta I \left[C_{ijt} > 0\right]$, i.e. a dummy variable taking the value one if location-pair ij starts trading between 2005 and 2010, zero if there is no change, and minus one if ij stops trading. Including all the zeros in the dataset results in an extremely high number of observations (71 million), so estimation with joint source and destination fixed effects becomes computationally infeasible. Table 11 therefore reports results with either source or destination fixed effects in columns (2) and (3) respectively. The estimates confirm the findings in the baseline specification; location-pairs with new stations are more likely to start trading (and less likely to stop trading) between 2005 and 2010.

Summing up, we find significant growth in firm linkages between regions connected by the bullet train. Moreover, this reallocation of the production network is consistent with the firm performance gains found in Section 5.1.

6 Conclusions

This paper examines how firm performance is related to the characteristics of the supply network with a special focus on geography. Using a comprehensive, unique data set on supplier-customer links among 950,000 Japanese firms, we develop a set of facts about the production network. Geographic proximity plays a key role for the matching of suppliers and customers as most connections cover relatively short distances. Large, more productive firms both have more suppliers and, on average, have suppliers that are farther away. While large firms have more suppliers, the trading partners of those large, well-connected firms on average are less-well connected themselves.

Guided by these facts, we develop a simple model where firms can outsource tasks and search for suppliers in different locations. Firms located in close proximity to other markets, and firms that face low search costs, will search more and find better suppliers. This in turn drives down the firm's marginal production costs. We test the theory by exploiting the opening of a high-speed train line in Japan which lowered the cost of passenger travel but left shipping costs unchanged.

We find compelling evidence that the supply network matters for firm performance. The infrastructure shock generated significant performance gains, especially for firms in industries that

[27] We also tested the hypothesis that the impact of distance is non-linear by including squared log distance. We found that the distance elasticity is smaller for higher distances, however the interaction terms were insignificant and close to zero.

have large shares of purchased inputs. We also provide evidence that these gains are related to new buyer-seller linkages as predicted by the model.

While there is a large literature on the link between infrastructure and improvements in regional economic outcomes, this paper provides the first direct evidence on the role of infrastructure on supply chains and firm performance. We highlight a novel transmission mechanism for the effects of improved infrastructure where reductions in search costs and buyer-seller inefficiencies allow firms to match with more and better suppliers, thus lowering the marginal cost of production. The resulting geographic variation in marginal costs for otherwise ex-ante identical firms yields systematic differences in economic activity across space. Firms in more geographically central locations have lower marginal costs and produce more.

This work has emphasized the role of domestic suppliers in explaining firm performance. Our results suggest that future research might fruitfully focus on how heterogeneous firms sort into locations, how reduced travel time affects intensive and extensive margins in both domestic and international buyer-supplier relationships, and how firm linkages form and evolve over time.

References

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, "The Network Origins of Aggregate Fluctuations," *Econometrica*, 2012, *80(5) September*, 1977-2016. 1

Amiti, Mary and Jozef Konings, "Trade Liberalization, Intermediate Inputs and Productivity: Evidence from Indonesia," *American Economic Review*, 2007, *97* (5), 1611-1638. 1

Antràs, Pol, *Global Production: A Contracting Perspective* 2014. 4.1, 4.2

_ **, Teresa C Fort, and Felix Tintelnot**, "The Margins of Global Sourcing: Theory and Evidence from U.S. Firms," 2014. 1, 4

Atalay, Enghin, Ali Hortacsu, James Roberts, and Chad Syverson, "Network Structure of Production," *Proceeding of the National Academy of Science*, 2011, *108(13)*, 5199-202. 5, 3

Bernard, Andrew B., Andreas Moxnes, and Karen Helene Ulltveit-Moe, "Two-sided Heterogeneity and Trade," Technical Report, CEPR Discussion Paper No. 9681 2013. 1, 3, 3.1, 3.1

Blum, Bernardo S., Sebastian Claro, and Ignatius J. Horstmann, "Import Intermediaries and Trade Costs: Theory and Evidence," mimeo, University of Toronto 2012. 1

Bøler, Esther Ann, Andreas Moxnes, and Karen Helene Ulltveit-Moe, "Technological Change, Trade in Intermediates and the Joint Impact on Productivity," 2014. 1

Carvalho, Vasco M., Makoto Nirei, and Yukiko U. Saito, "Supply Chain Disruptions: Evidence from Great East Japan Earthquake," Discussion Paper, Research Institute of Economy, Trade and Industry (RIETI) 2014. 1

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux, "The productivity advantages of large cities: Distinguishing agglomeration from firm selection," *Econometrica*, 2012, *80* (6), 2543-2594. 1

Comin, Diego, Mikhail A. Dmitriev, and Esteban Rossi-Hansberg, "The spatial diffusion of technology," 2012. 1

Coval, Joshua D. and Tobias J. Moskowitz, "Home Bias at Home: Local Equity Preference in Domestic Portfolios," *Journal of Finance*, December 1999, *54.* 2

_ **and** _ , "The Geography of Investment: Informed Trading and Asset Prices," *Journal of Political Economy*, August 2001, *109* (4). 2

Cristea, Anca, "Buyer-seller relationships in International trade: Evidence from U.S. States' exports and business-class travel," *Journal of International Economics*, 2011, *84*, 207-220. 1

Davis, Donald R. and Jonathan I. Dingel, "A Spatial Knowledge Economy," Technical Report, NBER Working Paper 18188 2012. 1

Donaldson, Dave, "Railrods of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, forthcoming. 1

Duranton, Gilles, Peter Morrow, and Matthew A. Turner, "Roads and Trad: Evidence from the US," *Review of Economic Studies*, 2013. 1

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, Sept 2002, *70* (5), 1741-1779. 4.1, 4.1

Fort, Teresa C, "Technology and Production Fragmentation: Domestic versus Foreign Sourcing," Technical Report, Tuck School of Business at Dartmouth 2014. 1

Giroud, Xavier, "Proximity and Investment: Evidence from Plant-level Data," *Quarterly Journal of Economics*, 2013, *128*, 861-915. 1, 21

Glaeser, Edward L. and David C. Maré, "Cities and Skills," *Journal of Labor Economics*, 2001, *19* (2), 316-42. 1

Goldberg, Pinelopi Koujianou, Amit Kumar Khandelwal, Nina Pavcnik, and Petia Topalova, "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India," *The Quarterly Journal of Economics*, November 2010, *125* (4), 1727-1767. 1

Halpern, Lazlo, Miklos Koren, and Adam Szeidl, "Imported Inputs and Productivity," CeFIG Working paper, Central European University 2011. 1

Hillberry, Russell and David Hummels, "Trade responses to geographic frictions: A decomposition using micro-data," *European Economic Review*, 2008, *52* (3), 527-550. 1

Jackson, Matthew O. and Brian W. Rogers, "Meeting Strangers and Friends of Friends: How Random Are Social Networks?," *American Economic Review*, 2007, *97* (3), 890-915. 3.1, 10, 11

Keller, Wolfgang and Stephen R Yeaple, "The Gravity of Knowledge," *American Economic Review*, 2013, *103(4)*, 1414-1444. 1

Klette, Tor Jakob, "Market power, scale economies and productivity: estimates from a panel of establishment data," *The Journal of Industrial Economics*, 1999, *47* (4), 451-476. G.1

Lerner, Josh, "Venture Capitalists and the Oversight of Private Firms,," *Journal of Finance*, March 1995, *50*, 301-318. 2

Oberfield, Ezra, "Business Networks, Production Chains, and Productivity: A Theory of Input-Output Architecture," Technical Report, Princeton University 2013. 1

Ohnishi, T., H. Takayasu, and M. Takayasu, "Network Motifs in Inter-firm Network," *Journal of Economic Interactions and Coordination*, 2010, *5(2)*, 171-180. 7

Olley, G. S. and A. Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64*, 1263-1297. 20, G.1

Petersen, Mitchell A. and Raghuram G. Rajan, "Does Distance Still Matter? The Information Revolution in Small Business Lending," *Journal of Finance*, December 2002, *57* (6), 2533-2570. 2

Redding, Stephen J. and Matthew A. Turner, "Transportation Costs and the Spatial Organization of Economic Activity," in "Handbook of Urban and Regional Economics" forthcoming, chapter Transportation Costs and the Spatial Organization of Economic Activity. 1

Saito, Yukiko U., Tsutomu Watanabe, and Mitsuru Iwamura, "Do larger firms have more interfirm relationships?," *Physica A: Statistical Mechanics and its Applications*, 2007, *383* (1), 158-163. 7

Serrano, Ma Angeles and Marian Boguna, "Topology of the World Trade Web," *The Accounting Review*, 2003, *68* (1). 11

Sveikauskas, Leo, "The productivity of cities," *The Quarterly Journal of Economics*, 1975, *89* (3), 393-413. 1

Uzzi, Brian, "The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect," *American Sociological Review*, 1996, *61(4)*, 674-698. 1

Table 1: Firm performance and distance to supplier.

	In-degree	Sales	Employment	Labor prod.
All firms:				
Distance	0.17***	0.19***	0.16***	0.03***
	(338.40)	(268.30)	(273.25)	(99.14)
R-sq	0.58	0.62	0.59	0.46
N	3,345,585	3,336,944	3,336,784	3,332,215
Single-plant firms:				
Distance	0.02***	0.02***	0.01***	0.01***
	(14.07)	(7.82)	(7.07)	(4.59)
R-sq	0.64	0.67	0.68	0.63
N	275,233	274,688	274,786	274,273
Buyer prefecture FE	Yes	Yes	Yes	Yes
Supplier FE	Yes	Yes	Yes	Yes

Note: Each column represents a dependent variable. t-statistics in parentheses. All variables in logs. Labor productivity is calculated as sales relative to the number of employees. 'All firms' refers to all firm pairs in the data. 'Single-plant firms' refers to firm-pairs where both buyer and seller are single-plant. *** significant at the 0.01 level, ** significant at the 0.05 level, * significant at the 0.1 level.

Table 2: Firm sales, in-degree, out-degree and distance to connections.

Dep. variable:	# sourcing cities	# customer cities	Median distance to suppliers	Median distance to customers	In-degree	Out-degree
Firm sales	0.31***	0.29***	0.02**	0.18***	0.36***	0.31***
	(49.92)	(23.37)	(2.17)	(14.36)	(34.14)	(22.06)
..×Single plant	0.00	0.00***	-0.01**	-0.01**	0.00	-0.00***
	(1.30)	(5.07)	(5.06)	(6.30)	(0.09)	(5.33)
..×Single industry	-0.01***	-0.01***	-0.01***	0.00	-0.02***	-0.01***
	(19.64)	(8.26)	(6.04)	(0.52)	(26.24)	(12.13)
# industries	401	401	401	401	403	403
# obs	429,160	429,160	428,891	428,891	441,156	441,156
R-sq	0.36	0.28	0.01	0.03	0.36	0.25

Note: Robust t-statistics in parentheses. All variables are in logs. Only firms with positive in- and out-degree are included in sample. JSIC 3 digit industry fixed effects included in all regressions. Single plant= 1 if the firm is a single-plant firm and zero otherwise. Single industry =1 if the firm only belongs to one 3-digit JSIC industry and zero otherwise (the data includes up to three 3-digit industries per firm). *** significant at the 0.01 level, ** significant at the 0.05 level, * significant at the 0.1 level.

Table 3: Firm Performance.

	(1) Sales	(2) Sales/employee	(3) TFPR
$Station_f \times H_j \times Post2004_t$	0.47**	0.42*	0.29**
	(2.12)	(1.76)	(2.44)
Firm and municipality controls	Yes	Yes	Yes
Prefecture-year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
# obs	148,264	146,466	145,058
# firms	18,068	18,068	18,018
R-sq	0.97	0.92	0.94

Note: Robust t-statistics in parentheses. Dependent variables are in logs and are measured relative to industry-year means. *** significant at the 0.01 level, ** significant at the 0.05 level, * significant at the 0.1 level.

Table 4: Firm Performance: Falsification test.

	(1) Sales	(2) Sales/employee	(3) TFPR
$Station_f \times H_j \times Post2000_t$	-0.30	-0.05	-0.02
	(1.05)	(0.22)	(0.17)
Firm and municipality controls	Yes	Yes	Yes
Prefecture-year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
# obs	66,756	66,756	66,487
# firms	14,165	14,165	14,158
R-sq	0.99	0.94	0.95

Note: Robust t-statistics in parentheses. Dependent variables are in logs and are measured relative to industry-year means. *** significant at the 0.01 level, ** significant at the 0.05 level, * significant at the 0.1 level.

Table 5: Firm Performance: Robustness I.

	(1) Sales	(2) Sales/ empl	(3) TFPR	(4) Sales	(5) Sales empl	(6) TFPR
$Station_f^{0-10} \times H_j \times Post2004_t$	0.60**	0.39*	0.29**			
	(2.28)	(1.80)	(2.57)			
$Station_f^{0-30} \times H_j \times Post2004_t$				0.47**	0.42*	0.29**
				(2.12)	(1.76)	(2.43)
$Station_f^{30-60} \times H_j \times Post2004_t$				0.08**	-0.13***	-0.09***
				(2.09)	(2.61)	(2.76)
Skill intensity controls	No	No	No	Yes	Yes	Yes
Firm and municipality controls	Yes	Yes	Yes	Yes	Yes	Yes
Prefecture-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
# obs	148,264	146,466	145,058	148,264	146,466	145,058
# firms	18,068	18,068	18,018	18,068	18,068	18,018
R-sq	0.97	0.92	0.94	0.97	0.92	0.94

Note: Robust t-statistics in parentheses. Dependent variables are in logs and are measured relative to industry-year means. *** significant at the 0.01 level, ** significant at the 0.05 level, * significant at the 0.1 level. $Station_f^{x-y} = 1$ if firm f is between x and y km from a new station. Columns (1)-(3): As baseline but use 10 km threshold instead of 30 km threshold. Columns (4)-(6): As baseline but add interactions with $Station_f^{30-60}$.

Table 6: Firm Performance: Robustness II.

	(1) Sales	(2) Sales /empl	(3) TFPR	(1) Sales	(2) Sales /empl	(3) TFPR
$Station_f \times H_j \times Post2004_t$	0.52**	0.43*	0.31***	0.44**	0.41*	0.28**
	(2.34)	(1.81)	(2.61)	(2.01)	(1.73)	(2.38)
$Station_f \times R\&D_j \times Post2004_t$				-0.05*	-0.02	-0.01
				(1.79)	(0.53)	(0.30)
Construction industry	No	No	No	Yes	Yes	Yes
Skill intensity controls	No	No	No	Yes	Yes	Yes
Firm and municipality controls	Yes	Yes	Yes	Yes	Yes	Yes
Prefecture-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
# obs	145,641	143,868	142,474	145,641	143,868	142,474
# firms	17,729	17,729	17,681	17,729	17,729	17,681
R-sq	0.97	0.92	0.94	0.97	0.92	0.94

Note: Robust t-statistics in parentheses. Dependent variables are in logs and are measured relative to industry-year means. *** significant at the 0.01 level, ** significant at the 0.05 level, * significant at the 0.1 level. Columns (1) to (3): Construction excluded. Columns (4) to (6): 2003 R&D indicator variable of industry as well as interactions with $Station_f$ and $Post2004_t$ included.

Table 7: Freight time, % change 2000 to 2010.

	Saga	Nagasaki	Kumamoto	Oita	Miyazaki	Kagoshima
Fukuoka	4	28	27	58	-25	**21**
Saga		-8	30	4	-52	5
Nagasaki			72	54	9	20
Kumamoto				88	-38	**24**
Oita					22	32
Miyazaki						-6

Note: Table shows the % change in freight time from 2000 to 2010 from the Net Freight Flow Census (NFFC) collected by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT). Numbers refer to average freight time across different modes (train, truck, air and sea).

Table 8: Connections: Descriptive statistics.

	Mean	Median	Std.dev.	min	max
C_{ij2005}	7.05	2	62.35	1	16507
C_{ij2010}	7.51	2	59.93	1	14808
$\Delta \ln C_{ij}$	0.08	0	0.56	-3.47	3.78
$Both_{ij}$	0.01	0	0.08	0	1
One_{ij}	0.02	0	0.14	0	1
Firms per cell	104.21	19	463.30	1	21,207
# sources	7,613				
# destinations	8,054				
# obs	386,294				

Table 9: Shinkansen: Growth in connections.

	(1)	(2)	(3)	(4)
$Both_{ij}$	0.07***	0.12***	0.39***	0.42***
	(5.91)	(7.91)	(20.12)	(7.93)
One_{ij}	-0.02***	-0.01	0.19***	0.15***
	(3.56)	(0.74)	(19.87)	(6.42)
$\ln Dist_{ij}$			-0.06***	-0.06***
			(71.32)	(81.98)
$Both_{ij} \times \ln Dist_{ij}$				-0.01
				(0.86)
$One_{ij} \times \ln Dist_{ij}$				0.01*
				(1.87)
Destination FE	No	Yes	Yes	Yes
Source FE	No	Yes	Yes	Yes
# obs	386,294	386,294	386,294	386,294
# sources		7,613	7,613	7,613
# destinations		8,054	8,054	8,054
R-sq	0.00	0.17	0.18	0.18

Note: Bootstrapped t-statistics in parentheses with 200 replications. Dependent variable is $\Delta \ln C_{ij} = \ln C_{ij2010} - \ln C_{ij2005}$. *** significant at the 0.01 level, ** significant at the 0.05 level, * significant at the 0.1 level.

Table 10: Shinkansen: Growth in connections. 10 km threshold.

	(1)	(2)	(3)	(4)
$Both_{ij}$	0.09***	0.14***	0.44***	0.43***
	(3.46)	(6.48)	(23.87)	(5.97)
One_{ij}	-0.05***	-0.01	0.20***	0.15***
	(6.02)	(1.07)	(23.58)	(4.76)
$\ln Dist_{ij}$			-0.06***	-0.06***
			(76.66)	(104.35)
$Both_{ij} \times \ln Dist_{ij}$				0.00
				(0.16)
$One_{ij} \times \ln Dist_{ij}$				0.01*
				(1.72)
Destination FE	No	Yes	Yes	Yes
Source FE	No	Yes	Yes	Yes
# obs	386,294	386,294	386,294	386,294
# sources		7,613	7,613	7,613
# destinations		8,054	8,054	8,054
R-sq	0.00	0.17	0.18	0.18

Note: Bootstrapped t-statistics in parentheses with 200 replications. Dependent variable is $\Delta \ln C_{ij} = \ln C_{ij2010} - \ln C_{ij2005}$. *** significant at the 0.01 level, ** significant at the 0.05 level, * significant at the 0.1 level.

Table 11: Shinkansen: Extensive margin connections.

$Both_{ij}$	0.71***	0.72***	0.69***
	(17.30)	(17.28)	(16.50)
One_{ij}	0.04***	0.05***	0.04***
	(9.71)	(7.87)	(6.03)
$\ln Dist_{ij}$	-0.26***	-0.26***	-0.26***
	(240.64)	(232.24)	(232.79)
Destination FE	No	No	Yes
Source FE	No	Yes	No
# obs	70,676,571	70,130,526	70,130,526
# sources		8,612	
# destinations			8,612
R-sq	0.00	0.00	

Notes: Coefficient estimates are multiplied by 100. t-statistics in parentheses. Dependent variable is $\Delta I\left[C_{ijt} > 0\right]$. *** significant at the 0.01 level, ** significant at the 0.05 level, * significant at the 0.1 level.

Appendix

A Optimal Search

The solution for $\bar{\tau}$ is

$$\bar{\tau}(z,j) = \kappa_1 \left(\frac{T}{w^{\sigma}} \frac{A}{f(j)} \right)^{1/\theta} \Phi(z,j)^{-k/\theta} z^{(\sigma-1)/\theta}.$$

Proof. The maximization problem of the firm is

$$\max_{\bar{\tau}} \{ \pi(z,j) - wf(j) n(z,j) \},$$

where $\pi(z,j) = Ap(z,j)^{1-\sigma}/\sigma$. A is the demand shifter for final goods, $A = w(1+\psi)\bar{L}Q^{\sigma-1}$, where Q is the CES price index for final goods. Given monopolistic competition and CES preferences, the firm charges a price that is a constant mark-up over marginal costs: $p(z,j) = \bar{m}w^{\alpha}P(z,j)^{1-\alpha}/z$, where $\bar{m} = \sigma/(\sigma-1)$. Inserting p and P into the profit function then yields

$$\pi(z,j) = \frac{\left(\bar{m}\lambda^{1-\alpha}\right)^{1-\sigma}}{\sigma} A w^{1-\sigma} \Phi(z,j)^{(\sigma-1)(1-\alpha)/\theta} z^{\sigma-1}.$$

The expressions for Φ and n are

$$\begin{aligned} \Phi(z,j) &= T_0 + \int_1^{\bar{\tau}(z,j)} T\tau^{-\theta} g(\tau,j)\, d\tau, \\ n(z,j) &= \int_1^{\bar{\tau}(z,j)} g(\tau,j)\, d\tau. \end{aligned}$$

Differentiating with respect to $\bar{\tau}$ yields

$$\begin{aligned} \frac{\partial \Phi(z,j)}{\partial \bar{\tau}} &= T\bar{\tau}^{-\theta} g(\bar{\tau},j), \\ \frac{\partial n(z,j)}{\partial \bar{\tau}} &= g(\bar{\tau},j). \end{aligned}$$

The first order condition is then

$$\frac{\left(\bar{m}\lambda^{1-\alpha}\right)^{1-\sigma}}{\sigma} A \frac{(\sigma-1)(1-\alpha)}{\theta} w^{1-\sigma} \Phi(z,j)^{(\sigma-1)(1-\alpha)/\theta - 1} z^{\sigma-1} T\bar{\tau}^{-\theta} = wf(j).$$

Rearranging,

$$\bar{\tau}(z,j) = \kappa_1 \left(\frac{T}{w^{\sigma}} \frac{A}{f(j)} \right)^{1/\theta} \Phi(z,j)^{-k/\theta} z^{(\sigma-1)/\theta},$$

where $\kappa_1 = \left(\frac{\left(\bar{m}\lambda^{1-\alpha}\right)^{1-\sigma}}{\sigma} \frac{(\sigma-1)(1-\alpha)}{\theta} \right)^{1/\theta}$ and $k = 1 - (\sigma-1)(1-\alpha)/\theta$.

The second order condition is

$$\frac{\left(\bar{m}\lambda^{1-\alpha}\right)^{1-\sigma}}{\sigma} A \frac{(\sigma-1)(1-\alpha)}{\theta} z^{\sigma-1} w^{1-\sigma} \left(-k\Phi^{-k-1} \left(\frac{\partial \Phi(z,j)}{\partial \bar{\tau}} \right)^2 + \frac{\partial^2 \Phi(z,j)}{\partial \bar{\tau}^2} \Phi(z,j)^{-k} \right) - wf(j) \frac{\partial^2 n(z,j)}{\partial \bar{\tau}^2} < 0.$$

Inserting the expressions for $\partial^2\Phi/\partial\bar{\tau}^2$ and $\partial n/\partial n^2$, this can be rewritten as

$$\frac{\left(\bar{m}\lambda^{1-\alpha}\right)^{1-\sigma}}{\sigma}A\frac{(\sigma-1)(1-\alpha)}{\theta}z^{\sigma-1}w^{1-\sigma}\Phi^{1-k}\frac{T\bar{\tau}^{-\theta}}{\Phi}\left(-\frac{k}{\Phi}T\bar{\tau}^{-\theta}g^2-\theta\bar{\tau}^{-1}g+g'\right)-wf(j)\,g'<0.$$

Using the first order condition, we know that the following must hold in optimum:

$$\pi(z,j)\,\underline{\chi}(z,j)=wf\frac{\theta}{(\sigma-1)(1-\alpha)},$$

where $\underline{\chi}$ is the trade share from the marginal location $\bar{\tau}$, $\underline{\chi}(z,j)=T\bar{\tau}(z,j)^{-\theta}/\Phi(z,j)$. This tells us that gross profits from the marginal location $\bar{\tau}$ equals the fixed search cost f multiplied by the factor $\theta/(\sigma-1)(1-\alpha)$. Exploiting this relationship gives us the second order condition

$$k\underline{\chi}(z,j)\,g(\bar{\tau},j)+\frac{\theta}{\bar{\tau}} \quad > \quad 0.$$

□

B Predictions of the Model

This section derives implications of the model described in Sections 4.3 and 4.5 of the main text.

B.1 The relationship between $\bar{\tau}$ and z

The cutoff $\bar{\tau}$ is increasing in z, $\partial\bar{\tau}/\partial z>0$.

Proof. Using equation (1), we have

$$\frac{\partial\Phi}{\partial z}=\frac{\partial\bar{\tau}}{\partial z}T\bar{\tau}^{-\theta}g(\bar{\tau},j)\,.$$

Using equation (2), we have

$$\frac{\partial\bar{\tau}}{\partial z} \quad = \quad \kappa_1\left(\frac{T}{w^{\sigma}}\frac{A}{f(j)}\right)^{1/\theta}\Phi(z,j)^{-k/\theta}\,z^{(\sigma-1)/\theta}\left(-\frac{k}{\theta}\Phi(z,j)^{-1}\frac{\partial\Phi}{\partial z}+\frac{\sigma-1}{\theta}z^{-1}\right).$$

Substituting in $\partial\Phi/\partial z$ and rearranging yields

$$\frac{\partial\bar{\tau}}{\partial z} \quad = \quad \frac{(\sigma-1)/z}{\theta/\bar{\tau}+k\bar{\chi}(z,j)\,g(\bar{\tau},j)},$$

which is positive given that the regularity condition $\theta/\bar{\tau}+k\bar{\chi}(z,j)\,g(\bar{\tau},j)>0$ holds. □

This also implies that $\partial o/\partial z>0$ because $\partial\Phi/\partial z>0$.

B.2 The relationship between $\bar{\tau}$ and $f(j)$

The cutoff $\bar{\tau}$ is decreasing in costs $f(j)$, $\partial\bar{\tau}/\partial f(j) < 0$.

Proof. Using equation (1), we have

$$\frac{\partial \Phi(z,j)}{\partial f(j)} = \frac{\partial \bar{\tau}}{\partial f} T\bar{\tau}^{-\theta} g(\bar{\tau}, j).$$

Using equation (2), we have

$$\frac{\partial \bar{\tau}(z,j)}{\partial f(j)} = -\frac{1}{\theta}\bar{\tau}(z,j)\frac{1}{f}\left(1 + k\frac{\partial \Phi}{\partial f}\frac{f}{\Phi(z,j)}\right).$$

Substituting in $\partial\Phi/\partial f$ and rearranging yields

$$\frac{\partial \bar{\tau}(z,j)}{\partial f(j)} = \frac{-1/f}{\theta/\bar{\tau} + k\bar{\chi}(z,j)\,g(\bar{\tau},j)},$$

which is negative given that the regularity condition $\theta/\bar{\tau} + k\bar{\chi}(z,j)\,g(\bar{\tau},j) > 0$ holds. This means that

$$\frac{\partial n(z,j)}{\partial f(j)} = \frac{\partial \bar{\tau}(z,j)}{\partial f(j)} g(\bar{\tau},j) < 0.$$

□

Note that we can also express

$$\frac{\partial \Phi(z,j)}{\partial f(j)}\frac{f(j)}{\Phi(z,j)} = -\frac{\underline{\chi}(z,j)\,g(\bar{\tau},j)}{\theta/\bar{\tau} + k\underline{\chi}(z,j)\,g(\bar{\tau},j)} < 0.$$

Furthermore, $\partial o/\partial f < 0$ because $\partial\Phi/\partial f < 0$.

B.3 Assortivity

The expected measure of buyers from j among suppliers in z's marginal market is decreasing in efficiency z.

Proof. The expected measure of buyers from j for a task ω in a location with trade costs τ (given the assumption of a unit continuum of tasks) is

$$\begin{aligned} c(\tau,j) &= m(j)\int_{\underline{z}(\tau,j)} \frac{T\tau^{-\theta}}{\Phi(z,j)}\lambda(z,j)\,dz \\ &= m(j)\,T\tau^{-\theta}\int_{\underline{z}(\tau,j)} \frac{\lambda(z,j)}{\Phi(z,j)}dz, \end{aligned}$$

where $\lambda(z,j)$ is the density of productivity in location j and $\underline{z}(\tau,j)$ is the minimum efficiency z required in order to source from a location with trade costs τ. The expected measure of buyers from j among suppliers in z_0's marginal market is therefore

$$\bar{c}(z_0,j) = m(j)\,T\bar{\tau}(z_0)^{-\theta}\int_{z_0} \frac{\lambda(z,j)}{\Phi(z,j)}dz.$$

We get

$$\frac{\bar{c}\left(z_0, j\right)}{\partial z_0}=-m\left(j\right) T \bar{\tau}\left(z_0\right)^{-\theta}\left[\frac{\theta}{\bar{\tau}\left(z_0\right)} \frac{\partial \bar{\tau}}{\partial z_0} \int_{z_0} \frac{\lambda\left(z, j\right)}{\Phi\left(z, j\right)} dz+\frac{\lambda\left(z_0, j\right)}{\Phi\left(z_0, j\right)}\right],$$

which is negative because $\partial \bar{\tau} / \partial z_0>0$ (see Section B.1). □

Hence, the average supplier in a downstream firm's marginal market is less well-connected if downstream productivity z_0 is higher.

B.4 The relationship between trade costs and expected productivity of a buyer

The expected productivity of a downstream buyer from j is increasing in trade costs between supplier and buyer.

Proof. The expected productivity of buyers from j for a task ω is

$$E\left[z\left(j, \tau\right)\right]=\int_{\underline{z}(\tau, j)} v\left(z\right) z \lambda\left(z, j\right) dz,$$

with weights

$$v\left(z\right)=\frac{\frac{T \tau^{-\theta}}{\Phi(z, j)}}{\int_{\underline{z}(\tau, j)} \frac{T \tau^{-\theta}}{\Phi(z, j)} \lambda\left(z, j\right) dz}=\frac{\Phi\left(z, j\right)^{-1}}{\int_{\underline{z}(\tau, j)} \Phi\left(z, j\right)^{-1} \lambda\left(z, j\right) dz}.$$

Differentiating with respect to τ yields

$$\frac{E\left[z\left(j, \tau\right)\right]}{\partial \tau} \quad = \quad \frac{\partial \underline{z}}{\partial \tau} \frac{\lambda\left(\underline{z}, j\right)}{\Phi\left(\underline{z}, j\right)} \frac{\int_{\underline{z}(\tau, j)} z \frac{\lambda(z, j)}{\Phi(z, j)} dz-\int_{\underline{z}(\tau, j)} \underline{z} \frac{\lambda(z, j)}{\Phi(z, j)} dz}{\left(\int_{\underline{z}(\tau, j)} \frac{\lambda(z, j)}{\Phi(z, j)} dz\right)^2}.$$

The sum of the two integrals in the numerator is positive because we integrate over $z>\underline{z}$. Moreover, $\partial \underline{z} / \partial \tau=1 /[\partial \bar{\tau} / \partial z]>0$, so that the $E\left[z\left(j, \tau\right)\right] / \partial \tau>0$. □

Hence, when comparing a supplier's connections in different markets, the expected productivity of a buyer is higher in markets with higher trade costs.

C Distributional Assumptions

Assume that τ is inversely Pareto distributed with support $[1, \tau_H]$ and shape $\gamma>\theta$: The density is $g\left(\tau\right)=\gamma \frac{\tau_H^{-\gamma}}{1-\tau_H^{-\gamma}} \tau_0^{\gamma-1}$ and the cdf is $G\left(\tau\right)=\left[\tau_H^{-\gamma} / \left(1-\tau_H^{-\gamma}\right)\right]\left(\tau_0^\gamma-1\right)$. An inverse Pareto captures the empirical fact that a location j has few nearby markets and many remote markets; we show in Appendix F that the inverse Pareto is a reasonable approximation of the empirical distribution of distance in our data. Note that a distribution with high upper bound τ_H first-order stochastically dominates a distribution with a low upper bound. Denote the two distributions (i) and (ii); then for $\tau_H^{(i)} > \tau_H^{(ii)}$, we have $G^{(i)}(\tau) < G^{(ii)}(\tau)$. This can be seen by differentiating the cdf:

$$\frac{\partial G}{\partial \tau_H} = -\frac{\gamma \tau_H^{-\gamma-1}}{\left(1-\tau_H^{-\gamma}\right)^2}\left(\tau^\gamma - 1\right) < 0.$$

In addition, we assume that a downstream firm's average productivity in task production, T_0, is related to the average cost of purchasing tasks in the marketplace as follows: $T_0 = \frac{T\tau_H^{-\gamma}}{1-\tau_H^{-\gamma}}\frac{\gamma}{\gamma-\theta}$.[28] Hence, a downstream firm cannot be too efficient in producing tasks itself, otherwise there would be no incentive to outsource. Given these additional assumptions, the hurdle $\bar{\tau}$, equilibrium market access Φ and measure of searched locations n are

$$\begin{aligned}
\bar{\tau}(z,j) &= \kappa_2 \left(\frac{A}{w^\sigma f(j)}\right)^{1/\omega} T^{(1-k)/\omega} z^{(\sigma-1)/\omega}, & (6)\\
\Phi(z,j) &= \frac{T\tau_H^{-\gamma}}{1-\tau_H^{-\gamma}}\frac{\gamma}{\gamma-\theta}\bar{\tau}(z,j)^{\gamma-\theta}, & (7)\\
n(z,j) &= \frac{\tau_H^{-\gamma}}{1-\tau_H^{-\gamma}}\left(\bar{\tau}(z,j)^\gamma - 1\right), & (8)
\end{aligned}$$

where $\omega = \theta + k(\gamma-\theta)$ and κ_2 is a constant.[29] The sourcing problem has an interior solution only if the second order condition, $\omega > 0$, is satisfied. Henceforth, we focus exclusively on the interior solution, i.e. $\omega > 0$.

D Propositions

Proposition 1 states that (i) Lower search costs $f(j)$ and trade costs τ_H lead to growth in sales among downstream firms in j. (ii) Sales growth is stronger in input-intensive (low α) relative to labor intensive (high α) industries.

Proof. Under the distributional assumption for $g(\tau, j)$, we get

$$\begin{aligned}
\frac{\partial \ln r(z,j)}{\partial \ln f(j)} &= -\frac{(\sigma-1)(1-\alpha)}{\theta}\frac{\gamma-\theta}{\omega} < 0,\\
\frac{\partial \ln r(z,j)}{\partial \ln \tau_H} &= -\frac{(\sigma-1)(1-\alpha)}{\theta}\frac{\gamma}{1-\tau_H^{-\gamma}} < 0,
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 \ln r(z,j)}{\partial \ln f(j)\,\partial(1-\alpha)} &= -\frac{\sigma-1}{\theta}\frac{\gamma-\theta}{\omega}\left(1+(1-\alpha)\frac{\sigma-1}{\theta}\frac{\gamma-\theta}{\omega}\right) < 0,\\
\frac{\partial^2 \ln r(z,j)}{\partial \ln \tau_H\,\partial(1-\alpha)} &= -\frac{\sigma-1}{\theta}\frac{\gamma}{1-\tau_H^{-\gamma}} < 0,
\end{aligned}$$

[28] Note that $\int_1^{\tau_H} T\tau^{-\theta} g(\tau,j)\,d\tau = \frac{T\tau_H^{-\gamma}}{1-\tau_H^{-\gamma}}\frac{\gamma}{\gamma-\theta}\left(\tau_H^{\gamma-\theta}-1\right)$, so T_0 equals $\int_1^{\tau_H} T\tau^{-\theta} g(\tau,j)\,d\tau / \left(\tau_H^{\gamma-\theta}-1\right)$.

[29] $\kappa_2 = \kappa_1^{\theta/\omega}\left(\frac{\tau_H^{-\gamma}}{1-\tau_H^{-\gamma}}\frac{\gamma}{\gamma-\theta}\right)^{-k/\omega}$

hence the elasticity of sales with respect to both fixed costs $f(j)$ and variable costs τ_H is negative, and the elasticity is more negative when $1-\alpha$ is high. $\square$

Proposition 2 states that lower search costs $f(j)$ and variable costs τ_H lead to more outsourcing $o(z,j)$ and suppliers from more locations $n(z,j)$ among downstream firms in j.

Proof. Section B.2 derives $\partial n(z,j)/\partial f(j) < 0$ and $\partial o(z,j)/\partial f(j) < 0$. Under the distributional assumption for $g(\tau,j)$, we get

$$\frac{\partial \Phi(z,j)}{\partial \tau_H} = -\Phi(z,j)\,\gamma \frac{\tau_H^{-1}}{1-\tau_H^{-\gamma}} < 0,$$

and therefore

$$\frac{\partial o(z,j)}{\partial \tau_H} = T_0 \Phi(z,j)^{-2} \frac{\partial \Phi}{\partial \tau_H} < 0.$$

Furthermore,

$$\frac{\partial n(z,j)}{\partial \tau_H} = -n(z,j)\,\gamma \frac{\tau_H^{-1}}{\left(1-\tau_H^{-\gamma}\right)} < 0.$$

Hence, both the share of tasks outsouced and the measure of locations searched increase as search costs $f(j)$ or variable costs τ_H fall. $\square$

E Sales per employee and TFPR

This section discusses the use of sales per employee (LP) and revenue productivity ($TFPR$) in the context of the model.

We start with sales per employee. By using the expression for the production function, sales per employee can be written as

$$LP = \frac{py}{l} = pz\left(\frac{v}{l}\right)^{1-\alpha}.$$

Inserting the first order condition, $v/l = [(1-\alpha)/\alpha]\, w/P$, we get

$$LP = \left(\frac{1-\alpha}{\alpha}\right)^{1-\alpha} pz \left(\frac{w}{P}\right)^{1-\alpha}.$$

Inserting the expression for prices, $p = \bar{m} k w^{\alpha} P^{1-\alpha}/z$, where $k = \alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)}$, we get

$$LP = \frac{\bar{m} w}{\alpha}.$$

Hence, sales per employee is constant across firms within the same industry. Note that if output prices p are sticky, then a fall in sourcing costs P lead to a rise in LP.

Revenue productivity is defined as

$$
\begin{aligned}
TFPR &= \frac{py}{(vP)^{1-\alpha} l^{\alpha}} \\
&= \frac{pz}{P^{1-\alpha}},
\end{aligned}
$$

where we in the second equality inserted the production function.[30] Inserting prices yields

$$TFPR = \bar{m}kw^{\alpha},$$

which is also constant across firms within the same industry. Note that if output prices p are sticky, then a fall in sourcing costs P lead to a rise in $TFPR$. Intuitively, this occurs because $TFPR$ controls for the value of inputs vP, instead of the quantity of inputs v, so that a fall in P translates into a rise in $TFPR$.

F The distribution of trade costs

This section provides empirical support for the assumption that trade costs are inversely Pareto distributed with density

$$g(\tau) \quad = \quad \gamma \frac{\tau_H^{-\gamma}}{1-\tau_H^{-\gamma}} \tau_0^{\gamma-1}. \tag{9}$$

Let distance d from location i be inversely Pareto distributed with support $[0, d_{Hi}]$ and shape parameter $\kappa > 0$. The cdf is

$$H_i(d) = \left(\frac{d}{d_{Hi}}\right)^{\kappa}.$$

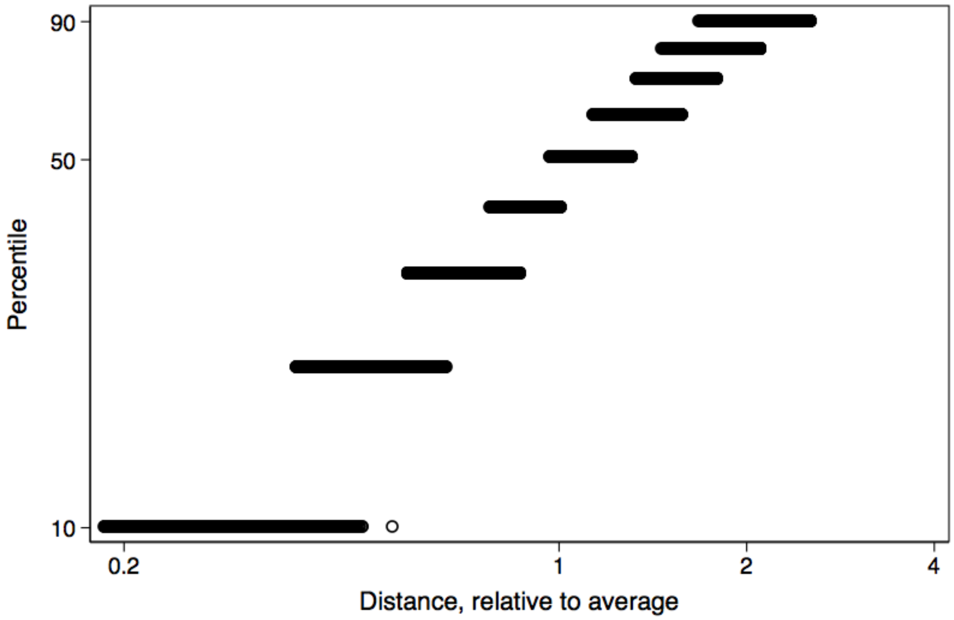
Consider the 500x500 grid dataset described in Section 5.2. We calculate distance for every location pair ij and the empirical distribution of distance for each location i. Due to the large number of location-pairs, we limit the calculations to the 1st, 2nd, .., 9th deciles of the distance distribution. From this, we obtain the k-th decile in location i, d_{ik}. If the distribution is inverse Pareto, the following must hold:

$$\ln H_{ik} = -\kappa \ln d_{iH} + \kappa \ln d_{ik}, \tag{10}$$

where H_{ik} takes the values 0.1 for $k = 1$, 0.2 for $k = 2$ and so on. The inverse Pareto should fit the data well if the relationship between H_{ik} and the d_{ik} is approximately log linear. Figure 9 plots H_{ik} against $d_{ik} - \bar{d}_i$, where $\bar{d}_i = (1/9)\sum_{k=1}^{9} d_{ik}$, on log axes. The normalization removes the constant term d_{iH} which may vary across locations. Overall, the relationship is close to linear, although there is clearly heterogeneity in the distribution across locations. Estimating equation (10) with location fixed effects produces a slope coefficient κ of 1.07.

[30] We have omitted capital in the definition of $TFPR$ because capital is not in the model. The analysis in this section would be similar if we included capital. When estimating $TFPR$, capital is controlled for, see Section G.1.

Figure 9: The cdf of distance across locations.



Note: Axes on log scales.

As is common in the literature, we assume that for large d, $\tau(d)$ is well approximated by the power law $\tau = (\alpha d)^{\rho}$ with $\alpha > 0$ and $\rho > 0$. Then τ inherits the distribution of d with shape parameter $\gamma = \kappa/\rho$. Because $\tau \geq 1$, the τ density has support $[1, \tau_H]$, which yields the expression of $g(\tau)$ given in equation 9.

G Data Appendix

G.1 Firm Performance

This section describes the measures of firm performance used in Section 5.1. As in Klette (1999), all firm level variables are demeaned relative to industry-year means, $\ln \hat{y}_{ijt} = \ln y_{ijt} - \overline{\ln y_{jt}}$, where $\ln \hat{y}_{ijt}$ refers to the demeaned variable for firm i in industry j at time t, $\ln y_{ijt}$ refers to the original log variable and $\overline{\ln y_{jt}}$ refers to the mean of the log variable in industry-year jt. The industry classification is 3-digit JSIC. Demeaning by industry-year has the benefit that it eliminates the need for deflating nominal variables; moreover it allows the technology of an industry to move freely over time.

$TFPR$ is estimated by the Olley and Pakes (1996) procedure. We estimate the gross production function

$$\ln Revenue_{it} = \beta_l \ln Labor_{it} + \beta_m \ln Materials_{it} + \beta_k \ln Capital_{it} + \omega_{it} + \eta_{it}, \qquad (11)$$

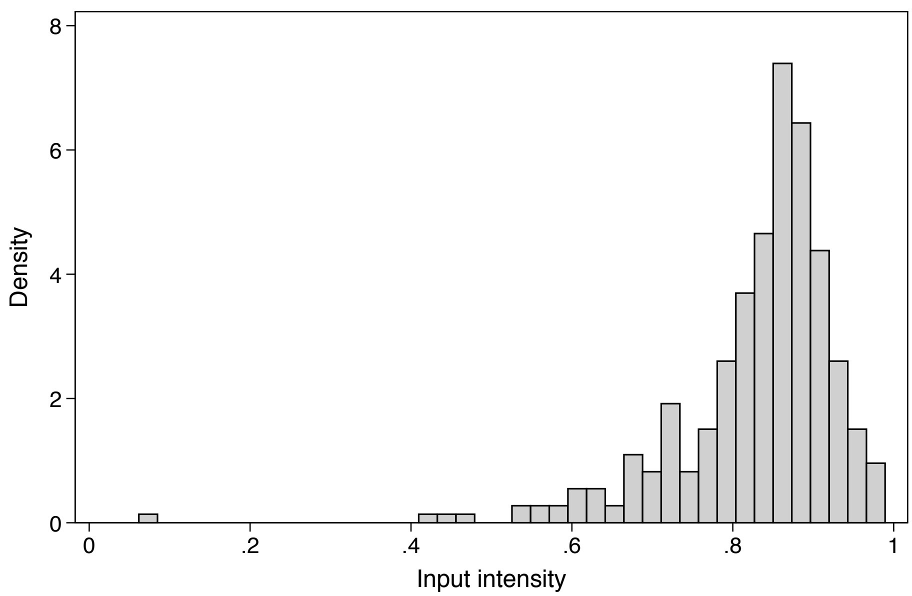
where ω_{it} is total factor productivity of the firm and η_{it} is either measurement error or a shock to productivity which is not forecastable during the period in which labor can be adjusted. After obtaining the estimates $\hat{\beta}_l$, $\hat{\beta}_m$ and $\hat{\beta}_k$, $TFPR$ is calculated by subtracting predicted output from actual output,

$$TFPR_{it} = \hat{\omega}_{it} = \ln Revenue_{it} - \widehat{\beta}_l \ln Labor_{it} - \widehat{\beta}_m \ln Materials_{it} - \widehat{\beta}_k \ln Capital_{it}.$$

G.2 Input intensity

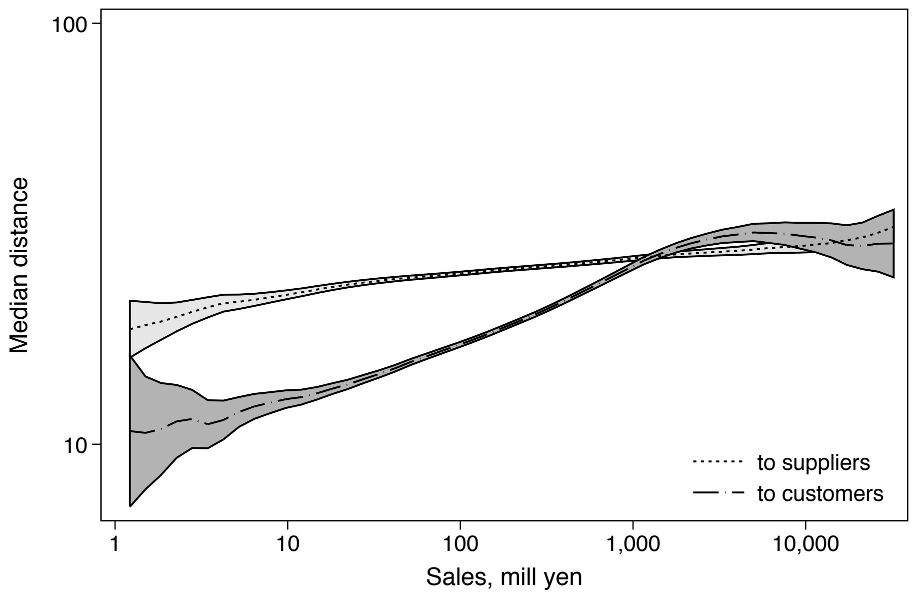
Input intensity H_j is calculated as input costs relative to total costs for each JSIC 3-digit industry j in year 2003. Specifically, denote WC_j total wage costs for industry j, $WC_j = \sum_{i \in j} \text{wage costs}_i$ and total costs $TC_j = \sum_{i \in j} \text{total costs}_i$. H_j is then $H_j = 1 - WC_j/TC_j$. Figure 10 shows the density of H_j across all 315 JSIC industries.

Figure 10: Density of input intensity H_j across industries.



Notes: 2003 data.

Figure 11: Size and median distance to connections: Single-plant firms



Note: 2005 data. The figure shows the kernel-weighted local polynomial regression of firm-level median log distance to the firm's connections (vertical axis) on log sales (horizontal axis). Firms with more than one plant are excluded. The two lines represent distance to suppliers and customers as separate regressions. Gray area denotes the 95 percent confidence bands. Sample is first trimmed by excluding the 0.1 percent lowest and highest observations of sales.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1434	Zack Cooper Stephen Gibbons Matthew Skellern	Does Competition from Private Surgical Centres Improve Public Hospitals' Performance? Evidence from the English National Health Service
1433	Nicholas Bloom Raffaella Sadun John Van Reenen	Management as a Technology?
1432	Andrew B. Bernard Toshihiro Okubo	Product Switching and the Business Cycle
1431	Fabrice Defever Alejandro Riaño	Protectionism through Exporting: Subsidies with Export Share Requirements in China
1430	Andrew B. Bernard Renzo Massari Jose-Daniel Reyes Daria Taglioni	Exporter Dynamics and Partial-Year Effects
1429	Cletus C. Coughlin Dennis Novy	Estimating Border Effects: The Impact of Spatial Aggregation
1428	Alan Manning	The Elusive Employment Effect of the Minimum Wage
1427	Decio Coviello Andrea Guglielmo Giancarlo Spagnolo	The Effect of Discretion on Procurement Performance
1426	Andrew B. Bernard Andreas Moxnes and Karen Helene Ulltveit-Moe	Two-sided Heterogeneity and Trade
1425	Stephen Machin Sandra McNally Martina Viarengo	"Teaching to Teach" Literacy

1424	Yatang Lin	Where does the Wind Blow? Green Preferences and Spatial Misallocation in the Renewable Energy Sector
1423	Andrew B. Bernard Valerie Smeets Frederic Warzynski	Rethinking Deindustrialization
1422	Stefano Gagliarducci Marco Manacorda	Politics in the Family: Nepotism and the Hiring Decisions of Italian Firms
1421	Jordi Blanes i Vidal Marc Möller	Team Adaptation
1420	Andrew B. Bernard J. Bradford Jensen Stephen J. Redding Peter K. Schott	Global Firms
1419	Marco Manacorda Andrea Tesei	Liberation Technology: Mobile Phones and Political Mobilization in Africa
1418	Jeremiah Dittmar Ralf R. Meisenzahl	State Capacity and Public Goods: Institutional Change, Human Capital and Growth in Early Modern Germany
1417	Holger Breinlich Swati Dhingra Gianmarco Ottaviano	How Have EU's Trade Agreements Impacted Consumers?
1416	Stefan Bender Nicholas Bloom David Card John Van Reenen Stefanie Wolter	Management Practices, Workforce Selection and Productivity

The Centre for Economic Performance Publications Unit
Tel 020 7955 7673 Fax 020 7404 0612
Email info@cep.lse.ac.uk Web site http://cep.lse.ac.uk