Production Networks, Geography and Firm Performance *

Andrew B. Bernard[†]
Tuck School of Business at Dartmouth, CEPR & NBER

Andreas Moxnes[‡]
Dartmouth College & NBER

Yukiko U. Saito[§]
RIETI

Preliminary, do not circulate. This Version: November 2013

Abstract

Firms operate in complex supplier-customer networks that potentially range over long distances. However, the effects of supplier networks and supplier location on firm performance are largely unknown. This paper characterizes the domestic production network in Japan using detailed buyer-supplier data on over 950,000 firms. Beyond describing the characteristics of the Japanese production network, the paper examines the geographic features of the network links. Greater geographic distance plays an important role in reducing the probability of buyer-seller relations between pairs of firms. For a given firm, greater distance is strongly associated with better performance measures of suppliers and customers. Geography, the density and the quality of network connections are strongly correlated with downstream (customer) firm performance. Labor productivity, size, credit score and sales growth of a downstream firm are strongly positively correlated with features of its upstream supply base including the number of suppliers and their average performance. In addition, geographic proximity of a firm's suppliers is associated with improved firm performance. The paper also provides the first evidence on the relationship between supplier network connections and downstream firm outcomes. Firm performance is better when its suppliers have more suppliers of their own. However, firm performance is lower when its suppliers are connected to more downstream customers.

Keywords: Suppliers, customers, gravity model, in-degree, out-degree, productivity
JEL codes: .

*Thanks go to Adam Kleinbaum, Yaniv Dover as well as seminar participants at DINR for helpful comments. We thank Keiko Ito for providing the HS-IO concordance data and 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 "XXX" undertaken at the Research Institute of Economy, Trade and Industry (RIETI).

[†]100 Tuck Hall, Hanover, NH 03755, USA; *tel:* +1 603 646 0302; *email:* andrew.b.bernard@tuck.dartmouth.edu
[‡]328 Rockefeller Hall, Hanover, NH 03755, USA; *tel:* +1 603 646 9345; *email:* andreas.moxnes@dartmouth.edu
[§]1-3-1 Kasumigaseki Chiyoda-ku, Tokyo 100-8901 Japan; *tel:* +81 3 3501-8312; *email:* saito-yukiko@rieti.go.jp

1 Introduction

In spite of the widespread perception that firms' success in part depends on their connections with suppliers and customers, 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. In this paper we use a comprehensive, unique data set on the production network in Japan. For the (large) majority of (big) firms in Japan, we can determine their location, suppliers, customers, owners and measures of performance. Our goal is to provide a new set of facts that can inform and discipline theory on the importance of geography and networks in firm outcomes.

While there has been an explosion of research on social and economic networks, their formation, and their effects on performance, 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. Our data provide supplier, customer and owner links between firms for over 950,000 firms in Japan for 2006. This set of firms accounts for the large majority of private sector economic activity in the country. The fundamental question motivating the research is whether firm performance is related to the characteristics of the supply network.

To address this question we must first understand what makes one network of suppliers better than another and how supplier networks interact with firm characteristics and location. While the data is extremely rich in terms of its coverage of firm connections, it does not allow us to see the value of sales between firms nor can we observe the evolution of the network over time.

We start by characterizing the features of the production network. As with many social and economic networks, the Japanese production network has feature of a "small worlds" network with short average path lengths between nodes and most firms belonging to the same component. The distribution of buyers per supplier and the number of suppliers per customer are well characterized by a Pareto, scale-free distribution except for relatively small deviations in the extreme tails of the distribution. The production network display negative assortivity, the trading partners of well-connected firms on average are less-well connected. This feature of big firms matching more often on average with smaller firms also is found in exporter-importer networks in international trade.

We pay particular attention to the geography of the domestic production network. Physical proximity plays an important role in the matching of suppliers and customers. Firms are more likely to form connections if they are in the same prefecture, on the same island and are close in terms of distance. Firm performance and distance are linked in many dimensions. Better firms have suppliers and customers that on average are farther away; the better the firm the greater the value it brings to a supplier or customer relationship and the more likely it is to cover the costs of transacting at a distance. This relationship between distance and firm quality also is present within

firms. For a given downstream (upstream) firm, the quality of its suppliers (customers) is increasing in distance.

The characteristics of supplier networks themselves are correlated with firm outcomes. High performing firms have more suppliers and higher average quality of those suppliers. Controlling for the number and quality of suppliers, downstream firm performance negatively correlated with average distance to suppliers. The position of a firm's suppliers in the overall network is also correlated with firm outcomes. Performance of a downstream firm is higher when its upstream suppliers have more suppliers of their own after controlling for the number of direct suppliers, their average quality and their geographic proximity. In contrast, firm outcomes are worse if suppliers are well-connected with other downstream firms. If a firm's suppliers also sell to many other customers then sales, labor productivity, credit scores and sales growth are lower.

2 Data

The data employed in this paper 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 for 2006 contains information on more than 950,000 firms, represents more than half of all the firms in Japan and covers all sectors of the economy.

Each firm provides rank-ordered lists of the most important suppliers (up to 24), customers (24), and shareholding firms (24).[1] TSR also collects information on employment, the number of establishments, the number of factories, up to three (4-digit) industries, three years of sales and profits and a physical address. In addition, the database records TSR's credit score for the firm.

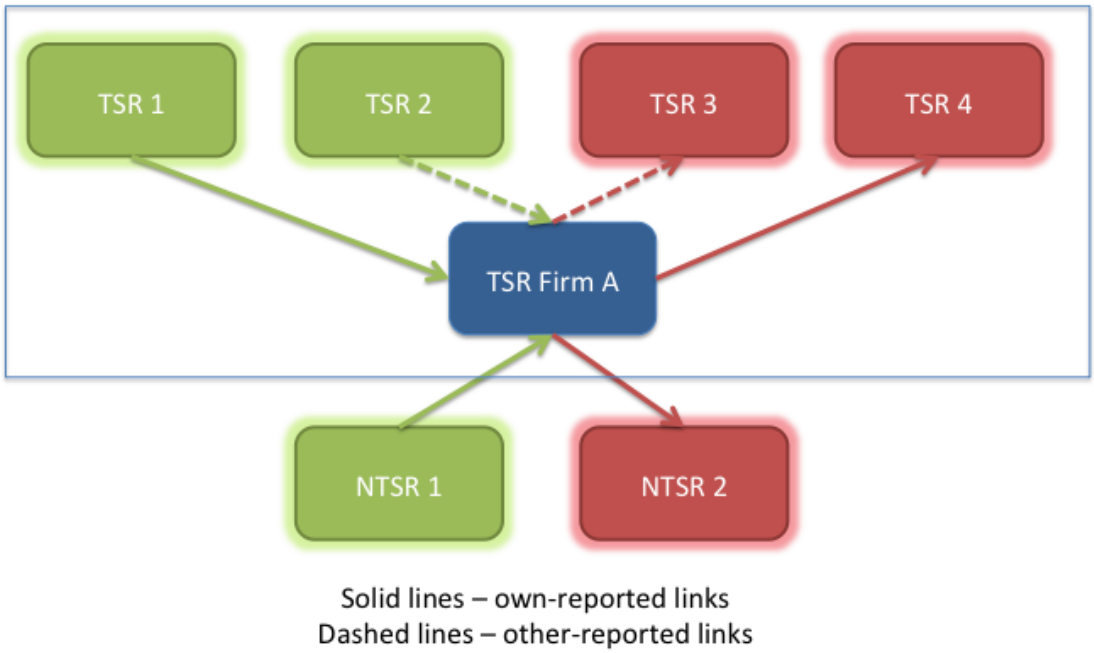
Using an address matching service provided by the Center for Spatial Information Science at the University of Tokyo, we are able to match a firm's address to longitude and latitude data.[2] The top 3 prefectures by counts of firms are Tokyo, Osaka and Aichi (Nagoya) while the top three 2-digit industries are General Construction Work, Specialist Construction Work and Equipment Installation. We use the geo-coded data to create a measure of great circle distance between firms.[3]

[1] The firms also report up to three banks with information on specific bank branches although we do not use the bank information in this paper.

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

[3] We recognize that actual distance required to travel between two points may be poorly approximated by great circle distances from geographic coordinates due to the distinct topography of Japan. Developing a more precise set of distance measures is a topic for future work.

Figure 1: Supplier and Customer Connections: An Example



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.[4] However, the 24-firm limit for suppliers, customers and owners 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. On average, two-way reporting is relatively rare, only 5.3% of A's own-reported customers also report A as a supplier. Importantly, some firms that are suppliers and customers are outside the TSR set of firms (NTSR), i.e. they are indeed 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's 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

[4]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.

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).

Many 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.[5] 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 Domestic Production Network

In this section we begin to explore the domestic production network in Japan.[6] 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. For firms with positive out-degree, the mean number of customers is 5.6 and the median is one.

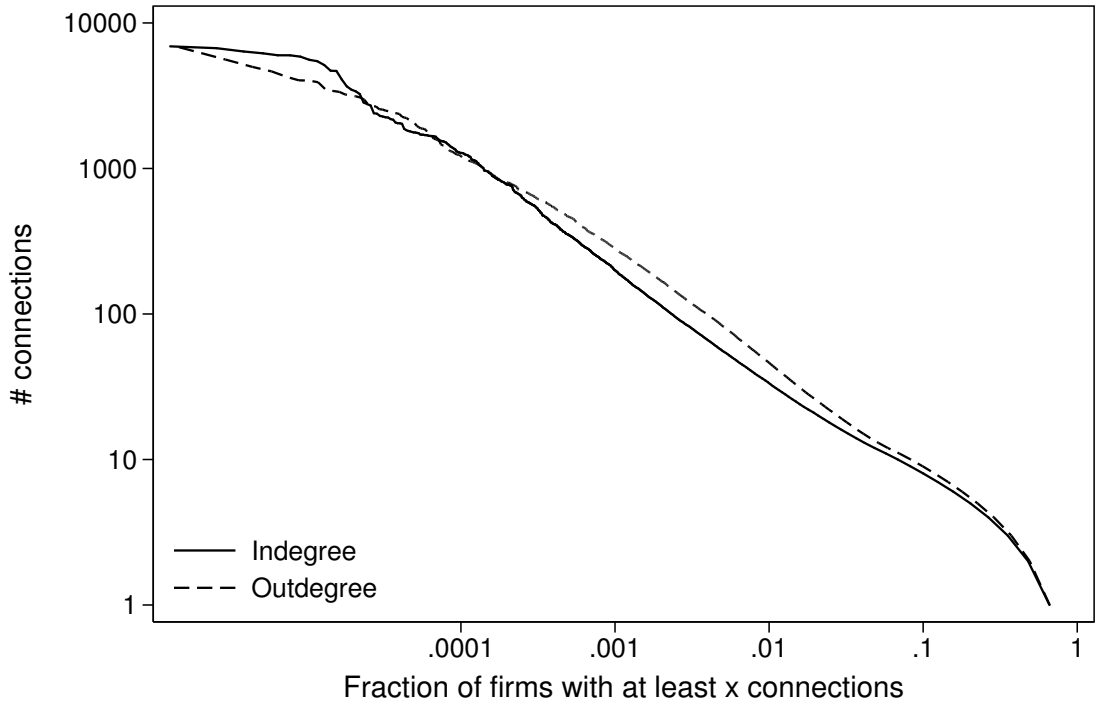
The cdf of the in-degree and out-degree distributions is given in Figure 2. The distributions are well-approximated by a Pareto (power law) distribution. The estimated Pareto shape parameter is -1.32 for the in-degree distribution and -1.50 for the out-degree distribution. Deviations from the Pareto are found in the extreme tails of the distribution. Firms with very large numbers of connections are somewhat under-represented while firms with small numbers of 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.

In other respects the domestic production network in Japan has many characteristics common

[5] 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.

[6] Some of these network characteristics are also presented in Saito et al. (2007), Ohnishi et al. (2010) and ...

Figure 2: In-degree and Out-degree CDFs



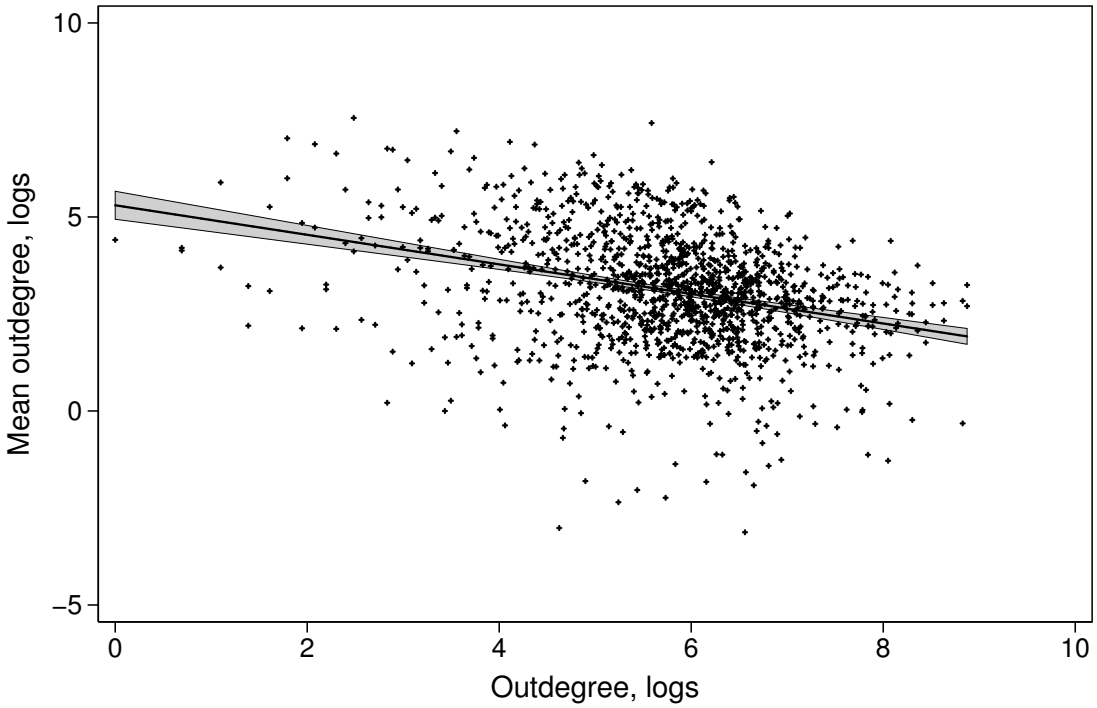
to a variety of socio-economic networks. It is characterized by features of a "small world" network as the mean, median and maximum path length are relatively small, 4.7,4, and 17 respectively. 99 percent of the nodes belong the same component. Clustering coefficients are quite small, total clustering is 0002 while average clustering is 0.064.

3.1 Assortivity

One distinguishing feature of networks is the extent to which a well-connected node is linked to other well-connected nodes, known as assortivity. While there is an extensive body of research on assortivity in technical and social networks, these relationships are less well documented in economics networks. We characterize sellers according to their number of buyers, and buyers according to their number of sellers. We find that the better connected a seller, the less well-connected is its average buyer. Figure 3 provides an overview of out-degree assortivity in the Japanese production network. The figure shows all possible values of the number of customer per Japanese firm, a_j, on the x-axis, and the average number of (customer) connection of these buyers, $b_j(a_j)$, on the y-axis. Both variables are in logs and demeaned.[7] The interpretation of a point with the coordinates (0.2,-0.2) is a Japanese firm that has 20 percent more customers than average sells to firms with 20 percent

[7]This Figure shows $\ln b_j(a_j) - \ln \bar{b}_j(a_j)$, where $\ln \bar{b}_j(a_j)$ is the average number of connections among all buyers in j.

Figure 3: Assortivity



fewer customers than average. The fitted regression line has a slope of -0.38, so a 10 percent increase in number of customers is associated with a 3.8 percent decline in average connections among the customers.

This results suggests that the best firms, those with the most customers, are selling to firms who on average have fewer customers 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).[8] In recent papers, Bernard et al. (2013) and Lu et al. (2013) also find negative assortivity between trading firms using Norwegian and Colombian importer-exporter trade data.

One concern is that the finding of negative 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 by asking whether firms in Tokyo (i) with many connections in Osaka (j) trading with less well-connected firms in Osaka. We include fixed effects for both the location of firm as well as its suppliers,

$$\overline{Outdegree}_{ij} = \alpha_i^1 + \alpha_j^2 + \beta Outdegree_{ij} + \epsilon_{ij} \qquad \beta = -0.096 \quad s.e.(0.002)$$

[8] 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.

Again we find negative assortivity, firms with more customers in a destination market are selling to firms with fewer customers. Controlling for destination countries, Bernard et al. (2013) estimate a comparable coefficient of 0.13 when considering the Norwegian exporter data. Our findings of negativity assortivity is not limited to measures of out-degree. We find similar relationships when using measures of in-degree or an undirected measure of the total number of connections.

3.2 Ownership

The TSR data is unique is that it not only indicates the directed links in the production network in Japan but it also includes information on the cross-ownership of the firms. There are over 165,000 ownership links involving 13.8% of the TSR firms. Across all firms, the mean (median) number of owners is 0.16 (0) but the average number of owners for firms with positive ownership links is 1.46. 2.3% of the supplier-customer connections involve an ownership link. If ownership were assigned randomly to firm pairs, the expected number of supplier-customer pairs with ownership links would be 0.0005 percent. Reversing the perspective, 48.8% of the ownership links also involve a direct supplier-customer link with the supplier and customer equally likely to be the owner.

4 Geography of the Production Network

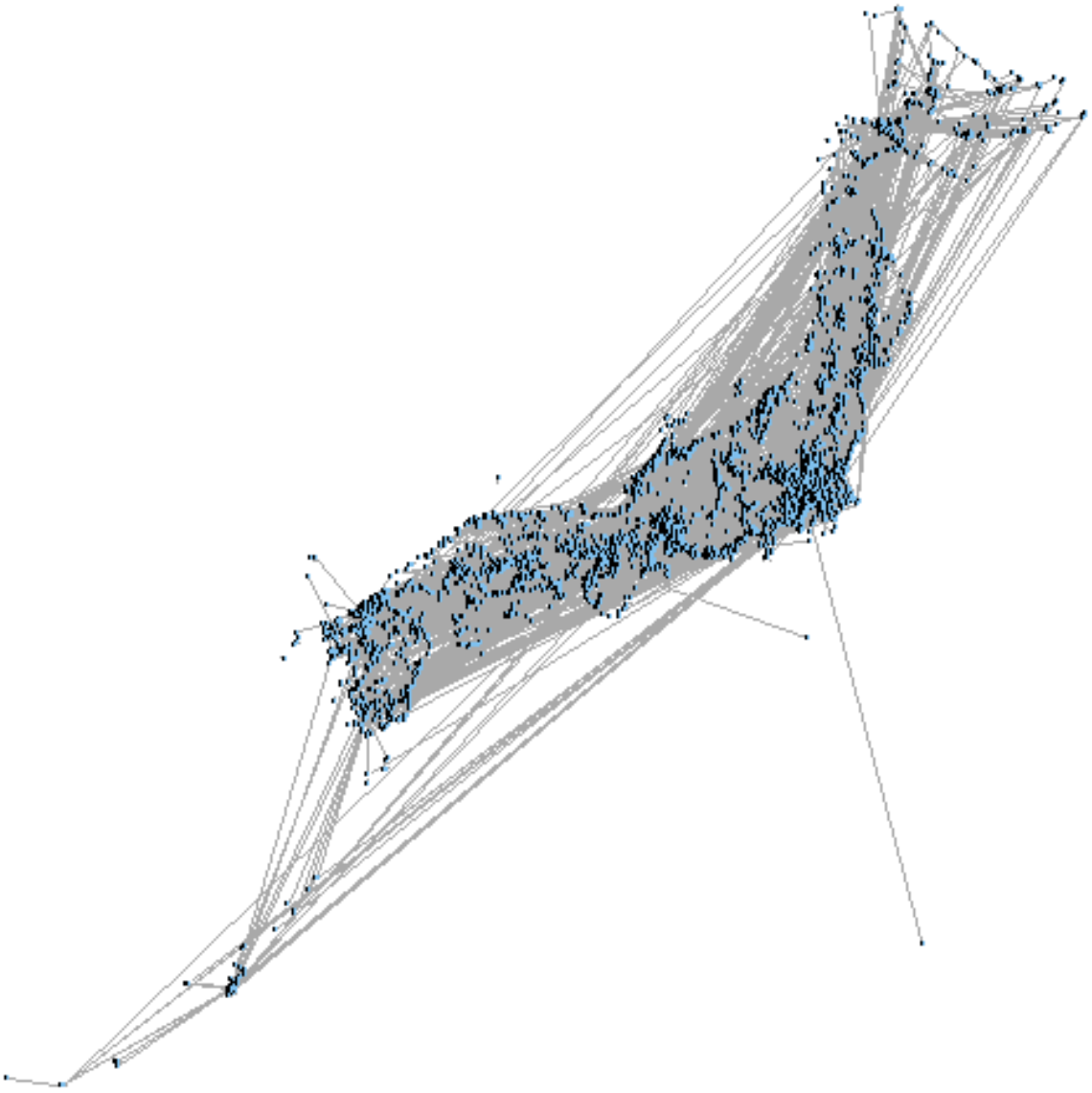
In this section, we explore the role of geography in the production network. First, we visualize the network using the geocodes associated with each firm. Second, we document that connections are typically formed with local suppliers and customers. Third, we show that only higher performance firms tend to form connections to geographically more remote suppliers and customers. These findings suggest that firms benefit the most from geographically local production networks, and that only the most productive firms have the capability to generate positive profits from a geographically dispersed supply chain.[9]

4.1 A Production Network Map

In Figure 4, we use the geocodes associated with each firm to plot a map of supplier connections. We draw 10,000 linked pairs at random and plot each end of the connection in physical space (latitude and longitude). The resulting network shows a striking resemblance to the population density map of Japan. There are relatively few connections between firms at the ends of the country while most supplier relationships are between firms that are physically proximate.

[9] Our findings are also consistent with the predictions of Bernard, Moxnes and Ulltveit-Moe (2013).

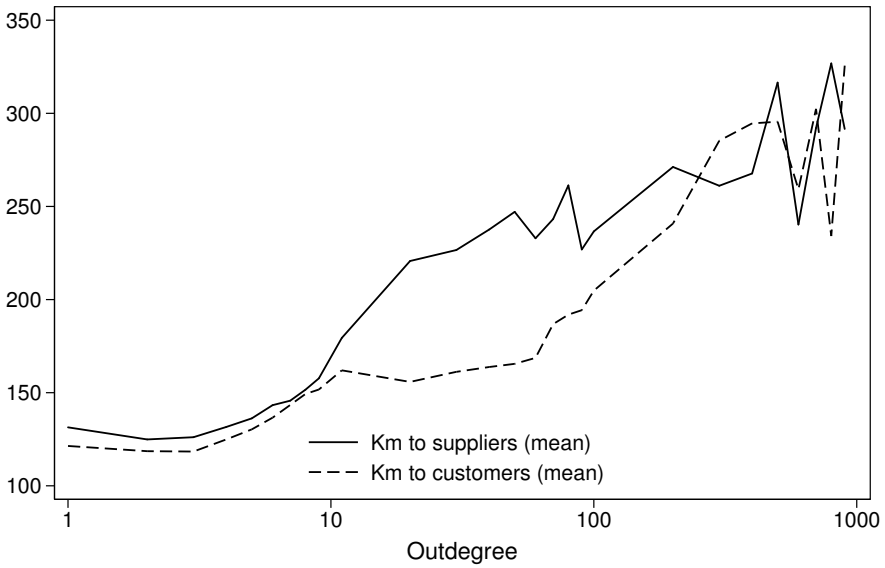
Figure 4: The Geography of the Production Network



4.2 Local Versus Remote Connections

We start by calculating the distance between any supplier-customer pair ij. 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. Next, we explore how connection distance varies with firm performance. Figure 5 plots the number of customers per firm on the horizontal axis (out-degree) and average distance to suppliers and customers on the vertical axis. Better connected firms tend to have more remote connections; distance to suppliers and customers is around twice as high for firms with more than 100 connections compared to firms with 1-10 connections.

Figure 5: Out-degree and Average Distance.



Note: The figure shows out-degree x on the horizontal axis and average distance to suppliers and customers across firms with out-degree x on the vertical axis. Out-degrees>10 are bins of 11-20, 21-30, .., 101-200, 201-300, .., 1000+.

4.3 Connection Quality and Distance

The previous section showed descriptive statistics indicating that high performance firms tend to form more remote connections. Here we provide additional evidence from a set of firm-level regressions. Consider the regression

$$q_{s(b)} = \alpha_b + \beta \ln Distance_{s(b)} + \gamma X_{s(b)} + \epsilon_{s(b)}, \tag{1}$$

where b is buyer, $s(b)$ is a supplier of b, $q_{s(b)}$ is a measure of firm s performance, α_b is a buyer fixed effect, $X_{s(b)}$ is a vector of controls and $\epsilon_{s(b)}$ is an error term. Firm performance $q_{s(b)}$ is either in-degree, out-degree, employment or sales per worker. The firm controls $X_{s(b)}$ are dummies for whether b and s are in the same industry, on the same island, in same prefecture. We also include supplier prefecture fixed effects. Hence, we compare two suppliers of firm b, in the same industry and prefecture and ask whether the long distance supplier has better firm performance $q_{s(b)}$ than the short distance supplier. Next, consider the regression

$$q_{b(s)} = \alpha_s + \beta \ln Distance_{b(s)} + \gamma X_{b(s)} + \epsilon_{b(s)}. \tag{2}$$

Here we instead condition on a seller s, and ask whether her long distance customers are associated with higher firm performance compared to her short distance customers.

The results are shown in Table 3. The upper panel uses buyer fixed effects, while the lower panel uses seller fixed effects. For every firm performance measure (in-degree, out-degree, employment and sales per worker), long distance suppliers and customers are associated with better firm performance.

Robustness. A potential concern is that that distance between supplier and customer is not always well defined for multi-plant firms.[10] We therefore remove all multi-plant firms from the sample, and re-estimate the two regressions.[11] Another concern is that the role of distance may vary across sectors. We therefore re-estimate the two regressions for four different data subsets: manufacturing (seller) - manufacturing (buyer), non-manufacturing (seller) - manufacturing (buyer), manufacturing (seller) - non-manufacturing (buyer) and non-manufacturing (seller) - non-manufacturing (buyer). All of these robustness checks produces very similar results compared to the baseline case.[12]

5 Probability of Links

A key issue in the study of social and economic networks is how links form and how the network evolves over time. Much of the literature on network formation considers link formation through preferential and random attachment.[13] Often researchers have only a single cross section of the network at a point in time and estimate the relative contributions from the degree distribution function. Our focus is somewhat different as we are interested in how firm characteristics and geography are correlated with the existence of supplier-buyer links in the Japanese production network.

We would like to estimate the relationship between firm characteristics, geography and the existence of a link. However, there is a problem due to the size of the network and the extremely large number of possible links. With $n = 770,000$ firms, there are $\left(n^2 - n\right)/2 \approx 296$ billion potential connections in an undirected network. We employ a solution suggested by Kleinbaum (2012) and include all firm ij pairs with links and an equal sized random sample of pairs with no link.[14] We then estimate a linear probability model where the dependent variable $Connect_{ij} = 1$ if firm i supplies firm j and is zero otherwise. The observations without links are weighted by the inverse of the sampling probability. In the data, the weight is 1 for connected pairs and ∼100,000 for non-connected pairs.

$$Connect_{ij} = \beta_0 + \beta_1 X_i + \beta_2 X_j + \beta_3 X_{ij} + \epsilon_{ij} \tag{3}$$

[10] The geolocation is based on the firm's headquarters.

[11] Removing multi-plant firms reduces the sample by 82 percent.

[12] Detailed results available upon request.

[13] Preferential attachment refers to the probability that a node is more likely to form a link with another node that is highly connected. Under random attachment, link formation does not depend on the degree of the node.

[14] See also King and Zeng (2001) and Imbens (1992).

We include measures of firm size (log employment) for both firm i and firm j as well as a dummy variable if both i and j are in the same 4-digit industry.[15] In some specifications, we include the interaction between firm size and dummies for cross-ownership, one variable if i owns j and another if j owns i. The geographic variables include the log of the great circle distance between i and j, a dummy if the firms are on the same island, and a dummy if they are in the same prefecture.

The findings are presented in Table 1. In the top panel, we include specifications with supplier (firm i) fixed effects and in the bottom panel we include customer (firm j) fixed effects. Potential partner size is positive and significant throughout. Supplier-buyer relationship are more frequent when the potential partner is larger. However, using only within-firm variation, we see that the effect of partner size is less important when the firm is large, e.g. the probability of selling to a large customer is decreasing in the size of the supplier (column 3). This finding mirrors the negative assortivity found in section 3.1, large firms connect on average with smaller partners. A similar pattern holds looking across suppliers controlling for customer fixed effects.

The relationship between ownership and the probability of a supplier-customer link is more complex. In the top panel, looking across potential buyers within a supplier, we find that the probability of a link is significantly higher if the supplier owns the customer. However, the probability of a link is lower if the customer owns the supplier. In the bottom panel, looking across potential suppliers for a given customer, we find that the probability of a link is significantly lower when there is any cross-ownership of the firms.

Geography is strongly correlated with the existence of buyer-supplier connections. The probability of a link falls with the distance between the firms. Even after controlling for distance, firms on the same island and in the same prefecture are more likely to be connected. We explore the implications of distance on the production network in more depth later n the paper.

5.1 Two-way Trade

Goods and services can flow both ways between a pair of firms. Such two-way trade occurs in 7.3 percent of the supplier-customer pairs, far more often that would occur by chance. To assess the importance of firm and geographic variables in two-way trade, we estimate a linear probability model

$$Twoway_{ij} = \beta_0 + \beta_1 X_i + \beta_2 X_j + \beta_3 X_{ij} + \epsilon_{ij} \tag{4}$$

where $Twoway_{ij} = 1$ when there is two-way trade and $Twoway_{ij} = 0$ when firm i supplies firm j but is not a customer of firm j. The covariates are similar to those in equation 3 with the exception of ownership which is given by a single dummy variable.

[15] IO tables across countries show large amounts of within industry trade.

Table 2 reports the results. Among firm i's customers, firm j is more likely to also be a supplier to firm i if it is large and in the same industry. However the interaction partner sizes is again negative. Cross-ownership enters with a large and significant coefficient. For a given supplier, an ownership connection increases the probability of two-way trade by 28 percent. Customers with an ownership links are three times as likely to be involved in two-way trade than are other customers of the firm.

The role of geography is somewhat different than in equation 3. Distance between the firms reduces the probability of two-way trade but suppliers and customers in the same prefecture are less likely to be involved in two-way trade. The results are symmetric looking across firm j's suppliers.

6 Gravity in the Domestic Production Network

In this section, we evaluate to what extent connections between regions of Japan is consistent with the predictions of the gravity equation for trade flows. The gravity model predicts that trade between regions i and j is a multiplicative function of the economic sizes of i and j and trade costs between i and j. Specifically,

$$X_{ij} = \kappa \frac{Y_i X_j}{\tau_{ij}^{\rho}},$$

where X_{ij} is trade between i and j, κ is a constant, $X_j = \sum_i X_{ij}$ is total purchases by j, Y_i is production in region i, τ_{ij} is a measure of trade costs between i and j and ρ is the elasticity. We focus on three main predictions of model. First, total imports in region j are proportional to total purchases, $Imports_j = \sum_{i \neq j} X_{ij} = \kappa X_j \sum_{i \neq j} \frac{Y_i}{\tau_{ij}^{\rho}}$ (Prediction 1). Second, j's share of purchases from i is increasing in i's production, $X_{ij}/X_j = \kappa Y_i / \tau_{ij}^{\rho}$ (Prediction 2). The relationship also depends on τ_{ij}, however, so that conditional on production Y_i, regions with lower trade costs have a higher trade share. Third, the so-called Head-Ries index,

$$HR_{ij} = \left(\frac{X_{ij} X_{ji}}{X_{ii} X_{jj}} \right)^{1/2},$$

is a function of trade costs alone, $HR_{ij} = \tau_{ij}^{-\rho}$ (Prediction 3).[16] We therefore expect that distance, a standard proxy for trade costs, should be highly correlated with the Head-Ries index. Also note that, given information about the elasticity ρ, HR_{ij} gives us a simple estimate of trade costs τ_{ij}.

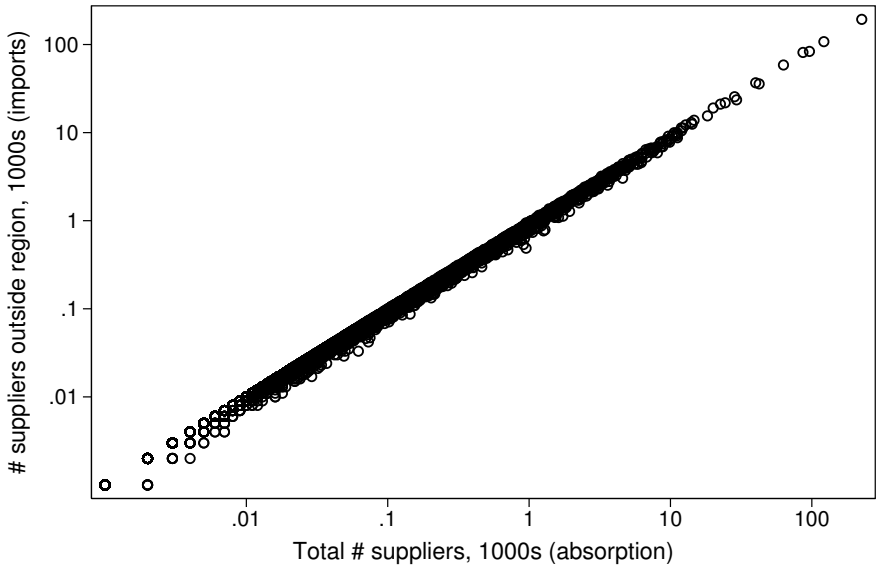
The dataset allows us to explore the gravity features of the data at any aggregation level. We start with highly disaggregated regions and divide Japan into a grid consisting of 500×500 cells. The data do not include information about the value of trade, but we do observe the number of links between regions. Hence, we define X_{ij} as the number of suppliers in i serving customers in j. As many cells are covering water and other non-populated areas, the dataset is reduced to 8,390

[16] Given $\tau_{ii} = 1$ and $\tau_{ij} = \tau_{ji}$. Without these restrictions, the expression is $HR_{ij} = \left[(\tau_{ij}\tau_{ji}) \,/\, (\tau_{ii}\tau_{jj}) \right]^{-\rho/2}$.

cells after removing these regions. Furthermore, not all ij pairs are trading; 565,929 pairs have positive X_{ij} among the number of potential pairs $8,390^2 = 70,392,100$. In the appendix, we repeat the analysis at the prefecture level, which gives us 47 regions and $47^2 = 2209$ ij pairs (no pairs with zero trade).

Prediction 1. We start by plotting the relationship between the total number of suppliers ($\sum_i X_{ij}$) and suppliers from outside the region ($\sum_{i \neq j} X_{ij}$) on log scales in Figure 6. According to Prediction 1 the two variables should be proportional. This is indeed the case; the fitted slope coefficient is 0.98, close to a coefficient of 1 as the theory would predict.

Figure 6: Outside and Total Number of Suppliers.



Note: The figure shows the total number of suppliers to a given region on the horizontal axis and the number of suppliers from outside the region on the vertical axis. Numbers are in 1000s and axes are in logs.

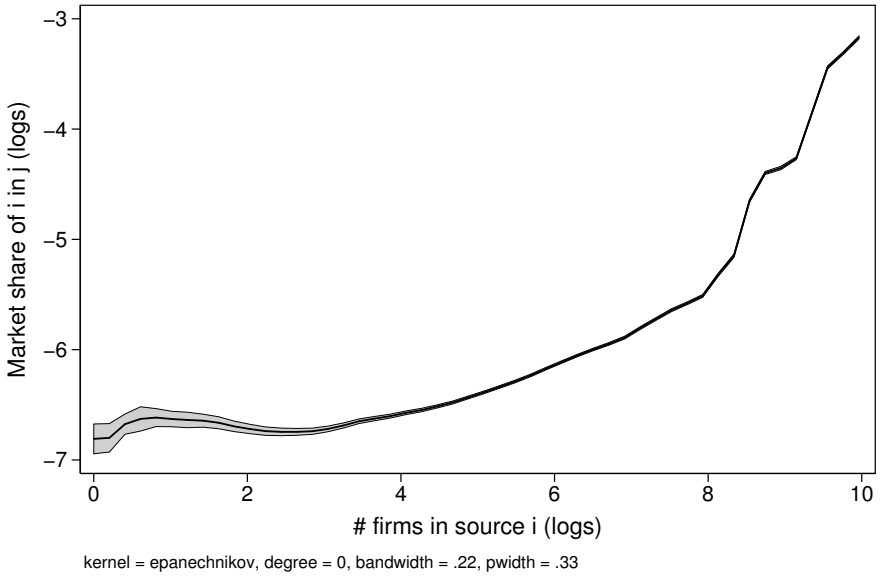
The mean and median import share, $\sum_{i \neq j} X_{ij}/X_j$, is 0.94 and 0.95 respectively, showing that the large majority of connections is formed to outside cells. We we will see below, the import share is much lower when regions are aggregated to larger geographic areas.

Prediction 2. The second prediction relates total production in i to the market share of i in j's purchases. We proxy production in i by the the number of firms operating in i. We then estimate a kernel-weighted local polynomial regression of the log market share of i in j, X_{ij}/X_j on the log number of firms in the source i.[17] The fitted line is displayed in Figure 7. Overall, larger regions

[17]We drop jj pairs from the regression as X_{ij}/X_j is typically much higher for $i = j$ than for $i \neq j$. Due to the large number of observations in the data, we do not show the scatter behind the fitted line.

have a greater market share in other regions. The relationship is somewhat stronger for larger regions in the sample.

Figure 7: Trade Share and Number of Source Firms.



Note: The figure shows the kernel-weighted local polynomial regression of the log market share of i in j, X_{ij}/X_j (vertical axis), on the log number of firms in the source region i (horizontal axis). The gray area denotes the 95 percent confidence bands. The linear regression slope is 0.32.

Prediction 3. Finally, we examine the relationship between the Head-Ries index and distance. We estimate a kernel-weighted local polynomial regression of the log Head-Ries index on the log distance between i and j. The fitted line is displayed in Figure 8.[18] The relationship is strongly negative, and the linear regression slope is -0.74. Note that this is quite close to distance coefficients estimated in international trade data, which are usually around -1. This suggests that geography impedes trade almost as much within countries as across countries. The fit of the linear regression model is also remarkably strong, with an $R^2 = 0.61$. Hence, distance alone is able to explain 61 percent of the bilateral variation in HR_{ij}, across 251,424 region pairs.[19]

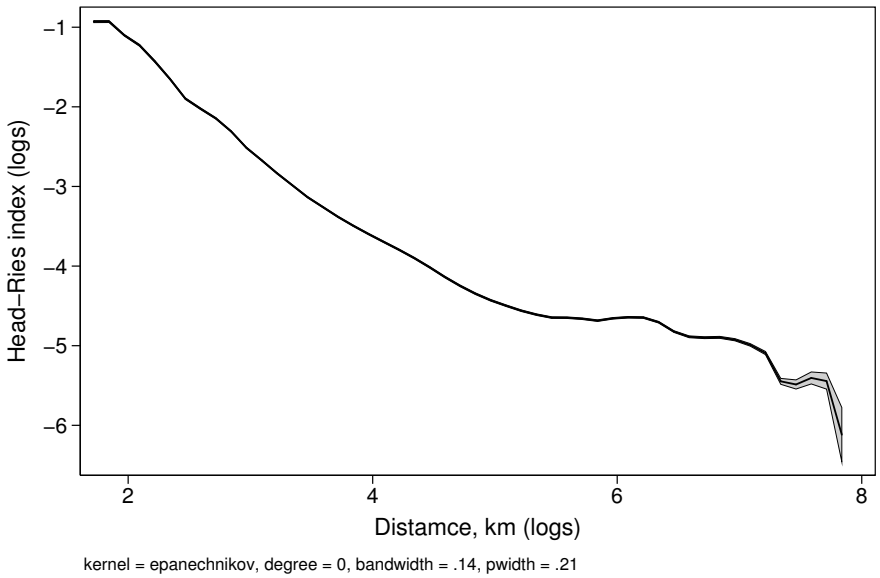
The absolute magnitude of the Head-Ries index is also of interest. In a broad class of models, iceberg trade costs between i and j relative to internal trade costs in i and j are $\tau_{ij} = HR_{ij}^{-1/\rho}$. Given knowledge about the trade elasticity ρ, one can therefore infer a model consistent measure of trade costs from the Head-Ries index. We choose $\rho = 4$, which is a common estimate in the literature (Simonovska and Waugh, forthcoming). The predicted τ_{ij}'s for the 5th, 25th, 50th, 75th

[18] Again, we do not show the underlying scatter, as the large number of data points obscures the fitted line.

[19] The number of observations is smaller than in the main sample because some regions have $X_{ii} = 0$, which makes $\ln HR_{ij}$ missing.

and 95th percentiles of the distance distribution are then 1.6, 2.3, 3.0, 3.2 and 3.4 respectively. The percentiles correspond to distances of 11 km, 43 km, 141 km, 403 km, 982 km respectively. Hence, even for very short distances of around 10 km, trade costs appear to be surprisingly high.

Figure 8: Trade Costs and Distance in the Production Network.



Note: The figure shows the kernel-weighted local polynomial regression of the log Head-Ries index (vertical axis) on log distance (horizontal axis). The gray area denotes the 95 percent confidence bands. The linear regression slope is -0.74.

7 Firm Performance and the Supplier Network

While the importance of both domestic and international supplier networks is acknowledged there is little direct evidence on the relationship between the network characteristics of suppliers and the downstream performance of customers.[20] In this section, we take a first step and examine the correlation between measures of firm (customer) performance and the characteristics of the firm's supplier network.[21]

The basic specification is

$$y_{fjr} = \delta_j + \delta_r + \beta \bar{y}_{fs} + \beta indeg_{fs} + \beta x_{fs} + \varepsilon_{frj}$$

where is y_{fjr} is a performance measure of firm f, $\bar{y}_{fs}$ is the average performance measure of firm f's suppliers, $indeg_{fs}$ is the log in-degree of firm f (# of suppliers) and x_{fs} is a vector of characteristics

[20] Javorcik XXX shows that trade and investment liberalization in service sectors has positive effects on downstream manufacturing productivity.

[21] We caution that one should not make causal inferences based on the results in this section

of the suppliers of firm f that can include the average distance to the suppliers, the average in-degree of firm f's suppliers and the average out-degree of firm f's suppliers, and δ_j and δ_r are 4-digit industry and prefecture fixed effects respectively.

Tables 4 and 5 report the results for a variety of performance measures including log sales, log sales per worker, log sales per customer per worker, credit score and growth in log sales. As expected, firm performance is positively correlated with the (log) number of suppliers and the (log) average performance of those suppliers, i.e. the direct measures of the firm's supply network.

More interesting, however, is the relationship between form performance and average distance to suppliers. Conditional on the number and quality of suppliers, distance is negatively correlated with measures of performance. The closer are the suppliers the better is the customer performance. We found earlier that "better" firms, as measured by the number of connections, have a higher average distance to their suppliers. We also showed that, within a firm, the quality of more distant suppliers is higher than that of close suppliers. In this result, we find that distance to the supply network is negatively correlated with firm performance. This result is consistent with models of economic geography and co-location that posit positive productive spillovers from firms operating in close proximity. Our findings emphasize the possibility of one form of that externality where the performance of downstream producers is better when suppliers are closer.

7.1 Network Effects of Suppliers

There is little formal theoretical guidance on the expected role of supplier networks on downstream firm performance. A supplier with a richer set of its own suppliers may be able to better aggregate and pass along (tacit) knowledge about the inputs it is providing. Such a connection would suggest that downstream firm performance would be positively related to the connectedness if its supply base. Models of geography and co-location often feature production externalities for firms in the same geographic but are silent about the importance of a dense supply chain. In contrast, a supplier serving more downstream customers may be sharing any such knowledge with other competing downstream firms, thus reducing the performance edge to its customers.

Controlling for the direct characteristics of a firm's suppliers (number, quality and distance), we add measures of supplier connectedness to the specification, for example columns 2-4, 6-8 etc. These supplier network characteristics include a measure of upstream connections and a measure of downstream connections of the firm's suppliers.

The average in-degree of suppliers is given by the average number of suppliers of a firm's suppliers. This measure captures how well-connected the suppliers are in the production network. It effectively looks back up the supply chain (upstream) and asks whether a customer's suppliers are themselves served by a greater number of firms. The other measure of supplier connectedness is the average out-degree of suppliers which is the average number of customers of a firm's suppliers.

This measure captures the customer base of the suppliers by looking forward (downstream) and is at the same level of the network as the (customer) firm itself.

The results in Tables 4 and 5 show a positive and significant correlation between downstream customer performance and the average in-degree of its suppliers. The greater the number of suppliers' suppliers the higher the performance measure. the magnitude of the effect can be benchmarked by comparing the point estimate to the coefficient on the firm's own in-degree. The effect of upstream connectedness is about one tenth that of direct connections by the firm.

The correlation between downstream firm performance and the average out-degree of its suppliers is negative and significant. the greater the number of customers of the firm's suppliers, the lower the measure of firm performance.

8 Identifying Networks Effects on Performance

In this section we use exogenous changes in the travel time between customers and suppliers to provide a better understanding of the causal relationships between geography, supplier connectedness and customer performance. We use the expansion of the high-speed (Shinkansen) rail network in recent years to help with identification. From 2004-present Japan has expanded its high-speed rail service, especially in the southern and northern parts of the country. This expansion has served to reduce travel times between locations that are now connected by Shinkansen trains.

Our identification strategy is similar to that employed by Giroud (2013) in examining the effects of travel time on plant investment and productivity outcomes within US firms. Japanese firms that see reductions in travel times to their suppliers due to the high-speed rail expansion should perform better than comparable firms that experience no travel time decline. The improvement in performance should be highest for customers who have travel time reduction to their best suppliers.

To be completed

9 Conclusion

We document the structure and geography of a large domestic production network and its relationship with firm performance. We have emphasized the role of domestic geographic frictions in shaping domestic relationships and the role of firm heterogeneity (productivity, firm size, number of customers, ownership) in overcoming these frictions.

References

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. 4, 3

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

Giroud, Xavier, "Proximity and Investment: Eveidence from Plant-level Data," *Quarterly Journal of Economics*, 2013, *128*, 861–915. 8

Imbens, Guido W., "An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling," *Econometrica*, 1992, *60(5)*, 1187–1214. 14

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, 8

King, Gary and Langche Zeng, "Logistic Regression in Rare Events Data," *Poliical Analysis*, 2001, *9(2)*, 137–163. 14

Kleinbaum, Adam, "Organizational Misfits and the Origins of Brokerage in Intrafirm Networks.," *Administrative Science Quarterly*, 2012, *57*, 407–452. 5

Lu, Dan, Asier Mariscal, and Luis Fernando Mejia, "Import Switching and the Impact of a Large Devaluation," Working Paper 2013. 3.1

Ohnishi, T., H. Takayasu, and M. Takayasu, "Network Motifs in Inter-firm Network," *Journal of Economic Internactions and Coordination*, 2010, *5(2)*, 171–180. 6

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. 6

Serrano, Ma Angeles and Marian Boguna, "Topology of the World Trade Web," *The Accounting Review*, 2003, *68* (1). 8

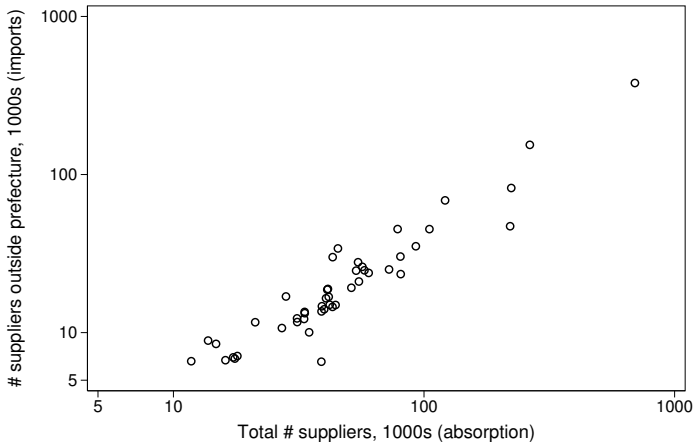
Simonovska, Ina and Michael Waugh, "The Elasticity of Trade: Estimates and Evidence,," *Journal of International Economics*, forthcoming. 6

Appendix

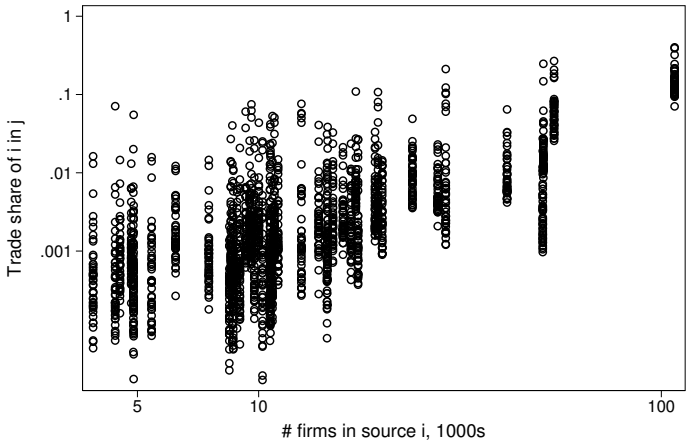
A Gravity in Prefecture Level Data

Figure 9: Gravity at the Prefecture Level.

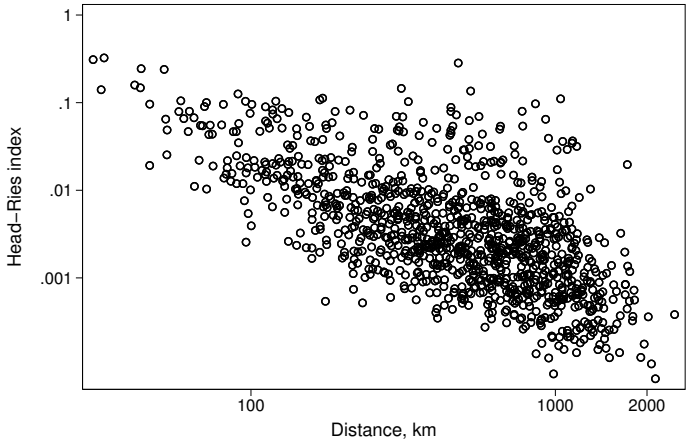
Panel A: Outside and Total Number of Suppliers.



Panel B: Trade Share of Number of Source Firms.



Panel C: Trade Costs and Distance in the Production Network.



In this section, we repeat the gravity analysis in the main text at a higher level of aggregation. We choose the prefecture level, which gives us 47 regions and $47^2 = 2209$ ij pairs (no pairs with zero trade). Panel A of Figure 9 shows the relationship between the total number of suppliers versus the number of suppliers from outside sources. The mean (median) import share is 0.43 (0.40). Panel B of Figure 9 shows the scatter between the number of firms in a source i versus the market share of i across prefectures. Panel C of Figure 9 shows the scatter between distance and the Head-Ries index. The fitted regression slope is -1.25.

Table 1: Probability of ij Links (firm i is a supplier)

	(1)	(2)	(3)	(4)
$Empl_i$	7.30^a			
$Empl_j$	5.89^a	7.96^a	20.10^a	20.10^a
$Empl_i \times Empl_j$			-3.04^a	-3.05^a
Ownership, i owns j				9.07^a
Ownership, j owns i				-2.50^a
Distance	-4.36^a	-5.81^a	-6.28^a	-6.25^a
Same island	27.1^a	30.3^a	31.4^a	31.4^a
Same prefecture	13.6^a	16.8^a	19.2^a	19.2^a
Same industry	14.5^a	19.6^a	20.4^a	20.3^a
i fixed effect	No	Yes	Yes	Yes
Prefecture fixed effect	Yes	Yes	Yes	Yes
N	6,129,065	6,154,031	6,129,065	6,129,065

	(2)	(3)	(4)
$Empl_i$	9.54^a	22.8^a	22.8^a
$Empl_i \times Empl_j$		-3.12^a	-3.11^a
Ownership, i owns j			-5.24^a
Ownership, j owns i			-3.14^a
Distance	-5.47^a	-6.07^a	-6.14^a
Same island	35.0^a	36.1^a	36.0^a
Same prefecture	19.1^a	21.1^a	20.9^a
Same industry	17.2^a	19.3^a	19.4^a
j fixed effect	Yes	Yes	Yes
Prefecture fixed effect	Yes	Yes	Yes
N	6,180,386	6,150,055	6,150,055

Note: Continuous variables in logs. All coefficients multiplied by 1e-6.

Table 2: Probability of Two-Way Trade

	(1)	(2)
$Empl_j$	2.95^a	
$Empl_i$		8.11^a
$Empl_i \times Empl_j$	-0.07^a	-0.35^a
Ownership link	284.99^a	274.37^a
Distance	-5.39^a	-4.25^a
Same industry	38.46^a	48.15^a
Same island	-0.55	-0.72
Same prefecture	-6.75^a	-8.48^a
i fixed effect	Yes	
j fixed effect		Yes
Prefecture fixed effect	Yes	Yes
N	3,074,057	3,093,664

Note: Continuous variables in logs. All coefficients multiplied by 1e-3.

Table 3: Connection Quality and Geography

Suppliers	In-degree	Out-degree	Employment	Sales/worker
Distance	0.12^a	0.17^a	0.15^a	0.04^a
Same industry	-0.29^a	$-.52^a$	-0.33^a	-0.08^a
Same island	-0.09^a	0.05^a	-0.08^a	0.04^a
Same prefecture	-0.40^a	-0.32^a	-0.45^a	-0.11^a
Seller prefecture FE	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes
N	3,080,414	3,345,585	3,340,659	3,335,784

Customers	In-degree	Out-degree	Employment	Sales/worker
Distance	0.07^a	0.06^a	0.06^a	0.01^a
	(.001)	(.001)	(.001)	(.000)
Same industry	-0.39^a	-0.15^a	-0.48^a	0.02^a
Same island	-0.17^a	-0.15^a	-0.17^a	-0.03^a
Same prefecture	-0.51^a	-0.45^a	-0.54^a	-0.14^a
Buyer prefecture FE	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes

Note: Each column represents a different dependent variable. Standard errors in parentheses. All variables in logs.

Table 4: Performance and the Supplier Network

	(1) Sales	(2) Sales	(3) Sales	(4) Sales	(5) Sales/Emp	(6) Sales/Emp	(7) Sales/Emp	(8) Sales/Emp
Sales (supplier avg.)	-0.00943^a	-0.0196^a	0.100^a	0.0678^a				
	(-12.86)	(-13.32)	(70.98)	(39.78)				
Sales/Emp (supplier avg.)					0.0501^a	0.0648^a	0.0991^a	0.0982^a
					(38.81)	(41.76)	(63.18)	(60.36)
Distance to suppliers (avg)	-0.0182^a	-0.0184^a	-0.0127^a	-0.0132^a	-0.0098^a	-0.0071^a	-0.0002	-0.00183^a
	(-18.83)	(-18.45)	(-13.24)	(-13.35)	(-15.92)	(-10.82)	(-0.316)	(-2.790)
# of Suppliers	1.012^a	1.021^a	0.997^a	1.003^a	0.278^a	0.281^a	0.273^a	0.273^a
	(612.9)	(604.9)	(604.5)	(595.2)	(254.1)	(252.1)	(248.7)	(243.9)
In-degree of suppliers (avg.)		0.0235^a		0.0960^a		-0.0110^a		0.0516^a
		(11.01)		(42.86)		(-13.01)		(40.53)
Out-degree of suppliers (avg.)			-0.171^a	-0.203^a			-0.0433^a	-0.0803^a
			(-90.71)	(-99.40)			(-54.79)	(-65.75)
Observations	653,815	632,587	653,815	632,587	652,083	631,027	652,083	631,027
R-squared	0.517	0.520	0.523	0.528	0.337	0.342	0.340	0.346

Table 5: Performance and the Supplier Network (cont.)

	(9) Score	(10) Score	(11) Score	(12) Score	(13) Growth	(14) Growth	(15) Growth	(16) Growth
S/C/E (supplier avg.)								
Score (supplier avg.)	0.0230^a	0.0327^a	0.0607^a	0.0485^a				
	(22.79)	(23.93)	(46.52)	(34.69)				
Growth (supplier avg.)					0.0422^a	0.0432^a	0.0431^a	0.0438^a
					(22.01)	(21.82)	(22.41)	(22.10)
Distance to suppliers (avg)	-0.111^a	-0.102^a	-0.0644^a	-0.0735^a	0.0004	0.0008^a	0.0009^a	0.0010^a
	(-26.12)	(-22.56)	(-14.75)	(-16.23)	(1.412)	(3.006)	(3.447)	(3.730)
# of Suppliers	2.652^a	2.678^a	2.615^a	2.629^a	0.0111^a	0.0111^a	0.0108^a	0.0107^a
	(353.2)	(347.5)	(346.9)	(339.3)	(24.96)	(24.56)	(24.21)	(23.66)
In-degree (supplier avg.)		-0.0539^a		0.280^a		-0.0011^a		0.0014^a
		(-8.103)		(30.42)		(-3.688)		(2.757)
Out-degree (supplier avg.)			-0.263^a	-0.430^a			-0.0017^a	-0.0029^a
			(-45.49)	(-52.29)			(-6.420)	(-6.027)
Observations	652,534	631,358	652,534	631,358	626,808	607,623	626,808	607,623
R-squared	0.263	0.262	0.266	0.265	0.014	0.015	0.014	0.015